Ruprecht-Karls-Universität Heidelberg
Institut für Angewandte Mathematik
Professor Dr. Jan JOHANNES

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Outline of the lecture course

STATISTICS 2

Winter semester 2022/23

Preliminary version: February 1, 2023

If you find **errors in the outline**, please send a short note
by email to johannes@math.uni-heidelberg.de.

MΛTHEMΛTIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
phone: +49 6221 54.14.190 – fax: +49 6221 54.14.101
email: johannes@math.uni-heidelberg.de
webpage: sip.math.uni-heidelberg.de

Table of contents

Chapter 1

Asymptotic properties of M- and Z-estimators

Asymptotic properties of M- and Z-estimators are presented generalising the minimum contrast approach introduced in the lecture Statistik 1. For a more detailed exposition we refer to the text book van der Vaart [1998].

Overview

§01 Introduction

§01|01 Motivation / illustration

§01.01 **Example** *(Linear model).* The dependence of the variation of a real random variable Y_1 (response) on the variation of a random vector $X_1 = (X_{1j})_{i\in[\![k]\!]}$ in $\mathbb{R}^k$ (explanatory variable) is often described by a linear relationship $\mathbb{E}\big(Y_1\big|X_1\big) = \sum_{j\in[\![k]\!]}\gamma_j X_{1j} = X_1^t\gamma$ or equivalently $Y_1 = X_1^t\gamma+\varepsilon_1$ where ε_1 is a real random error satisfying $\mathbb{E}\big(\varepsilon_1\big|X_1\big) = 0$. We aim to infer on the unknown parameter of interest $\gamma \in \mathbb{R}^k$ from $n \in \mathbb{N}$ i.i.d. copies (Y_i, X_i), $i \in [\![n]\!]$. Writing $Y := (Y_i)_{i\in[\![n]\!]}$ and $X^t = (X_1 \cdots X_n)$ we have $\mathbb{E}\big(Y\big|X\big) = X\gamma$. Any (measurable) choice

$$\widehat{\gamma} \in \operatorname*{arg\,inf}_{\gamma\in\mathbb{R}^k} \widehat{\mathrm{M}}_n(\gamma) \quad \text{with } \widehat{\mathrm{M}}_n(\gamma) := \tfrac{1}{n}\sum_{i\in[\![n]\!]}(Y_i - X_i^t\gamma)^2 = \tfrac{1}{n}\|Y - X\gamma\|^2 \tag{01.01}$$

is called a Least Squares Estimator (LSE), where arg inf denotes the subset of vectors in $\mathbb{R}^k$ attaining the function's smallest value. If $X^tX = \sum_{i\in[\![n]\!]} X_iX_i^t$ is strictly positive definite (hence, invertible) then $\widehat{\gamma} = (X^tX)^{-1}X^tY = \big(\sum_{i\in[\![n]\!]} X_iX_i^t\big)^{-1}\sum_{i\in[\![n]\!]} Y_iX_i$ is the unique LSE. Under "usual" conditions (Example §20.14) holds $\frac{1}{n}\sum_{i\in[\![n]\!]} X_iX_i^t \xrightarrow{\mathbb{P}} \mathbb{E}(X_1X_1^t) =: \Omega$ (LLN). If in addition $\mathbb{E}(\varepsilon_i^2|X_i) = \sigma^2$, then $\frac{1}{\sqrt{n}}\sum_{i\in[\![n]\!]}\varepsilon_iX_i \xrightarrow{d} \mathrm{N}_{(0,\sigma^2\Omega)}$ (CLT). Applying Slutzky's lemma §20.10 and the continuous mapping theorem §20.09 holds $\sqrt{n}(\widehat{\gamma}-\gamma) \xrightarrow{d} \mathrm{N}_{(0,\sigma^2\Omega^{-1})}$ for $\Omega > 0$. Further inference on $\widehat{\gamma}$ (hypothesis testing, confidence intervals, etc.) is typically based on this asymptotic result. However, a linear relationship $\mathbb{E}\big(Y\big|X\big) = X\gamma$ is often too restrictive. □

§01.02 **Example** *(Generalised linear model).* Consider a real random variable Y_1 and a random vector X_1 in $\mathbb{R}^k$ obeying $\mathbb{E}\big(Y_1\big|X_1\big) = g(X_1^t\gamma)$ for a known link function $g : \mathbb{R} \to \mathbb{R}$. We aim to infer on the unknown parameter of interest $\gamma \in \mathbb{R}^k$ from $n \in \mathbb{N}$ i.i.d. copies (Y_i, X_i), $i \in [\![n]\!]$.

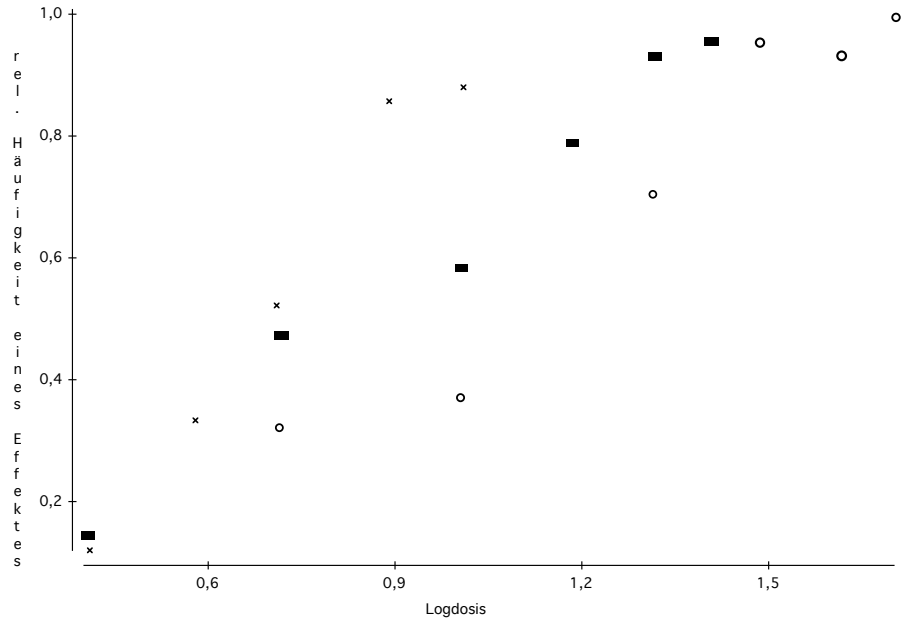
As an illustration let us consider the effect of three different drugs on the behaviour of certain animals. In a trial each drug is given in different dose to certain animals and the number of effected animals is counted. The Table 1.1 summarises the results. Let Y_{jk} denote the counts of an effect among n_{jk} animals applying a log-dose X_{jk}, $j \in [\![J_k]\!]$ of the drug $k \in [\![K]\!]$. Assuming an "independent and identical" behaviour of the n_{jk} animals it seems reasonable to model Y_{jk} as Binomial-distributed random variable, $Y_{jk} \sim \mathrm{Bin}_{(n_{jk},\pi_{jk})}$ for short, with unknown percentage $\pi_{jk} \in (0,1)$. It may be reasonable to assume that $n_{jk}\pi_{jk} = \mathbb{E}\big(Y_{jk}\big|X_{jk}\big) = g(\gamma_k + \gamma_0 X_{jk})$ where $(\gamma_k)_{k\in[\![K]\!]}$ is a drug specific factor and γ_0 is a common effect of the log-dose for all drugs. The model is called "probit" and "logit", respectively, if g is the standard-normal distribution function and the logit-distribution function $(x \mapsto \frac{e^x}{1+e^x})$. As in Example §01.01 inference on $\gamma = (\gamma_k)_{k\in[\![0,K]\!]}$ is often based on a LSE, i.e., any (measurable) choice $\widehat{\gamma} \in \arg\inf_{\gamma\in\mathbb{R}^{K+1}} \widehat{\mathrm{M}}_n(\gamma)$ with $\widehat{\mathrm{M}}_n(\gamma) := \frac{1}{K}\sum_{k\in[\![K]\!]} \frac{1}{J_K}\sum_{j\in[\![J_k]\!]}(Y_{jk} - g(\gamma_k + \gamma_0 X_{jk}))^2$.

Table 01 [§01]

drug	log-dose	effect	no effect	drug	log-dose	effect	no effect
1	1.01	44	6	2	1	18	30
1	0.89	42	7	2	0.71	16	33
1	0.71	24	22	3	1.4	48	2
1	0.58	16	32	3	1.31	43	3
1	0.41	6	44	3	1.18	38	10
2	1.7	48	0	3	1	27	19
2	1.61	47	3	3	0.71	22	24
2	1.48	47	2	3	0.4	7	40
2	1.31	34	14				

Number of animals exhibit an (no) effect in dependence of the drug's log-dose.

Figure 01 [§01]

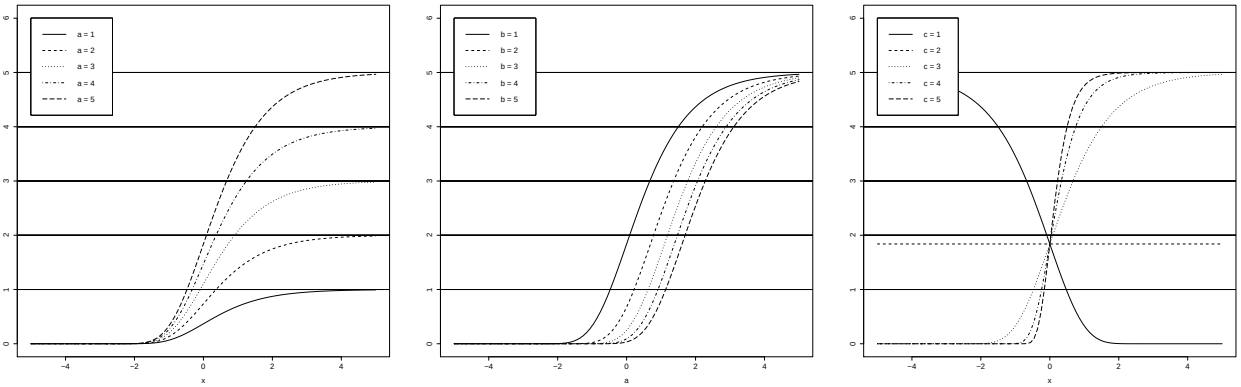


Relative frequency of the effects in dependence of the log-dose, drug 1: x; 2: ○; 3: -. □

§01.03 **Example** *(Nonlinear regression).* Consider a real random variable Y_1 and a random vector X_1 in

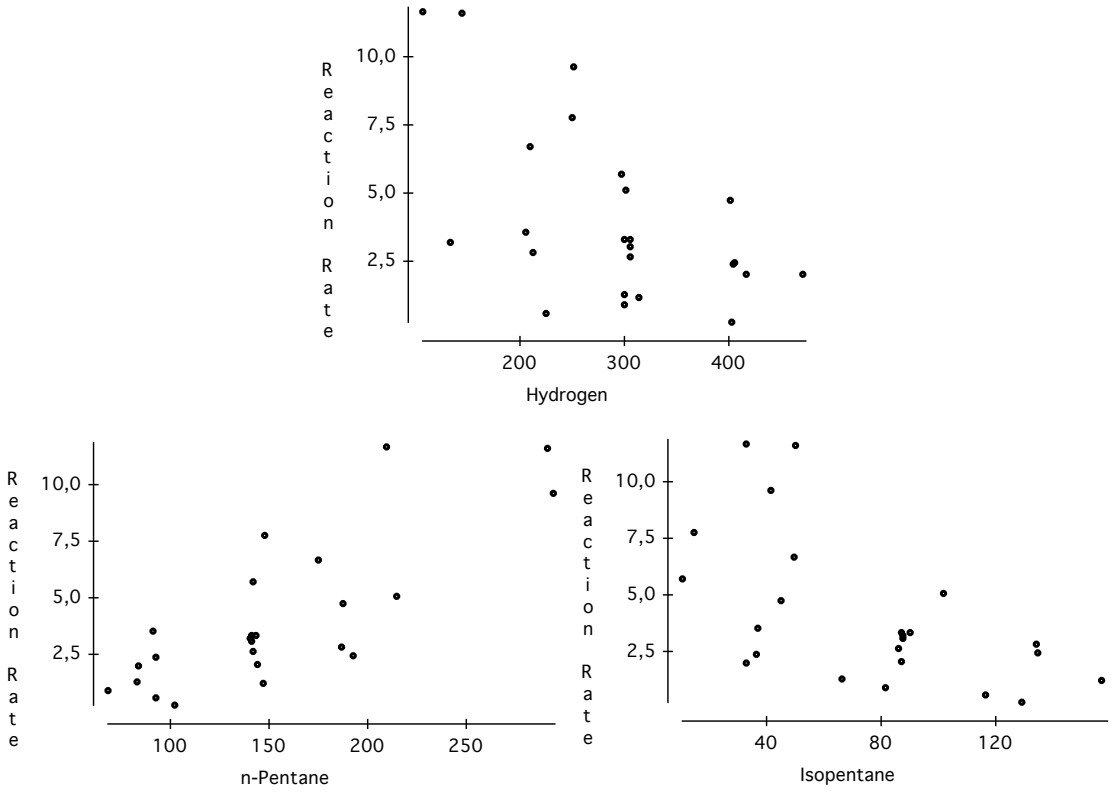
$\mathbb{R}^k$ obeying $\mathbb{E}\big(Y_1\big|X_1\big) = g(X_1, \gamma)$ for a given link function $g : \mathbb{R}^k \times \mathbb{R}^p \to \mathbb{R}$. We aim to infer on the unknown parameter $\gamma \in \mathbb{R}^p$ from $n \in \mathbb{N}$ i.i.d. copies (Y_i, X_i), $i \in [\![n]\!]$. The next figure shows the widely used Gompertz function $g(x, (a, b, c)) = a \exp(-b \exp(x \log(c)))$.

Figure 02 [§01]



As an illustration consider the following data of a reaction rate of a catalytic isomerisation of n-pentane into an isopentane given the partial pressure of hydrogen, n-pentane, and isopentane (see Carr [1960]). Isomerisation is a chemical process where a complex chemical product is transformed into basic elements. The reaction rate depends on several factors as for example, the partial pressure and the concentration of a catalyser (hydrogen).

Figure 03 [§01]



Reaction rate in dependence of the partial hydrogen, n-pentane and isopentane pressure.

Table 02 [§01]

Reaction rate	Partial pressure hydrogen	n-pentane	isopentane	Reaction rate	Partial pressure hydrogen	n-pentane	isopentane
3,541	205,8	90,9	37,1	5,686	297,3	142,2	10,5
2,397	404,8	92,9	36,3	1,193	314	146,7	157,1
6,694	209,7	174,9	49,4	2,648	305,7	142	86
4,722	401,6	187,2	44,9	3,303	300,1	143,7	90,2
0,593	224,9	92,7	116,3	3,054	305,4	141,1	87,4
0,268	402,6	102,2	128,9	3,302	305,2	141,5	87
2,797	212,7	186,9	134,4	1,271	300,1	83	66,4
2,451	406,2	192,6	134,9	11,648	106,6	209,6	33
3,196	133,3	140,8	87,6	2,002	417,2	83,9	32,9
2,021	470,9	144,2	86,9	9,604	251	294,4	41,5
0,896	300	68,3	81,7	7,754	250,3	148	14,7
5,084	301,6	214,6	101,7	11,59	145,1	291	50,2

Isomerisation reaction rate of an n-pentane into an isopentane.

A commonly used modelling for a reaction rate Y is the Hougen-Watson model where a special case is given by

$$\mathbb{E}\big(Y_i\big|(X_{i1},X_{i2},X_{i3})\big)=\frac{\gamma_1\gamma_3(X_{i2}-X_{i3}/1.632)}{1+\gamma_2X_{i1}+\gamma_3X_{i2}+\gamma_4X_{i3}},\quad i\in[\![n]\!], \tag{01.02}$$

where X_{i1}, X_{i2} and X_{i3} is the partial pressure of hydrogen, isopentane and n-pentane, respectively, and $(\gamma_j)_{j\in[\![4]\!]}$ is the unknown parameter of interest. As in Example §01.01 inference on γ is often based on a LSE, i.e., any (measurable) choice $\widehat{\gamma}\in\arg\inf_{\gamma\in\mathbb{R}^4}\widehat{\mathrm{M}}_n(\gamma)$ with $\widehat{\mathrm{M}}_n(\gamma):=\frac{1}{n}\sum_{i\in[\![n]\!]}(Y_i-g(X_i,\gamma))^2$. □

§01.04 **Example** *(Quantile regression).* Consider a real random variable Y_1 and a random vector X_1 in $\mathbb{R}^k$ obeying $Y_1=X_1^t\gamma+\varepsilon_1$ with quantile condition $\mathbb{P}\big(\varepsilon_1\leqslant 0\big|X_1\big)=\alpha$ for a given probability $\alpha\in(0,1)$ or equivalently $\mathbb{P}\big(Y_1\leqslant X_1^t\gamma\big|X_1\big)=\alpha$ meaning that the conditional-α-quantile of Y_1 given X_1 equals $X_1^t\gamma$. Let q_α denote the α-quantile of $\mathbb{P}^Z\in\mathcal{W}(\mathscr{B})$, i.e., $\mathbb{P}(Z\leqslant q_\alpha)=\alpha$. Define $\tau_\alpha(z):=(1-\alpha)z^-+\alpha z^+$ where $\tau_\alpha(z)=(1-\alpha)|z|$ if $z\leqslant 0$ and $\tau_\alpha(z)=\alpha z$ otherwise. Under regularity conditions the function $q\mapsto\mathbb{E}(\tau_\alpha(Z-q))$ attains its minimum at the value $q=q_\alpha$. Roughly, the α-quantile satisfies $0=\frac{\partial}{\partial q}\mathbb{E}(\tau_\alpha(Z-q))\big|_{q=q_\alpha}$, since

$$\begin{aligned}\frac{\partial}{\partial q}\mathbb{E}(\tau_\alpha(Z-q))&=(1-\alpha)\frac{\partial}{\partial q}\int_{-\infty}^{q}(q-z)f(z)dz+\alpha\frac{\partial}{\partial q}\int_{q}^{\infty}(z-q)f(z)dz\\&=(1-\alpha)\int_{-\infty}^{q}f(z)dz-\alpha\int_{q}^{\infty}f(z)dz\\&=(1-\alpha)\mathbb{P}(Z\leqslant q)-\alpha\mathbb{P}(Z>q)=\mathbb{P}(Z\leqslant q)-\alpha.\end{aligned}$$

Thereby, a reasonable estimator of γ is any (measurable) choice $\widehat{\gamma}\in\arg\inf_{\gamma\in\mathbb{R}^k}\widehat{\mathrm{M}}_n(\gamma)$ with $\widehat{\mathrm{M}}_n(\gamma)=\frac{1}{n}\sum_{i\in[\![n]\!]}\tau_\alpha(Y_i-X_i^t\gamma)$. □

§01.05 **Example** *(Generalised Method of Moments).* Given a random vector Z_1 in $\mathbb{R}^p$ and a function $\mathrm{h}^J=(\mathrm{h}_j)_{j\in[\![J]\!]}:\mathbb{R}^k\times\mathbb{R}^p\to\mathbb{R}^J$ let the unknown parameter of interest $\gamma\in\mathbb{R}^k$ satisfy $\mathbb{P}^{Z_1}\mathrm{h}_j(\gamma)=\mathbb{E}\big(\mathrm{h}_j(\gamma,Z_1)\big)=0$ for all $j\in[\![J]\!]$, or $\mathbb{P}^{Z_1}\mathrm{h}^J(\gamma)=\mathbb{E}\big(\mathrm{h}^J(\gamma,Z_1)\big)=0$ for short. Supposing an

i.i.d. sample $(Z_i)_{i\in[\![n]\!]}$ any (measurable) choice $\widehat{\gamma}$ satisfying $\widehat{\mathbb{P}}_n \mathrm{h}_j(\widehat{\gamma}) = \frac{1}{n}\sum_{i\in[\![n]\!]} \mathrm{h}_j(\widehat{\gamma}, Z_i) = 0$ for all $j \in [\![J]\!]$, or $\widehat{\mathrm{H}}_n(\widehat{\gamma}) = 0$ with $\widehat{\mathrm{H}}_n(\gamma) := \widehat{\mathbb{P}}_n \mathrm{h}^J(\gamma) = \frac{1}{n}\sum_{i\in[\![n]\!]} \mathrm{h}^J(\gamma, Z_i)$, $\gamma \in \mathbb{R}^k$, for short, is called moment estimator. In case a moment estimator does not exist, setting $\widehat{\mathrm{M}}_n(\gamma) := (\widehat{\mathbb{P}}_n \mathrm{h}^J(\gamma))^t W_n (\widehat{\mathbb{P}}_n \mathrm{h}^J(\gamma))$ for a given weighting matrix W_n one might consider any (measurable) choice $\widehat{\gamma} \in \arg\inf_{\gamma\in\mathbb{R}^k} \widehat{\mathrm{M}}_n(\gamma)$ called a Generalised Method of Moments (GMM) estimator. □

§01|02 Notation / definition

§01.06 **Reminder**. Denote by $\mathcal{W}(\mathscr{X})$ the set of all probability measures on a measurable space $(\mathcal{X}, \mathscr{X})$. For a non-empty index set Θ a family $\mathbb{P}_\Theta := (\mathbb{P}_\theta)_{\theta\in\Theta}$ of probability measures on $\mathscr{X}$ is formally defined by the map $\Theta \to \mathcal{W}(\mathscr{X})$ with $\theta \mapsto \mathbb{P}_\theta$. Here and subsequently, for each $\theta \in \Theta$ denotes $\mathbb{E}_\theta$ the expectation with respect to $\mathbb{P}_\theta$. For a random variable X taking its values in $(\mathcal{X}, \mathscr{X})$ we write shortly $X \leadsto \mathbb{P}_\Theta$, if $X \sim \mathbb{P}_\theta$ for some $\theta \in \Theta$. If the random variables $(X_i)_{i\in[\![n]\!]}$ form an *independent and identically distributed* (i.i.d.) sample of $X \sim \mathbb{P}$ with values in $(\mathcal{X}, \mathscr{X})$, then $\mathbb{P}^{\otimes n} = \otimes_{i\in[\![n]\!]}\mathbb{P}$ denotes the joint product probability measure of the family $(X_i)_{i\in[\![n]\!]}$ taking its values in the measurable product space $(\mathcal{X}^n, \mathscr{X}^{\otimes n})$. We write $(X_i)_{i\in[\![n]\!]} \overset{i.i.d.}{\sim} \mathbb{P}$ or $(X_i)_{i\in[\![n]\!]} \sim \mathbb{P}^{\otimes n}$ for short. We denote by $\mathbb{P}_\Theta^{\otimes n} := (\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}$ a family of product probability measures on $\mathscr{X}^{\otimes n}$. Any random variable S on $(\mathcal{X}, \mathscr{X})$ taking values in a measurable space $(\mathcal{S}, \mathscr{S})$, i.e., $\mathscr{X}$-$\mathscr{S}$-measurable function $S : \mathcal{X} \to \mathcal{S}$, is called *observation* or *statistic*. We denote by $\mathbb{P}_\Theta^S := (\mathbb{P}_\theta^S)_{\theta\in\Theta}$ the family of probability measures on $(\mathcal{S}, \mathscr{S})$ induced by S. A map $\gamma : \Theta \to \Gamma$ and its value $\gamma(\theta)$ for each $\theta \in \Theta$ is called *parameter* and *parameter value of interest*, respectively. A *parameter of interest* $\gamma : \Theta \to \Gamma$ is called *identifiable*, if for any $\theta_1, \theta_2 \in \Theta$ from $\gamma(\theta_1) \neq \gamma(\theta_2)$ follows $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. □

§01.07 **Definition**. The triple $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ is called a *statistical experiment* or *statistical model*. The non-empty set Θ and $\mathcal{X}$ is called *parameter* and *sample space*, respectively. A statistical model $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ is called *adequate* for a random variable X, if $X \leadsto \mathbb{P}_\Theta$. Given a family $\mathbb{P}_\Theta^{\otimes n}$ of product probability measures $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ is called a *statistical product experiment*. We denote by $(\mathcal{S}, \mathscr{S}, \mathbb{P}_\Theta^S)$ the statistical model induced by a $(\mathcal{S}, \mathscr{S})$-valued statistic S on $(\mathcal{X}, \mathscr{X})$. A statistic $\widehat{\gamma}$ on $(\mathcal{X}, \mathscr{X})$ with values in the measurable space $(\Gamma, \mathscr{G})$ is called *estimator* or *estimation function* for the identifiable parameter of interest γ. A statistical model $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ (and the family $\mathbb{P}_\Theta$) is called *dominated*, if a σ-finite measure μ on $\mathscr{X}$ exists, $\mu \in \mathcal{M}_\sigma(\mathscr{X})$ for short, such that for each $\theta \in \Theta$ the probability measure $\mathbb{P}_\theta$ is absolutely continuous with respect to μ, i.e., $\mathbb{P}_\theta \ll \mu$. We write shortly $\mathbb{P}_\Theta \ll \mu$. Any version of the Radon-Nikodym densities

$$\mathrm{L}(\theta, x) := \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mu}(x) \quad x \in \mathcal{X},\ \theta \in \Theta$$

considered as function of θ parametrised by x is called *likelihood* or *likelihood function* where typically it is understand as a random function $\mathrm{L} : \Theta \to \overline{\mathbb{R}}^+$ with $\theta \mapsto \mathrm{L}(\theta) := \mathrm{L}(\theta, \bullet)$. Its logarithm $\ell := \log \mathrm{L}$ (with convention $\log(0) := -\infty$) is called *log-likelihood* or *log-likelihood function*. The *likelihood* and *log-likelihood* in the corresponding dominated product experiment $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ are $\prod_{i\in[\![n]\!]} \mathrm{L}(\theta, x_i)$ and $\sum_{i\in[\![n]\!]} \ell(\theta, x_i)$, $\theta \in \Theta$, $x^n \in \mathcal{X}^n$, respectively. □

§01.08 **Reminder**. Let $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ be dominated by $\mu \in \mathcal{M}_\sigma(\mathscr{X})$. If μ is finite, then $\mu \ll \mathbb{P}_\mu := \frac{1}{\mu(\mathcal{X})}\mu \in \mathcal{W}(\mathscr{X})$ and hence $\mathbb{P}_\Theta$ is also dominated by $\mathbb{P}_\mu$. If μ is not finite, then there exists a countable and measurable partition $\{\mathcal{X}_m, m \in \mathbb{N}\}$ of $\mathcal{X}$ with $\mu(\mathcal{X}_m) \in \mathbb{R}^+_{\setminus 0}$ for all $m \in \mathbb{N}$. For each $m \in \mathbb{N}$ define $\mathbb{P}_\mu(\bullet\,|\mathcal{X}_m) \in \mathcal{W}(\mathscr{X})$ with $A \mapsto \mathbb{P}_\mu(A|\mathcal{X}_m) := \frac{\mu(A\cap\mathcal{X}_m)}{\mu(\mathcal{X}_m)}$. Then we have $\mu \ll \mathbb{P}_\mu := \sum_{m\in\mathbb{N}} 2^{-m}\mathbb{P}_\mu(\bullet\,|\mathcal{X}_m) \in \mathcal{W}(\mathscr{X})$, since $\mathbb{P}_\mu(A) = 0$ implies $\mu(A \cap \mathcal{X}_m) = 0$ for all $m \in \mathbb{N}$ and thus

$\mu(A) = 0$. Therewith, we have shown, that for each $\mu \in \mathcal{M}_\sigma(\mathscr{X})$ there is $\mathbb{P}_\mu \in \mathcal{W}(\mathscr{X})$ with $\mu \ll \mathbb{P}_\mu$ which automatically dominates $\mathbb{P}_\Theta$ too. On the other hand, there is a probability measure $\mathbb{P}_o = \sum_{i\in\mathbb{N}} c_i \mathbb{P}_{\theta_i}$ with $c_i \in \mathbb{R}^+$, $\theta_i \in \Theta$ for all $i \in \mathbb{N}$ and $\sum_{i\in\mathbb{N}} c_i = 1$, and thus $\mathbb{P}_o \ll \mu$, such that $\mathbb{P}_\theta \ll \mathbb{P}_o$ for all $\theta \in \Theta$ (e.g. Statistik 1, Satz §08.04). We call any such probability measure $\mathbb{P}_o$ *privileged dominating measure*. Therefore, we eventually assume with out loss of generality that the dominating measure is indeed a probability measure. □

§01.09 **Example** *(MLE).* Let $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ be a statistical model dominated by $\mu \in \mathcal{M}_\sigma(\mathscr{X})$ with likelihood $\mathrm{L}(\theta) = \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mu$ and log-likelihood $\ell(\theta) = \log \mathrm{L}(\theta)$ for $\theta \in \Theta$ and let $(\Theta, \mathscr{T})$ be a measurable space. Any statistic $\widehat{\theta}$ on $(\mathcal{X}, \mathscr{X})$ with values in $(\Theta, \mathscr{T})$ is called Maximum-Likelihood-Estimator (MLE) for θ, if $\mathrm{L}(\widehat{\theta}) = \sup_{\theta\in\Theta} \mathrm{L}(\theta)$ μ-a.s. meaning $\mathrm{L}(\widehat{\theta}(x), x) = \sup_{\theta\in\Theta} \mathrm{L}(\theta, x)$ for μ-a.e. $x \in \mathcal{X}$, or equivalently $\ell(\widehat{\theta}) = \sup_{\theta\in\Theta} \ell(\theta)$ μ-a.s.. Considering a statistical product experiment $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ dominated by $\mu^{\otimes n} \in \mathcal{M}_\sigma(\mathscr{X}^{\otimes n})$ and setting $\widehat{\mathrm{M}}_n(\theta) := -\widehat{\mathbb{P}}_n\ell(\theta)$, i.e. $\widehat{\mathrm{M}}_n(\theta, x^n) = -\frac{1}{n}\sum_{i\in[\![n]\!]} \ell(\theta, x_i)$ for $x^n \in \mathcal{X}^n$, the MLE $\widehat{\theta}$ is determined by $\widehat{\theta} \in \arg\inf_{\theta\in\Theta} \widehat{\mathrm{M}}_n(\theta)$ μ-a.s.. However, in general it is not guaranteed that MLE is unique or even exits. The MLE depends on the version of the likelihood, but there exists often a canonical choice. Furthermore, $\gamma(\widehat{\theta})$ is called MLE for a parameter of interest $\gamma : \Theta \to \Gamma$, if $\gamma(\widehat{\theta})$ is a statistic on $(\mathcal{X}^n, \mathscr{X}^{\otimes n})$ with values in $(\Gamma, \mathscr{G})$. □

§01.10 **Remark.** In all the examples the estimator $\widehat{\gamma}$ of the parameter of interest γ is determined by $\widehat{\gamma} \in \arg\inf_{\gamma\in\Gamma} \widehat{\mathrm{M}}_n(\gamma)$ for some random function $\gamma \mapsto \widehat{\mathrm{M}}_n(\gamma) \in \overline{\mathscr{X}}$ of the data. Obviously, rather than minimising (or maximising) a criterion function we might search for a zero of the associated normal or estimating equations, that is, $\widehat{\gamma}$ is determined as a zero of a random vector function $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma) \in \overline{\mathscr{X}}^k$. Note that estimator is defined $\mathbb{P}_\Theta$-a.s. only, meaning that one can change the estimator on a $\mathbb{P}_\Theta$-zero set N, i.e., $\mathbb{P}_\theta(N) = 0$ for all $\theta \in \Theta$. □

§01.11 **Definition.** Let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n = (\mathbb{P}_\theta^n)_{\theta\in\Theta}$ for all $n \in \mathbb{N}$ be a statistical model over the same parameter space Θ and let $\gamma : \Theta \to \Gamma$ be a parameter of interest. We call a function $\mathrm{M} : \Theta\times\Gamma \to \overline{\mathbb{R}}$ and $\mathrm{H} : \Theta \times \Gamma \to \overline{\mathbb{R}}^k$ *criterion function*, if for all $\theta \in \Theta$ the function $\mathrm{M}(\theta) : \gamma \mapsto \mathrm{M}(\theta, \gamma)$, respectively $\mathrm{H}(\theta) : \gamma \mapsto \mathrm{H}(\theta, \gamma)$, has in $\gamma(\theta)$ an unique minimum, respectively an unique zero. A sequence $(\widehat{\mathrm{M}}_n)_{n\in\mathbb{N}}$ and $(\widehat{\mathrm{H}}_n)_{n\in\mathbb{N}}$ of functions $\widehat{\mathrm{M}}_n : \Gamma \times \mathcal{X}_n \to \overline{\mathbb{R}}$ and $\widehat{\mathrm{H}}_n : \Gamma \times \mathcal{X}_n \to \overline{\mathbb{R}}^k$ is called *random criterion function* or *criterion process*, if the following two conditions are satisfied:

(CP1) For all $\gamma \in \Gamma$ is $\widehat{\mathrm{M}}_n(\gamma) : x \mapsto \widehat{\mathrm{M}}_n(\gamma, x)$, respectively $\widehat{\mathrm{H}}_n(\gamma) : x \mapsto \widehat{\mathrm{H}}_n(\gamma, x)$, a statistic, that is, $\widehat{\mathrm{M}}_n(\gamma) \in \overline{\mathscr{X}_n}$, respectively $\widehat{\mathrm{H}}_n(\gamma) \in \overline{\mathscr{X}_n}^k$.

(CP2) For all $\gamma \in \Gamma$ and $\theta \in \Theta$ it holds $\widehat{\mathrm{M}}_n(\gamma) \xrightarrow{\mathbb{P}_\theta^n} \mathrm{M}(\theta, \gamma)$, respectively $\widehat{\mathrm{H}}_n(\gamma) \xrightarrow{\mathbb{P}_\theta^n} \mathrm{H}(\theta, \gamma)$.

Every (measurable) choice $\widehat{\gamma}_n : \mathcal{X}_n \to \Gamma$ (if it exists) is called a *M*-estimator, respectively a *Z*-estimator, if it satisfies

$$\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) = \inf_{\gamma\in\Gamma} \widehat{\mathrm{M}}_n(\gamma) \quad \mathbb{P}_\Theta^n\text{-a.s.}, \qquad \text{respectively} \qquad \widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = 0 \quad \mathbb{P}_\Theta^n\text{-a.s.},$$

or more generally, if it is, respectively, a near minimum and near zero, that is, $\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) \leqslant \inf_{\gamma\in\Gamma} \widehat{\mathrm{M}}_n(\gamma) + \mathrm{o}_{\mathbb{P}^n}(1)$ and $\widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \mathrm{o}_{\mathbb{P}^n}(1)$. □

§01.12 **Remark.** There exists a measurable version of a minimum of an almost surely continuous function on a compact set (see Witting and Müller-Funk [1995], Satz 6.7). Note that in Definition §01.11 the criterion process $\widehat{\mathrm{M}}_n$ (respectively $\widehat{\mathrm{H}}_n$) is defined for each $n \in \mathbb{N}$ on a different measurable space. We write, however, shortly $\widehat{\mathrm{M}}_n(\gamma) \xrightarrow{\mathbb{P}_\theta^n} \mathrm{M}(\theta, \gamma)$, if for each $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ holds $\mathbb{P}_\theta^n(|\widehat{\mathrm{M}}_n(\gamma) - \mathrm{M}(\theta, \gamma)| > \varepsilon) \xrightarrow{n\to\infty} 0$. Let us briefly consider a sample $(X_i)_{i\in[\![n]\!]} \backsim \mathbb{P}_\Theta^{\otimes n}$ of a

random variable $X \bigcirc\!\!\!\sim \mathbb{P}_\ominus$. Keeping Notation §19.05 in mind $\mathbb{P}f$ and $\widehat{\mathbb{P}}_n f$ denotes the integral of $f \in \mathcal{L}_1(\mathscr{X}, \mathbb{P})$ with respect to $\mathbb{P}$ and the empirical measure $\widehat{\mathbb{P}}_n(x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]} \delta_{x_i}$, $x^n \in \mathcal{X}^n$, respectively. Revisiting each of the Examples §01.01 to §01.04 there is a function $\mathrm{m} : \Gamma \times \mathcal{X} \to \overline{\mathbb{R}}$ with $\mathrm{m}(\gamma) \in \mathcal{L}_1(\mathscr{X})$, $\gamma \in \Gamma$, such that the criterion process $\widehat{\mathrm{M}}_n$ and the associated criterion function M is for each $\gamma \in \Gamma$ given by $\widehat{\mathrm{M}}_n(\gamma) = \widehat{\mathbb{P}}_n\mathrm{m}(\gamma)$, i.e. $\widehat{\mathrm{M}}_n(\gamma, x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]} \mathrm{m}(\gamma, x_i)$, $x^n \in \mathcal{X}^n$, and $\mathrm{M}(\theta, \gamma) = \mathbb{P}_\theta\mathrm{m}(\gamma) = \int_{\mathcal{X}} \mathrm{m}(\gamma, x)\mathbb{P}_\theta(dx)$, respectively. Analogously, a moment estimator as in Example §01.05 is a Z-estimator. By construction in each example is the condition (CP1) and with the help of the LLN (see Remark §20.06) also the condition (CP2) satisfied. Note that the GMM estimator in Example §01.05 is also a M-estimator with criterion process satisfying (CP1) and (CP2). □

§01.13 **Definition**. For two probability measure $\mathbb{P}_0$ and $\mathbb{P}_1$ on a measurable space $(\mathcal{X}, \mathscr{X})$ is the function

$$\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) = \begin{cases} \mathbb{P}_0\big(\log \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1}\big) = \int \log\big(\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1}\big)\mathrm{d}\mathbb{P}_0, & \text{if } \mathbb{P}_0 \ll \mathbb{P}_1, \\ +\infty, & \text{otherwise} \end{cases}$$

called *Kullback-Leibler-divergence* of $\mathbb{P}_0$ with respect to $\mathbb{P}_1$. □

§01.14 **Reminder**. The Kullback-Leibler-divergence satisfies $\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) \geqslant 0$ as well as $\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) = 0$ if and only if $\mathbb{P}_0 = \mathbb{P}_1$, but $\mathrm{KL}(\bullet|\bullet)$ is not symmetric. Moreover, for product measures holds $\mathrm{KL}(\mathbb{P}_{0,1} \otimes \mathbb{P}_{0,2}|\mathbb{P}_{1,1} \otimes \mathbb{P}_{1,2}) = \mathrm{KL}(\mathbb{P}_{0,1}|\mathbb{P}_{1,1}) + \mathrm{KL}(\mathbb{P}_{0,2}|\mathbb{P}_{1,2})$ (e.g. Statistik 1, Lemma §17.03). □

§01.15 **Example** *(MLE, §01.09 continued.)*. Let $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\ominus^{\otimes n})$ be a statistical product experiment dominated by a privileged measure $\mathbb{P}_o \in \mathcal{W}(\mathscr{X})$ (see Reminder §01.08) with likelihood $\mathrm{L}(\theta) = \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_o$, log-likelihood $\ell = \log(\mathrm{L})$ and parameter of interest θ (i.e., $\gamma = \mathrm{id}_\Theta$). Furthermore, for all $\theta, \theta_o \in \Theta$ let $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta_o}$ be mutually dominated (i.e. $\mathbb{P}_\theta \ll \mathbb{P}_{\theta_o}$ and $\mathbb{P}_{\theta_o} \ll \mathbb{P}_\theta$, for short $\mathbb{P}_\theta \lll\ggg \mathbb{P}_{\theta_o}$), which implies $\mathbb{P}_{\theta_o} \lll\ggg \mathbb{P}_o$, and hence $-\mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_o) = \mathrm{KL}(\mathbb{P}_o|\mathbb{P}_{\theta_o})$. Then $\widehat{\mathrm{M}}_n(\theta) := -\widehat{\mathbb{P}}_n\ell(\theta) \in \overline{\mathscr{X}}^{\otimes n}$ with

$$x^n \mapsto \widehat{\mathrm{M}}_n(\theta, x^n) = -\frac{1}{n}\sum_{i\in[\![n]\!]} \ell(\theta, x_i)$$

is a criterion process associated to the criterion function $\mathrm{M}(\theta_o, \theta) := \mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_\theta) - \mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_o)$ assuming here and subsequently that the parameter θ is identifiable, that is, from $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ follows $\theta_1 = \theta_2$. Identifiability is a natural condition since it is a necessary condition for the existence of a consistent estimator. However, if θ is identifiable then $\theta \mapsto \mathrm{M}(\theta_o, \theta)$ attains its minimum $\mathrm{M}(\theta_o, \theta_o) = -\mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_o)$ uniquely at θ_o (keeping Reminder §01.14 in mind). The corresponding M-estimator is thus just a MLE. □

§02 Consistency

Here and subsequently, let (Γ, d) be a metric space endowed with its Borel-σ-algebra $\mathscr{G} := \mathscr{B}_\Gamma$, let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\ominus^n = (\mathbb{P}_\theta^n)_{\theta\in\Theta})$ for all $n \in \mathbb{N}$ be a statistical model over the parameter space Θ and let $\gamma : \Theta \to \Gamma$ be an identifiable parameter of interest.

§02.01 **Reminder**. For each $n \in \mathbb{N}$ let $\widehat{\gamma}_n$ be an estimator of γ, i.e. a statistic on $(\mathcal{X}_n, \mathscr{X}_n)$ with values in $(\Gamma, \mathscr{G})$. The sequence $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ of estimators is called *(weakly) consistent*, if for all $\varepsilon \in \mathbb{R}_{\setminus 0}^+$ holds $\mathbb{P}_\theta^n(d(\widehat{\gamma}_n, \gamma(\theta)) > \varepsilon) = \mathrm{o}(1)$ as $n \to \infty$ for all $\theta \in \Theta$. Note that the estimator $\widehat{\gamma}_n$ can be defined for each $n \in \mathbb{N}$ on a different measurable space. We write, however, shortly $d(\widehat{\gamma}_n, \gamma(\theta)) = \mathrm{o}_{\mathbb{P}^n}(1)$ as $n \to \infty$. Moreover, saying *„$\widehat{\gamma}_n$ is consistent“* always means the sequence $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ is (weakly) consistent. □

Consider an M-estimator $\widehat{\gamma}_n$ for a random criterion function $\widehat{\mathrm{M}}_n$ with associated criterion function M, that is, $\widehat{\mathrm{M}}_n(\gamma) \xrightarrow{\mathbb{P}_\theta^n} \mathrm{M}(\theta,\gamma)$ holds point-wise for each $\gamma \in \Gamma$. For example, due to the LLN $\widehat{\mathrm{M}}_n(\gamma) = \widehat{\mathbb{P}}_n\mathrm{m}(\gamma) \xrightarrow{\mathbb{P}_\theta^{\otimes n}} \mathbb{P}_\theta\mathrm{m}(\gamma) = \mathrm{M}(\theta,\gamma)$ provided $\mathrm{m}(\gamma) \in \mathcal{L}_1(\mathscr{X},\mathbb{P}_\theta)$. The hope is that a minimising value of $\widehat{\mathrm{M}}_n(\gamma)$ then converges to the minimising value of $\mathrm{M}(\theta,\gamma)$. However, in general point-wise convergence will not be sufficient.

§02.02 **Theorem**. *Under the assumptions and notations of Definition §01.11 any M-estimator $\widehat{\gamma}_n$ of γ, i.e., $\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) \leqslant \widehat{\mathrm{M}}_n(\gamma(\theta)) + \mathrm{o}_{\mathbb{P}_\theta^n}(1)$, is consistent, i.e., $d(\widehat{\gamma}_n,\gamma(\theta)) = \mathrm{o}_{\mathbb{P}_\theta^n}(1)$, if in addition the following two conditions are satisfied:*

(CO1) $\sup_{\gamma\in\Gamma}|\widehat{\mathrm{M}}_n(\gamma) - \mathrm{M}(\theta,\gamma)| = \mathrm{o}_{\mathbb{P}_\theta^n}(1)$ *(uniform convergence in probability);*

(CO2) $\inf_{\gamma\in\Gamma:d(\gamma,\gamma(\theta))\geqslant\varepsilon} \mathrm{M}(\theta,\gamma) > \mathrm{M}(\theta,\gamma(\theta))$ *for any* $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ *(identification).*

§02.03 **Proof** of Theorem §02.02. is given in the lecture. □

§02.04 **Corollary**. *Under the assumptions and notations of Definition §01.11 any Z-estimator $\widehat{\gamma}_n$ of γ, i.e., $\widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \mathrm{o}_{\mathbb{P}_\theta^n}(1)$, is consistent, i.e., $d(\widehat{\gamma}_n,\gamma(\theta)) = \mathrm{o}_{\mathbb{P}_\theta^n}(1)$, if in addition the following two conditions are satisfied:*

(CO1) $\sup_{\gamma\in\Gamma}\|\widehat{\mathrm{H}}_n(\gamma) - \mathrm{H}(\theta,\gamma)\| = \mathrm{o}_{\mathbb{P}_\theta^n}(1)$ *(uniform convergence in probability);*

(CO2) $\inf_{\gamma\in\Gamma:d(\gamma,\gamma(\theta))\geqslant\varepsilon} \|\mathrm{H}(\theta,\gamma)\| > 0 = \|\mathrm{H}(\theta,\gamma(\theta))\|$ *for any* $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ *(identification).*

§02.05 **Proof** of Corollary §02.04. is given in the lecture. □

§02.06 **Lemma**. *If* (i) Γ *is compact,* (ii) $\mathrm{M}(\theta,\gamma) > \mathrm{M}(\theta,\gamma(\theta))$ *for all* $\gamma \in \Gamma\setminus\{\gamma(\theta)\}$*, and* (iii) $\gamma \mapsto \mathrm{M}(\theta,\gamma)$ *is continuous, then* (CO2) *in Theorem §02.02 holds.*

§02.07 **Proof** of Lemma §02.06. is left as an exercise. □

§02.08 **Example** *(MLE, §01.15 continued).* Assuming in addition that the parameter space Θ is compact and that the criterion function $\theta \mapsto \mathrm{M}(\theta_o,\theta) := \mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_\theta) - \mathrm{KL}(\mathbb{P}_{\theta_o}|\mathbb{P}_o)$ is continuous then employing Lemma §02.06 the condition (CO2) of Theorem §02.02 is satisfied. □

§02.09 **Lemma**. (CO1) *in Theorem §02.02 is satisfied, if the following conditions hold:*

(i) (Γ,d) *is a compact metric space,*

(ii) $\gamma \mapsto \mathrm{M}(\theta,\gamma)$ *is continuous and* $\widehat{\mathrm{M}}_n(\gamma) = \mathrm{M}(\theta,\gamma) + \mathrm{o}_{\mathbb{P}_\theta^n}(1)$ *for all* $\gamma \in \Gamma$*, and*

(iii) $\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbb{P}_\theta^n\big(\sup_{\gamma_1,\gamma_2\in\Gamma:d(\gamma_1,\gamma_2)\leqslant\delta} |\widehat{\mathrm{M}}_n(\gamma_1) - \widehat{\mathrm{M}}_n(\gamma_2)| \geqslant \varepsilon\big) = 0$ *for all* $\varepsilon \in \mathbb{R}^+_{\setminus 0}$.

§02.10 **Proof** of Lemma §02.09. is given in the lecture. □

§02.11 **Example**. Given $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ and $\gamma : \Theta \to \Gamma$ for each $\gamma \in \Gamma$ let $\mathrm{m}(\gamma) \in \mathscr{X}$ be a real function $x \mapsto \mathrm{m}(\gamma,x)$ belonging to $\mathcal{L}_1(\mathscr{X},\mathbb{P}_\theta)$. Consider $\widehat{\mathrm{M}}_n(\gamma) := \widehat{\mathbb{P}}_n\mathrm{m}(\gamma)$, i.e. $\widehat{\mathrm{M}}_n(\gamma,x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]}\mathrm{m}(\gamma,x_i)$, $x^n \in \mathcal{X}^n$, and $\mathrm{M}(\theta,\gamma) := \mathbb{P}_\theta\mathrm{m}(\gamma)$ where due to the LLN §20.06 $\widehat{\mathrm{M}}_n(\gamma) = \mathrm{M}(\theta,\gamma) + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ for each $\gamma \in \Gamma$. Suppose in addition the following conditions:

(i) (Γ,d) is a compact metric space,

(ii) $\gamma \mapsto \mathrm{m}(\gamma,x)$ is continuous for $\mathbb{P}_\theta$-a.e. $x \in \mathcal{X}$,

(iii) there is $H \in \mathcal{L}_1(\mathscr{X},\mathbb{P}_\theta)$ with $\sup_{\gamma\in\Gamma}|\mathrm{m}(\gamma,x)| \leqslant |H(x)|$ for $\mathbb{P}_\theta$-a.e. $x \in \mathcal{X}$, or equivalently, $\sup_{\gamma\in\Gamma}|\mathrm{m}(\gamma)|$ belongs to $\mathcal{L}_1(\mathscr{X},\mathbb{P}_\theta)$.

Then, (I) $\gamma \mapsto \mathbb{P}_\vartheta \mathrm{m}(\gamma) = \mathrm{M}(\vartheta,\gamma)$ is continuous and (CO1) $\sup_{\gamma\in\Gamma}|\widehat{\mathrm{M}}_n(\gamma) - \mathrm{M}(\vartheta,\gamma)| = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$. Indeed, by dominated convergence (see §20.20) (ii) and (iii) imply together (I). Consider (CO1). Define the random variable $\Delta_\delta^n := \sup_{\gamma_1,\gamma_2\in\Gamma:d(\gamma_1,\gamma_2)\leqslant\delta}|\widehat{\mathrm{M}}_n(\gamma_1) - \widehat{\mathrm{M}}_n(\gamma_2)| \in \overline{\mathscr{X}}^{\otimes n}$. We show below for all $\varepsilon,\eta \in \mathbb{R}_{\setminus 0}^+$ exists $\delta \in \mathbb{R}_{\setminus 0}^+$ with $\limsup_{n\to\infty}\mathbb{P}_\vartheta^{\otimes n}\big(\Delta_\delta^n \geqslant \varepsilon\big) \leqslant \eta$ which in turn by Lemma §02.09 implies the claim (CO1). Let $\varepsilon,\eta \in \mathbb{R}_{\setminus 0}^+$. Keeping $\Delta_\delta^1 \in \overline{\mathscr{X}}$ with $x \mapsto \Delta_\delta^1(x) = \sup_{\gamma_1,\gamma_2\in\Gamma:d(\gamma_1,\gamma_2)\leqslant\delta}|\mathrm{m}(\gamma_1,x) - \mathrm{m}(\gamma_2,x)|$ in mind and applying the elementary triangular inequality we have $\Delta_\delta^n \leqslant \widehat{\mathbb{P}}_n\Delta_\delta^1$ point-wise on $\mathcal{X}^n$. Moreover, due to (i) and (ii) for $\mathbb{P}_\vartheta$-a.e. $x \in \mathcal{X}$ the function $\gamma \mapsto \mathrm{m}(\gamma,x)$ is uniformly continuous on Γ, and thus $\lim_{\delta\to 0}\Delta_\delta^1(x) = 0$. Therewith, dominated convergence (see §20.20), which can be applied due to (iii), implies $\lim_{\delta\to 0}\mathbb{P}_\vartheta\Delta_\delta^1 = 0$. In particular there is $\delta \in \mathbb{R}_{\setminus 0}^+$ such that $\mathbb{P}_\vartheta\Delta_\delta^1 \leqslant \eta\varepsilon$, which in turn implies $\mathbb{P}_\vartheta^{\otimes n}\Delta_\delta^n \leqslant \mathbb{P}_\vartheta^{\otimes n}(\widehat{\mathbb{P}}_n\Delta_\delta^1) = \mathbb{P}_\vartheta\Delta_\delta^1 \leqslant \eta\varepsilon$. Employing Markov's inequality §20.18 the last estimate implies the claim, that is, for all $\varepsilon,\eta \in \mathbb{R}_{\setminus 0}^+$ exists $\delta \in \mathbb{R}_{\setminus 0}^+$ with $\limsup_{n\to\infty}\mathbb{P}_\vartheta^{\otimes n}\big(\Delta_\delta^n \geqslant \varepsilon\big) \leqslant \eta$. If in addition to (i)-(iii) and, hence (I)

(iv) there is $\gamma(\vartheta) \in \Gamma$ with $\mathrm{M}(\vartheta,\gamma) > \mathrm{M}(\vartheta,\gamma(\vartheta))$ for all $\gamma \in \Gamma\setminus\{\gamma(\vartheta)\}$,

then applying Lemma §02.06 it holds (CO2) $\inf_{\gamma\in\Gamma:d(\gamma,\gamma(\vartheta))\geqslant\varepsilon}\mathrm{M}(\vartheta,\gamma) > \mathrm{M}(\vartheta,\gamma(\vartheta))$. To summarise, with (CO1) and (CO2) the conditions of Theorem §02.02 are satisfied. Consequently, any *M-estimator* $\widehat{\gamma}_n$, i.e., $\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) \leqslant \inf_{\gamma\in\Gamma}\widehat{\mathrm{M}}_n(\gamma) + \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$, and thus $\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) \leqslant \widehat{\mathrm{M}}_n(\gamma(\vartheta)) + \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$, is a *consistent estimator of* γ, i.e., $d(\widehat{\gamma}_n,\gamma(\vartheta)) = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$. □

§02.12 **Lemma**. (CO1) *in Corollary §02.04 is satisfied, if the following conditions hold:*

(i) (Γ,d) *is a compact metric space,*

(ii) $\gamma \mapsto \mathrm{H}(\vartheta,\gamma)$ *is continuous and* $\|\widehat{\mathrm{H}}_n(\gamma) - \mathrm{H}(\vartheta,\gamma)\| = \mathrm{o}_{\mathbb{P}_\vartheta^n}(1)$ *for all* $\gamma \in \Gamma$, *and*

(iii) $\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}_\vartheta^n\big(\sup_{\gamma_1,\gamma_2\in\Gamma:d(\gamma_1,\gamma_2)\leqslant\delta}\|\widehat{\mathrm{H}}_n(\gamma_1) - \widehat{\mathrm{H}}_n(\gamma_2)\| \geqslant \varepsilon\big) = 0$ *for all* $\varepsilon \in \mathbb{R}_{\setminus 0}^+$.

§02.13 **Proof** of Lemma §02.12. is left as an exercise. □

§02.14 **Example**. Given $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$, $\gamma : \Theta \to \Gamma$ and $(X_i)_{i\in[\![n]\!]} \sim \mathbb{P}_\vartheta^{\otimes n}$ for $\vartheta \in \Theta$, for each $\gamma \in \Gamma$ let $\mathrm{h}(\gamma) \in \overline{\mathscr{X}}^k$ be a numerical function belonging to $\mathcal{L}_1^k(\mathbb{P}_\vartheta)$ for all $\gamma \in \Gamma$. Consider $\widehat{\mathrm{H}}_n(\gamma) := \widehat{\mathbb{P}}_n\mathrm{h}(\gamma)$, i.e. $\widehat{\mathrm{H}}_n(\gamma,x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]}\mathrm{h}(\gamma,x_{\bullet}i)$, $x^n \in \mathcal{X}^n$, and $\mathrm{H}(\vartheta,\gamma) := \mathbb{P}_\vartheta\mathrm{h}(\gamma)$ where due to the LLN §20.06 $\|\widehat{\mathrm{H}}_n(\gamma) - \mathrm{H}(\vartheta,\gamma)\| = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$ for each $\gamma \in \Gamma$. Suppose in addition the following conditions:

(i) (Γ,d) is a compact metric space,

(ii) $\gamma \mapsto \mathrm{h}(\gamma,x)$ is continuous for $\mathbb{P}_\vartheta$-a.e. $x \in \mathcal{X}$,

(iii) $\sup_{\gamma\in\Gamma}\|\mathrm{h}(\gamma)\|$ belongs to $\mathcal{L}_1(\mathbb{P}_\vartheta)$.

Then, arguing line by line as in Example §02.11 (I) $\gamma \mapsto \mathbb{P}_\vartheta\mathrm{h}(\gamma) = \mathrm{H}(\vartheta,\gamma)$ is continuous and (CO1) $\sup_{\gamma\in\Gamma}\|\widehat{\mathrm{H}}_n(\gamma) - \mathrm{H}(\vartheta,\gamma)\| = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$. If in addition to (i)-(iii) and hence (I)

(iv) there is $\gamma(\vartheta) \in \Gamma$ with $\|\mathrm{H}(\vartheta,\gamma)\| > 0 = \|\mathrm{H}(\vartheta,\gamma(\vartheta))\|$ for all $\gamma \in \Gamma\setminus\{\gamma(\vartheta)\}$,

then applying Lemma §02.06 it holds (CO2) $\inf_{\gamma\in\Gamma:d(\gamma,\gamma(\vartheta))\geqslant\varepsilon}\|\mathrm{H}(\vartheta,\gamma)\| > 0 = \|\mathrm{H}(\vartheta,\gamma(\vartheta))\|$. To summarise, with (CO1) and (CO2) the conditions of Corollary §02.04 are satisfied. Consequently, any *Z-estimator* $\widehat{\gamma}_n$, i.e., $\widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$ is a *consistent estimator of* γ, i.e., $d(\widehat{\gamma}_n,\gamma(\vartheta)) = \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$. □

§02.15 **Remark**. The conditions (CO1) and (CO2) of Corollary §02.04 (respectively, (CO1) and (CO2) of Theorem §02.02) being sufficient to ensure consistency might be weakened in specific situations as we see next. □

§02.16 **Proposition**. *Let* $\Gamma \subseteq \mathbb{R}$ *and* $\widehat{\mathrm{H}}_n(\gamma) = \mathrm{H}(\theta,\gamma) + \mathrm{o}_{\mathbb{P}^n}(1)$ *for all* $\gamma \in \Gamma$ *where* H *is a deterministic function. Assume in addition that either*

(Ia) $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma)$ *is continuous and has exactly one zero* $\widehat{\gamma}_n$*, or*

(Ib) $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma)$ *is non-decreasing with* $\widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \mathrm{o}_{\mathbb{P}^n}(1)$*,*

and that (II) $\mathrm{H}(\theta,\gamma(\theta)-\varepsilon) < 0 < \mathrm{H}(\theta,\gamma(\theta)+\varepsilon)$ *for every* $\varepsilon \in \mathbb{R}^+_{\setminus 0}$*. Then,* $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}^n}(1)$.

§02.17 **Proof** of Proposition §02.16. is given in the lecture. □

§02.18 **Example**. Consider $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ and $\mathrm{h}(\gamma,t) := \mathrm{sign}(t-\gamma)$ with $\mathrm{sign}(t) := \mathbb{1}_{\{t\geqslant 0\}} - \mathbb{1}_{\{t\leqslant 0\}}$ for all $\gamma, t \in \mathbb{R}$. The sample median $\widehat{\gamma}_n$ is a (near) zero of the map $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma) := \widehat{\mathbb{P}}_n\mathrm{h}(\gamma)$, i.e. $\widehat{\mathrm{H}}_n(\gamma, x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]} \mathrm{h}(\gamma, x_i)$, $x^n \in \mathbb{R}^n$. Considering $\mathrm{H}(\gamma) = \mathbb{P}\mathrm{h}(\gamma) = \mathbb{P}((\gamma,\infty)) - \mathbb{P}((-\infty,\gamma))$ we have obviously $\widehat{\mathrm{H}}_n(\gamma) = \mathrm{H}(\gamma) + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$ for each $\gamma \in \Gamma$. Keeping in mind that $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma)$ is non-increasing from Proposition §02.16 follows consistency of the sample median $\widehat{\gamma}_n$, i.e., $\widehat{\gamma}_n = \gamma_o + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$, if for any $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ in addition $\mathrm{H}(\gamma_o-\varepsilon) > 0 > \mathrm{H}(\gamma_o+\varepsilon)$ or equivalently $\mathbb{P}((-\infty, <\gamma_o-\varepsilon)) < 1/2 < \mathbb{P}((-\infty,\gamma_o+\varepsilon))$. In other words, the sample median $\widehat{\gamma}_n$ is a consistent estimator of the population median, if it is unique. □

§03 Asymptotic normality

Here and subsequently, for $k, n \in \mathbb{N}$ let $\Gamma \subseteq \mathbb{R}^k$ be endowed with its Borel-σ-algebra $\mathscr{G} := \mathscr{B}_\Gamma$, let $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ be a statistical product experiment over the parameter space Θ and let $\gamma : \Theta \to \Gamma$ be an identifiable parameter of interest.

§03.01 **Heuristics**. Consider $\widehat{\mathrm{H}}_n(\gamma) = \widehat{\mathbb{P}}_n\mathrm{h}(\gamma)$, i.e. $\widehat{\mathrm{H}}_n(\gamma, x^n) = \frac{1}{n}\sum_{i\in[\![n]\!]} \mathrm{h}(\gamma, x_i)$, $x^n \in \mathcal{X}^n$, and $\mathrm{H}(\theta,\gamma) = \mathbb{P}_\theta\mathrm{h}(\gamma)$ for $\gamma \in \Gamma$ and $\theta \in \Theta$. Let $\widehat{\gamma}_n$ be a zero of $\gamma \mapsto \widehat{\mathrm{H}}_n(\gamma)$, i.e., $\widehat{\gamma}_n$ is a Z-estimator. Assume in addition that $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$ where $\gamma(\theta)$ is a zero of $\gamma \mapsto \mathrm{H}(\theta,\gamma)$. Heuristically, consider a *Taylor expansion* of a real-valued H around $\gamma(\theta) \in \Gamma \subseteq \mathbb{R}$, that is, $0 = \widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \widehat{\mathrm{H}}_n(\gamma(\theta)) + (\widehat{\gamma}_n - \gamma(\theta))\dot{\widehat{\mathrm{H}}}_n(\gamma(\theta)) + \frac{1}{2}(\widehat{\gamma}_n-\gamma(\theta))^2\ddot{\widehat{\mathrm{H}}}_n(\widetilde{\gamma}_n)$ for some $\widetilde{\gamma}_n$ between $\gamma(\theta)$ and $\widehat{\gamma}_n$. Thus, rewriting the last identity $\sqrt{n}(\widehat{\gamma}_n-\gamma(\theta)) = -\sqrt{n}\widehat{\mathrm{H}}_n(\gamma(\theta))\big(\dot{\widehat{\mathrm{H}}}_n(\gamma(\theta)) + \frac{1}{2}(\widehat{\gamma}_n-\gamma(\theta))\ddot{\widehat{\mathrm{H}}}_n(\widetilde{\gamma}_n)\big)^{-1}$. If $\mathrm{h}(\gamma(\theta))$ belongs to $\mathcal{L}_2(\mathbb{P}_\theta)$, then due to the CLT it holds $-\sqrt{n}(\widehat{\mathrm{H}}_n(\gamma(\theta)) - \mathrm{H}(\theta,\gamma(\theta))) = -\sqrt{n}(\widehat{\mathbb{P}}_n\mathrm{h}(\gamma(\theta)) - \mathbb{P}_\theta\mathrm{h}(\gamma(\theta))) \xrightarrow{d} \mathrm{N}_{(0,\mathbb{P}_\theta\mathrm{h}^2(\gamma(\theta)))}$. If moreover $\dot{\mathrm{h}}(\gamma(\theta)) \in \mathcal{L}_1(\mathbb{P}_\theta)$, then by the LLN $\dot{\widehat{\mathrm{H}}}_n(\gamma(\theta)) = \widehat{\mathbb{P}}_n\dot{\mathrm{h}}(\gamma(\theta)) = \mathbb{P}_\theta\dot{\mathrm{h}}(\gamma(\theta)) + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$. If in addition $\ddot{\widehat{\mathrm{H}}}_n(\widetilde{\gamma}_n) = \mathrm{O}_{\mathbb{P}^{\otimes n}}(1)$ then employing Slutzky's lemma §20.10 it follows $\sqrt{n}(\widehat{\gamma}_n-\gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,(\mathbb{P}_\theta\dot{\mathrm{h}}(\gamma(\theta)))^{-2}\mathbb{P}_\theta\mathrm{h}^2(\gamma(\theta)))}$. In the sequel, γ is a vector and h vector-valued. Consequently, $\dot{\mathrm{h}}(\gamma(\theta))$ is a matrix and we denote by $\|\dot{\mathrm{h}}(\gamma(\theta))\|_F$ its *Frobenius norm*, where $\|M\|_F := \big(\sum_{j\in[\![J]\!]}\sum_{k\in[\![K]\!]} M_{jk}^2\big)^{1/2}$ for any matrix $M = (M_{jk}) \in \mathbb{R}^{(J,K)}$. □

§03.02 **Theorem**. *Under the assumptions and notations of Definition §01.11 with* $\Gamma \subseteq \mathbb{R}^k$ *let* $\widehat{\gamma}_n$ *be a consistent Z-estimator of* γ*, i.e.* $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$*, with* $\widehat{\mathrm{H}}_n(\widehat{\gamma}_n) = \mathrm{o}_{\mathbb{P}^{\otimes n}}(n^{-1/2})$*. Assume the criterion process* $\widehat{\mathrm{H}}_n$ *is continuous differentiable in a neighbourhood* U *of* $\gamma(\theta) \in \mathrm{int}(\Gamma)$ *with derivative* $\dot{\widehat{\mathrm{H}}}_n := \frac{\partial}{\partial\gamma}\widehat{\mathrm{H}}_n \in \overline{\mathscr{X}}^{(k,k)}$ *and satisfies the following two conditions:*

(AN1) $\sqrt{n}\widehat{\mathrm{H}}_n(\gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\Omega_\theta)}$ *under* $\mathbb{P}_\theta^{\otimes n}$ *for some positive semidefinite* $\Omega_\theta \in \mathbb{R}^{(k,k)}$*,*

(AN2) $\sup_{\gamma\in U}\|\dot{\widehat{\mathrm{H}}}_n(\gamma) - \dot{\mathrm{H}}(\theta,\gamma)\|_F = \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$ *for some continuous matrix-valued function* $\gamma \mapsto \dot{\mathrm{H}}(\theta,\gamma)$ *with regular* $\dot{\mathrm{H}}(\theta,\gamma(\theta))$ *having* $\dot{\mathrm{H}}_\theta^{-1}$ *as inverse.*

Then $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta)) + \sqrt{n}\dot{\mathrm{H}}_\theta^{-1}\widehat{\mathrm{H}}_n(\gamma(\theta)) = \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$ *and* $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\dot{\mathrm{H}}_\theta^{-1}\Omega_\theta(\dot{\mathrm{H}}_\theta^{-1})^t)}$.

§03.03 **Proof** of Theorem §03.02. is given in the lecture. □

§03.04 **Corollary**. *Under the assumptions and notations of Definition §01.11 with* $\Gamma \subseteq \mathbb{R}^k$ *let* $\widehat{\gamma}_n$ *be a consistent* M*-estimator of* γ*, i.e.* $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$*, with* $\widehat{\mathrm{M}}_n(\widehat{\gamma}_n) = \inf_{\gamma\in\Gamma} \widehat{\mathrm{M}}_n(\gamma)$*. Assume the criterion process* $\widehat{\mathrm{M}}_n$ *is twice continuously differentiable in a neighbourhood* U *of* $\gamma(\theta) \in \mathrm{int}(\Gamma)$ *with derivatives* $\dot{\widehat{\mathrm{M}}}_n := \frac{\partial}{\partial\gamma}\widehat{\mathrm{M}}_n \in \overline{\mathscr{X}}^k$ *(score function) and* $\ddot{\widehat{\mathrm{M}}}_n := \frac{\partial^2}{\partial^2\gamma}\widehat{\mathrm{M}}_n \in \overline{\mathscr{X}}^{(k,k)}$ *and satisfies in addition the following two conditions:*

(AN1) $\sqrt{n}\dot{\widehat{\mathrm{M}}}_n(\gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\Omega_\theta)}$ *under* $\mathbb{P}_\theta^{\otimes n}$ *for some positive semidefinite* $\Omega_\theta \geqslant 0$*,*

(AN2) $\sup_{\gamma\in U}\|\ddot{\widehat{\mathrm{M}}}_n(\gamma) - \ddot{\mathrm{M}}(\theta,\gamma)\|_F = \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ *for some continuous matrix-valued function* $\gamma \mapsto \ddot{\mathrm{M}}(\theta,\gamma)$ *with regular* $\ddot{\mathrm{M}}(\theta,\gamma(\theta))$ *having* $\ddot{\mathrm{M}}_\theta^{-1}$ *as inverse.*

Then $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\ddot{\mathrm{M}}_\theta^{-1}\Omega_\theta\ddot{\mathrm{M}}_\theta^{-1})}$.

§03.05 **Proof** of Corollary §03.04. is given in the lecture. □

§03.06 **Example** *(§02.11 continued).* Given $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ and $\gamma : \Theta \to \Gamma$ for each $\gamma \in \Gamma$ let $\mathrm{m}(\gamma) \in \mathcal{L}_1(\mathbb{P}_\theta)$ be a real function. Consider $\widehat{\mathrm{M}}_n(\gamma) = \widehat{\mathbb{P}}_n\mathrm{m}(\gamma)$ and $\mathrm{M}(\theta,\gamma) = \mathbb{P}_\theta\mathrm{m}(\gamma)$ where due to the LLN $\widehat{\mathrm{M}}_n(\gamma) = \mathrm{M}(\theta,\gamma) + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ for each $\gamma \in \Gamma$. Suppose in addition that

(i) Γ is compact,

(ii) $\gamma \mapsto \mathrm{m}(\gamma, x)$ is twice continuously differentiable in a neighbourhood U of $\gamma(\theta) \in \mathrm{int}(\Gamma)$ for $\mathbb{P}_\theta$-a.e. $x \in \mathcal{X}$ with derivatives $\dot{\mathrm{m}} := \frac{\partial}{\partial\gamma}\mathrm{m}$ and $\ddot{\mathrm{m}} := \frac{\partial^2}{\partial^2\gamma}\mathrm{m}$

(iii) $\dot{\mathrm{m}}(\gamma(\theta)) \in \mathcal{L}_2(\mathbb{P}_\theta)$ with $\mathbb{P}_\theta\dot{\mathrm{m}}(\gamma(\theta)) = 0$ and $\Omega_\theta := \mathbb{P}_\theta\dot{\mathrm{m}}(\gamma(\theta))\dot{\mathrm{m}}(\gamma(\theta))^t \geqslant 0$,

(iv) $\sup_{\gamma\in U}\|\ddot{\mathrm{m}}(\gamma)\|_F \in \mathcal{L}_1(\mathbb{P}_\theta)$ and $\ddot{\mathrm{M}}_\theta := \mathbb{P}_\theta\ddot{\mathrm{m}}(\gamma(\theta))$ is regular with inverse $\ddot{\mathrm{M}}_\theta^{-1}$.

hold true. If the M-estimator satisfies $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ then $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\ddot{\mathrm{M}}_\theta^{-1}\Omega_\theta\ddot{\mathrm{M}}_\theta^{-1})}$ due to Corollary §03.04 since the conditions (AN1)-(AN2) are satisfied. Indeed, following Example §02.11, (iv) implies the condition (AN2) and due to the CLT the condition (AN1) follows from (iii). However, estimators of $\ddot{\mathrm{M}}_\theta$ and Ω_θ are necessary in order to use the asymptotic distribution to conduct inference. A typical approach to obtain these estimators is as follows. First replacing $\mathbb{P}_\theta$ by $\widehat{\mathbb{P}}_n$, the quantity $\ddot{\widehat{\mathrm{M}}}_n(\gamma) := \widehat{\mathbb{P}}_n\ddot{\mathrm{m}}(\gamma)$ and $\widehat{\Omega}_n(\gamma) = \widehat{\mathbb{P}}_n\dot{\mathrm{m}}(\gamma)\dot{\mathrm{m}}(\gamma)^t$ is just an empirical counterpart of $\ddot{\mathrm{M}}_\gamma(\gamma) = \mathbb{P}_\theta\ddot{\mathrm{m}}(\gamma)$ and $\Omega_\theta(\gamma) = \mathbb{P}_\theta\dot{\mathrm{m}}(\gamma)\dot{\mathrm{m}}(\gamma)^t$, respectively. Secondly, replace γ by its estimator $\widehat{\gamma}_n$ we obtain $\ddot{\widehat{\mathrm{M}}}_n := \ddot{\widehat{\mathrm{M}}}_n(\widehat{\gamma}_n)$ and $\widehat{\Omega}_n := \widehat{\Omega}_n(\widehat{\gamma}_n)$ as estimator of $\ddot{\mathrm{M}}_\theta = \ddot{\mathrm{M}}_\theta(\gamma(\theta))$ and $\Omega_\theta = \Omega_\theta(\gamma(\theta))$, respectively. If in addition to (i)-(iv) the following condition holds

(v) $\sup_{\gamma\in U}\|\dot{\mathrm{m}}(\gamma)\|$ belongs to $\mathcal{L}_2(\mathbb{P}_\theta)$.

Then $\sup_{\gamma\in U}\|\ddot{\widehat{\mathrm{M}}}_n(\gamma) - \ddot{\mathrm{M}}_\theta(\gamma)\|_F = \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ and $\sup_{\gamma\in U}\|\widehat{\Omega}_n(\gamma) - \Omega_\theta(\gamma)\|_F = \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ following line by line the arguments in Example §02.11. From these uniform convergences and $\widehat{\gamma}_n = \gamma(\theta) + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ follows $\ddot{\widehat{\mathrm{M}}}_n = \ddot{\mathrm{M}}_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ and $\widehat{\Omega}_n = \Omega_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ which in turn implies $\widehat{V}_n := \ddot{\widehat{\mathrm{M}}}_n^{-1}\widehat{\Omega}_n\ddot{\widehat{\mathrm{M}}}_n^{-1} = \ddot{\mathrm{M}}_\theta^{-1}\Omega_\theta\ddot{\mathrm{M}}_\theta^{-1} + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$. Consequently, by applying Slutzky's lemma §20.10 we have $\sqrt{n}\widehat{V}_n^{-1/2}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0,\mathrm{Id}_k)}$. □

§03.07 **Example** *(MLE, §01.15 continued).* Let $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ with $\mathbb{P}_\theta \ll\gg \mathbb{P}_o$ for all $\theta \in \Theta$, likelihood $\mathrm{L}(\theta) = \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_o$, log-likelihood $\ell = \log \mathrm{L}$ and parameter of interest θ (i.e., $\gamma = \mathrm{id}_\Theta$) as in Example §01.15. Consider the MLE $\widehat{\theta}_n$ which maximises the (joint) log-likelihood $\theta \mapsto \widehat{\mathbb{P}}_n\ell(\theta)$. Let the following conditions be satisfied:

(i) (Θ, d) is a compact metric space,

(ii) the parameter θ is identifiable, i.e., $\theta_1 \neq \theta_2$ implies $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$

(iii) the map $\theta \mapsto \ell(\theta, x)$ is continuous for $\mathbb{P}_\vartheta$-a.e. $x \in \mathcal{X}$,

(iv) $\sup_{\theta\in\Theta} |\ell(\theta)|$ belongs to $\mathcal{L}_1(\mathbb{P}_\vartheta)$.

Then combining the arguments in the Examples §02.08 and §02.11 the conditions (CO1) and (CO2) of Theorem §02.02 are satisfied, which in turn implies consistency of the MLE $\widehat{\theta}_n = \theta + \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$. In addition let the following conditions be fulfilled

(v) for $\mathbb{P}_\vartheta$-a.e. $x \in \mathcal{X}$ the map $\theta \mapsto \ell(\theta, x)$ is twice continuously differentiable in a neighbourhood U of $\theta \in \Theta$ with derivatives $\dot{\ell}_\theta := \frac{\partial}{\partial\theta}\ell$ and $\ddot{\ell}_\theta := \frac{\partial^2}{\partial^2\theta}\ell$,

(vi) $\sup_{\theta\in U}\|\dot{\ell}_\theta\| \in \mathcal{L}_2(\mathbb{P}_\vartheta)$ and $\sup_{\theta\in U}\|\ddot{\ell}_\theta\|_F \in \mathcal{L}_1(\mathbb{P}_\vartheta)$,

(vii) the Fisher-information matrix $\mathcal{I}_\theta := \mathbb{P}_\theta(\dot{\ell}_\theta\dot{\ell}_\theta^t)$ is strictly positive definite.

Then the conditions (AN1) and (AN2) of Corollary §03.04, and the identity $\mathcal{I}_\theta = -\mathbb{P}_\theta\ddot{\ell}_\theta$ are satisfied (for details see Statistik 1 Satz §17.22). Therewith, the MLE satisfies $\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}\mathcal{I}_\theta^{-1}\widehat{\mathbb{P}}_n\dot{\ell}_\theta + \mathrm{o}_{\mathbb{P}_\vartheta^{\otimes n}}(1)$ and, consequently, $\sqrt{n}(\widehat{\theta}_n - \theta_o) \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_\theta^{-1})}$. □

§03.08 **Remark.** The conditions (v) and (vi) in Example §03.07 can be weakened replacing differentiability by Hellinger-differentiability. Keeping the *Hellinger-distance* $H(\mathbb{P}_\theta, \mathbb{P}_{\theta_o}) = \|\mathrm{L}^{1/2}(\theta) - \mathrm{L}^{1/2}(\theta_o)\|_{\mathcal{L}_2(\mathbb{P})}$ in mind, where $\mathrm{L}^{1/2}(\theta) \in \mathcal{L}_2(\mathbb{P}_\vartheta)$ using $\|\mathrm{L}^{1/2}(\theta)\|^2_{\mathcal{L}_2(\mathbb{P})} = \mathbb{P}_\vartheta(\mathrm{L}(\theta)) = 1 < \infty$, the family $\mathbb{P}_\Theta$ is called *Hellinger-differentiable with derivative* $\dot{\ell}_{\theta_o}$ in $\theta_o \in int(\Theta) \subseteq \mathbb{R}^k$, if $\dot{\ell}_{\theta_o} \in \mathcal{L}_2^k(\mathbb{P}_{\theta_o})$ and hence $\dot{\ell}_{\theta_o}\mathrm{L}^{1/2}(\theta_o) \in \mathcal{L}_2^k(\mathbb{P}_\vartheta)$ such that

$$\lim_{\theta\to\theta_o}\int_{\mathcal{X}}\left|\frac{\mathrm{L}^{1/2}(\theta,x) - \mathrm{L}^{1/2}(\theta_o,x) - \frac{1}{2}\langle\dot{\ell}_{\theta_o}(x), \theta-\theta_o\rangle\mathrm{L}^{1/2}(\theta_o,x)}{\|\theta-\theta_o\|}\right|^2\mathbb{P}_\vartheta(dx)$$
$$= \lim_{h\to 0}\frac{\|\mathrm{L}^{1/2}(\theta_o+h) - \mathrm{L}^{1/2}(\theta_o) - \frac{1}{2}\langle\dot{\ell}_{\theta_o}, h\rangle\mathrm{L}^{1/2}(\theta_o)\|^2_{\mathcal{L}_2(\mathbb{P})}}{\|h\|^2} = 0$$

The map $x \mapsto \dot{\ell}_{\theta_o}(x)$ is also called *score function*. Keeping $\dot{\ell}_{\theta_o} \in \mathcal{L}_2^k(\mathbb{P}_{\theta_o})$ in mind the *Fisher-information* matrix $\mathcal{I}_{\theta_o} = \mathbb{P}_{\theta_o}(\dot{\ell}_{\theta_o}\dot{\ell}_{\theta_o}^t)$ is well-defined. Note that, the score function and the Fisher-information matrix are independent of the dominating measure $\mathbb{P}_o$. □

§03|01 Testing procedures

§03.09 **Heuristics.** Let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)$ for all $n \in \mathbb{N}$ be a statistical model over the parameter space Θ and let $\gamma : \Theta \to \Gamma$ be an identifiable parameter of interest. Given a map $A : \Gamma \to \mathbb{R}^p$ we eventually test the hypothesis $H_0 : A(\gamma) = 0$ against the alternative $H_1 : A(\gamma) \neq 0$. Typical examples include $A(\gamma) = \gamma - \gamma_o$ for a given value γ_o, or more generally, linear hypothesis $A(\gamma) = M\gamma - a_o$ for a given value a_o and matrix M. It covers in particular testing the j-th coordinate of $\gamma = (\gamma^j)_{j\in[\![k]\!]}$, i.e., $A(\gamma) = \gamma^j - \gamma_o^j$. Under regularity conditions it seems reasonable to assume an estimator $\widehat{\gamma}_n$ of γ having under $\mathbb{P}_\theta^n$ the property $\sqrt{n}(A(\widehat{\gamma}_n) - A(\gamma(\theta))) \xrightarrow{d} \mathrm{N}_{(0,\Sigma_\theta)}$ with invertible asymptotic covariance matrix Σ_θ. If we have in addition an estimator $\widehat{\Sigma}_n = \Sigma_\theta + \mathrm{o}_{\mathbb{P}_\theta^n}(1)$ at hand. Then under the hypothesis H_0, i.e., for $\mathbb{P}_\theta^n$ with $A(\gamma(\theta)) = 0$, a *Wald test* exploits the property $\widehat{W}_n := nA(\widehat{\gamma}_n)^t\widehat{\Sigma}_n^{-1}A(\widehat{\gamma}_n) \xrightarrow{d} \chi^2_p$ where χ^2_p is a Chi-square-distribution with p degrees of freedom. Precisely, a *Wald test* rejects the hypothesis $H_0 : A(\gamma) = 0$ if $\widehat{W}_n$ exceeds the 1-α-Quantile $\chi^2_{p,1-\alpha}$ of a χ^2_p-distribution. Obviously, the Wald test does exactly meets the asymptotic level α, i.e., $\lim_{n\to\infty}\mathbb{P}_\theta^n(\widehat{W}_n \geqslant \chi^2_{p,1-\alpha}) = \mathbb{P}(W \geqslant \chi^2_{p,1-\alpha}) = \alpha$ where $W \sim \chi^2_p$. However, the behaviour of the test statistic $\widehat{W}_n$ under the alternative H_1 is still an open questions, which we intent to study in the next sections. □

§03.10 **Example** *(§03.06 continued).* Let $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$, $\gamma : \Theta \to \Gamma$ be an identifiable parameter of interest and let $\mathrm{m}(\gamma) \in \mathcal{L}_1(\mathbb{P}_\theta)$ for all $\gamma \in \Gamma$. For each $\gamma \in \Gamma$ let $\widehat{\mathrm{M}}_n(\gamma) = \widehat{\mathbb{P}}_n \mathrm{m}(\gamma)$ and $\mathrm{M}(\theta, \gamma) = \mathbb{P}_\theta \mathrm{m}(\gamma)$. Under the conditions (i)-(v) in Example §03.06 an M-estimator $\widehat{\gamma}_n \in \arg\inf_{\gamma\in\Gamma} \widehat{\mathrm{M}}_n(\gamma)$ satisfies $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{d} \mathrm{N}_{(0, \ddot{\mathrm{M}}_\theta^{-1}\Omega_\theta\ddot{\mathrm{M}}_\theta^{-1})}$ under $\mathbb{P}_\theta^{\otimes n}$. Moreover, we have eventually access to estimators $\widehat{\ddot{\mathrm{M}}}_n = \ddot{\mathrm{M}}_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ and $\widehat{\Omega}_n = \Omega_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$. Let $A : \Gamma \to \mathbb{R}^p$ be continuously differentiable in a neighbourhood of $\gamma(\theta)$ then applying the delta method §20.16 we obtain $\sqrt{n}(A(\widehat{\gamma}_n) - A(\gamma(\theta))) \xrightarrow{d} \mathrm{N}_{(0,\Sigma_\theta)}$ under $\mathbb{P}_\theta^{\otimes n}$ with $\Sigma_\theta := \dot{A}_{\gamma(\theta)}\ddot{\mathrm{M}}_\theta^{-1}\Omega_\theta\ddot{\mathrm{M}}_\theta^{-1}\dot{A}_{\gamma(\theta)}^t$. From $\dot{A}_{\widehat{\gamma}_n} = \dot{A}_{\gamma(\theta)} + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ follows $\widehat{\Sigma}_n := \dot{A}_{\widehat{\gamma}_n}\widehat{\ddot{\mathrm{M}}}_n^{-1}\widehat{\Omega}_n\widehat{\ddot{\mathrm{M}}}_n^{-1}\dot{A}_{\widehat{\gamma}_n}^t = \Sigma_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$ and, thus $\sqrt{n}\widehat{\Sigma}_n^{-1/2}(A(\widehat{\gamma}_n) - A(\gamma(\theta))) \xrightarrow{d} \mathrm{N}_{(0,\mathrm{Id}_p)}$ which under H_0, i.e., for $\mathbb{P}_\theta^{\otimes n}$ with $A(\gamma(\theta)) = 0$, implies $\widehat{W}_n := nA(\widehat{\gamma}_n)^t\widehat{\Sigma}_n^{-1}A(\widehat{\gamma}_n) \xrightarrow{d} \chi^2_p$. □

Chapter 2

Asymptotic properties of tests

Asymptotic properties of tests under local alternatives are presented complementing the Neyman-Pearson theory introduced in the lecture Statistik 1. For a more detailed exposition we refer to the text books Witting and Müller-Funk *[1995] and* van der Vaart *[1998].*

Overview

§04 Contiguity

§04|01 Preliminaries: likelihood ratios and differentiable models

§04.01 **Motivation.** Considering a statistical model $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)$, a parameter of interest $\gamma : \Theta \to \Gamma$, a partition $\{\mathscr{H}^0, \mathscr{H}^1\}$ of the parameter values of interests $\Gamma = \mathscr{H}^0 \biguplus \mathscr{H}^1$ (i.e. $\Gamma = \mathscr{H}^0 \cup \mathscr{H}^1$, $\emptyset = \mathscr{H}^0 \cap \mathscr{H}^1$ and $\mathscr{H}^0 \neq \emptyset \neq \mathscr{H}^1$) we are interested in a (randomised) test $\varphi_n \in \mathscr{X}_n^+$ (i.e. $\varphi_n : \mathcal{X}_n \to [0,1]$) of the hypothesis $H_0 : \mathscr{H}^0$ against the alternative $H_1 : \mathscr{H}^1$. Under regularity conditions we may have at hand an estimator $\widehat{\gamma}_n$ of γ with known asymptotic distribution. Typically the estimator $\widehat{\gamma}_n$ allows us to construct a test statistic T_n with known asymptotic distribution under H_0, i.e. under $\mathbb{P}_\theta^n$ with $\gamma(\theta) \in \mathscr{H}^0$. Exploiting the asymptotic distribution an associated test $\varphi_n = \mathbb{1}_{\{T_n \notin C_\alpha\}}$ does eventually not exceed asymptotically a given level $\alpha \in (0,1)$ under the hypothesis H_0. However, we like to investigate also its power under the alternative H_1, i.e. under a specific $\mathbb{P}_\theta^n$ with $\gamma(\theta) \in \mathscr{H}^1$. □

§04.02 **Reminder.** Let ν and μ be measures on $(\mathcal{X}, \mathscr{X})$.

(a) For any positive numerical function $\mathbb{f} \in \overline{\mathscr{X}}^+$ the map $B \mapsto \mathbb{f}\mu(B) := \mu(\mathbb{1}_B \mathbb{f}) = \int_B \mathbb{f}\, d\mu$ defines a measure $\mathbb{f}\mu$ on $(\mathcal{X}, \mathscr{X})$. Any $\mathbb{f} = d\nu/d\mu \in \overline{\mathscr{X}}^+$ satisfying $\nu = \mathbb{f}\mu$ is called *density* of ν with respect to μ, or *μ-density* for short.

(b) We say ν is *dominated* by μ, symbolically $\nu \ll \mu$, if for each $B \in \mathscr{X}$ with $\mu(B) = 0$ follows $\nu(B) = 0$. The measures μ and ν are called *equivalent* or *mutually dominated*, symbolically $\mu \lll \nu$, if both $\nu \ll \mu$ and $\mu \ll \nu$.

(c) We say ν and μ are *orthogonal* or *singular*, symbolically $\nu \perp \mu$, if there exists $\mathcal{X} = \mathcal{X}_\mu \biguplus \mathcal{X}_\nu$ with $\mathcal{X}_\mu, \mathcal{X}_\nu \in \mathscr{X}$ and $\mu(\mathcal{X}_\nu) = 0 = \nu(\mathcal{X}_\mu)$. Evidently, we have $\nu \perp \mu$ if and only if there exists $N \in \mathscr{X}$ with $\mu(N) = 0$ such that $\nu = \mathbb{1}_N \nu$.

We note that $g \in \mathcal{L}_1(\mathfrak{f}\mu)$ if and only if $g\mathfrak{f} \in \mathcal{L}_1(\mu)$. In this case holds $\mathfrak{f}\mu(g) = \int g\, d(\mathfrak{f}\mu) = \int (g\mathfrak{f})\, d\mu = \mu(g\mathfrak{f})$ (Klenke [2012], Satz 4.15, p. 93). Let additionally $\nu \in \mathcal{M}_\sigma(\mathscr{X})$ be a σ-finite measure on $(\mathcal{X}, \mathscr{X})$. If $\mathfrak{f}_1\mu = \nu = \mathfrak{f}_2\mu$ for $\mathfrak{f}_1, \mathfrak{f}_2 \in \overline{\mathscr{X}}^+$, then $\mathfrak{f}_1 = \mathfrak{f}_2$ μ-a.e.. In other words a density is unique up to μ-a.e. equivalence (Klenke [2012], Satz 7.29, p. 159). If in addition $\mu \in \mathcal{M}_\sigma(\mathscr{X})$, then by *Lebesgue's decomposition theorem* there exists $\nu^\mu, \nu^\perp \in \mathcal{M}_\sigma(\mathscr{X})$ such that $\nu = \nu^\mu + \nu^\perp$ with $\nu^\perp \perp \mu$ and $\nu^\mu = \mathfrak{f}\mu$ where $\mathfrak{f} \in \overline{\mathscr{X}}^+$ and $\mathfrak{f} \in \mathbb{R}^+$ μ-a.e.. (Klenke [2012], Satz 7.33, p. 160) Furthermore, there is a *Radon-Nikodym-density* $\mathfrak{f} \in \overline{\mathscr{X}}^+$ with $\nu = \mathfrak{f}\mu$ and $\mathfrak{f} \in \mathbb{R}^+$ μ-a.e. if and only if $\nu \ll \mu$ (Klenke [2012], Korollar 7.34, p. 161). If $\mathfrak{f} \in \mathscr{X}^+$ is a Radon-Nikodym-density of ν with respect to μ, i.e. $\nu = \mathfrak{f}\mu$, then the positive real function $\mathfrak{f}\mathbb{1}_{\{\mathfrak{f}\in\mathbb{R}^+\}} \in \mathscr{X}^+$ is it too. Consequently, without loss of generality we consider here and subsequently a positive real version of the Radon-Nikodym-density. Furthermore, given $\mathfrak{f} = \mathrm{d}\,\nu^\mu/\mathrm{d}\mu \in \mathscr{X}^+$ let us define a numerical function $\mathrm{L} := \mathfrak{f}\mathbb{1}_{N^c} + \infty\mathbb{1}_N \in \overline{\mathscr{X}}^+$ with $\mu(N) = 0 = \nu^\perp(N^c)$ where $\{\mathrm{L} = \infty\} = N$ and the Lebesgue decomposition writes $\nu = \mathrm{L}\mu + \mathbb{1}_{\{\mathrm{L}=\infty\}}\nu$, i.e. for all $A \in \mathscr{X}$ we have $\nu(A) = \mu(\mathbb{1}_A\mathrm{L}) + \nu(A \cap \{\mathrm{L} = \infty\})$. □

§04.03 **Definition**. Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathscr{X})$ be probability measures on $(\mathcal{X}, \mathscr{X})$. Any positive numerical random variable $\mathrm{L} \in \overline{\mathscr{X}}^+$ satisfying

$$\mathbb{P}_0(\mathrm{L} = \infty) = 0 \text{ and } \mathbb{P}_1 = \mathrm{L}\mathbb{P}_0 + \mathbb{1}_{\{\mathrm{L}=\infty\}}\mathbb{P}_1 \quad \text{for all } B \in \mathscr{X} \tag{04.01}$$

is called a *likelihood ratio (LR)* of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$, symbolically $\mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0 := \mathrm{L}$. □

Here and subsequently, let $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathscr{X})$ and $\mathrm{L} := \mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0$ be a likelihood ratio of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$. We first note that $\mathbb{P}_0(\mathrm{L}) = \mathbb{P}_1(\mathrm{L} < \infty) \in [0, 1]$ and $\mathbb{P}_0(\mathrm{L} \in \mathbb{R}^+) = 1$ by definition, and also $\mathbb{P}_1(\mathrm{L} = 0) = \mathrm{L}\mathbb{P}_0(\mathrm{L} = 0) + \mathbb{P}_1(\{\mathrm{L} = 0\} \cap \{\mathrm{L} = \infty\}) = 0$.

§04.04 **Property**.

(i) $\mathbb{P}_0 \perp \mathbb{P}_1 \Leftrightarrow \exists B \in \mathscr{X} : \mathbb{P}_0(B) = 0$ *(hence* $\mathrm{L}\mathbb{P}_0(B) = 0$*) and* $\mathbb{P}_1(B) = 1$ *(hence* $\mathbb{P}_1(B \cap \{\mathrm{L} = \infty\}) = 1$*)* $\Leftrightarrow \mathbb{P}_1(\mathrm{L} = \infty) = 1 \Leftrightarrow \mathbb{P}_0(\mathrm{L}) = 0$*;*

(ii) $\mathbb{P}_0 \not\perp \mathbb{P}_1 \Leftrightarrow \forall B \in \mathscr{X} : \mathbb{P}_0(B) = 0$ *implies* $\mathbb{P}_1(B) < 1$ *(particularly for* $B = \{\mathrm{L} = \infty\}$*)* $\Leftrightarrow \mathbb{P}_1(\mathrm{L} = \infty) < 1 \Leftrightarrow \mathbb{P}_0(\mathrm{L}) > 0$*;*

(iii) $\mathbb{P}_1 \ll \mathbb{P}_0 \Leftrightarrow \forall B \in \mathscr{X} : \mathbb{P}_0(B) = 0$ *implies* $\mathbb{P}_1(B) = 0$ *(particularly for* $B = \{\mathrm{L} = \infty\}$*)* $\Leftrightarrow \mathbb{P}_1(\mathrm{L} = \infty) = 0 \Leftrightarrow \mathbb{P}_0(\mathrm{L}) = 1$.

§04.05 **Remark**. Note that both $\mathbb{P}_0$ and $\mathbb{P}_1$ are dominated by $\mathbb{P}_\mu := \frac{1}{2}(\mathbb{P}_0 + \mathbb{P}_1) \in \mathcal{W}(\mathscr{X})$. Let $\mathfrak{f}_i \in \mathscr{X}^+$ denote a $\mathbb{P}_\mu$-density of $\mathbb{P}_i$, $i \in \{0, 1\}$ (c.f. Reminder §04.02), then

$$\mathrm{L}_* = \frac{\mathfrak{f}_1}{\mathfrak{f}_0}\mathbb{1}_{\{\mathfrak{f}_0\in\mathbb{R}^+_{\setminus 0}\}} + \infty\mathbb{1}_{\{\mathfrak{f}_0=0\}\cap\{\mathfrak{f}_1\in\mathbb{R}^+_{\setminus 0}\}} \tag{04.02}$$

is a likelihood ratio of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$, i.e., $\mathrm{L}_* = \mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0$. Indeed, $\mathrm{L}_* \in \overline{\mathscr{X}}^+$ satisfies $\mathbb{P}_0(\mathrm{L}_* = \infty) \leqslant \mathbb{P}_0(\mathfrak{f}_0 = 0) = 0$ and for all $B \in \mathscr{X}$

$$\begin{aligned}\mathrm{L}_*\mathbb{P}_0(B) + \mathbb{P}_1(B \cap \{\mathrm{L}_* = \infty\}) &= \mathfrak{f}_0\mathbb{P}_\mu\big(\tfrac{\mathfrak{f}_1}{\mathfrak{f}_0}\mathbb{1}_{B\cap\{\mathfrak{f}_0\in\mathbb{R}^+_{\setminus 0}\}}\big) + \mathbb{P}_1(B \cap \{\mathfrak{f}_0 = 0\} \cap \{\mathfrak{f}_1 \in \mathbb{R}^+_{\setminus 0}\}) \\ &= \mathfrak{f}_1\mathbb{P}_\mu(B \cap \{\mathfrak{f}_0 \in \mathbb{R}^+_{\setminus 0}\}) + \mathbb{P}_1(B \cap \{\mathfrak{f}_0 = 0\}) = \mathbb{P}_1(B).\end{aligned}$$

Consequently, L_* is always a version of the likelihood ratio $\mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0$. In general the likelihood ratio $\mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0$ (and similar $\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P}_1$) is uniquely determined by (04.01) up to $(\mathbb{P}_0 + \mathbb{P}_1)$-a.e. equivalence (e.g. WTheorie 1, Lemma §03.15 or Witting [1985] Satz 1.110 a), p. 112). Moreover, the positive numerical random variable $\mathrm{L}_*^{-1} = \frac{\mathfrak{f}_0}{\mathfrak{f}_1}\mathbb{1}_{\{\mathfrak{f}_1\in\mathbb{R}^+_{\setminus 0}\}} + \infty\mathbb{1}_{\{\mathfrak{f}_1=0\}\cap\{\mathfrak{f}_0\in\mathbb{R}^+_{\setminus 0}\}}$ is a version of the likelihood ratio

$\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P}_1$ switching the roles of $\mathbb{P}_0$ and $\mathbb{P}_1$. Consequently, (iii) can equivalently be written as $\mathbb{P}_1 \ll \mathbb{P}_0 \Leftrightarrow \mathbb{P}_1(\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P}_1 = 0) = \mathbb{P}_1(\mathrm{L}_*^{-1} = 0) = \mathbb{P}_1(\mathrm{L}_* = \infty) = 0$. However, given any version $\mathrm{L} = \mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_0$ of the likelihood ratio the measure $\mathbb{P}_1$ can be written as a sum $\mathbb{P}_1 = \mathbb{P}_1^a + \mathbb{P}_1^\perp$ of two measures $\mathbb{P}_1^a, \mathbb{P}_1^\perp \in \mathcal{M}_\sigma(\mathscr{X})$ where $\mathbb{P}_1^a := \mathrm{L}\mathbb{P}_0$ and $\mathbb{P}_1^\perp := \mathbb{1}_{\{\mathrm{L}=\infty\}}\mathbb{P}_1$ with $\mathbb{P}_1^\perp(B) = \mathbb{P}_1(B \cap \{\mathrm{L} = \infty\})$, $B \in \mathscr{X}$ is, respectively, the absolute continuous and singular part of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$ (Lebesgue decomposition). □

§04.06 **Property**. *The two measures* $\mathbb{P}_1^a := \mathrm{L}\mathbb{P}_0$ *and* $\mathbb{P}_1^\perp := \mathbb{1}_{\{\mathrm{L}=\infty\}}\mathbb{P}_1$ *in* $\mathcal{M}_\sigma(\mathscr{X})$ *satisfy*

(i) $\mathbb{P}_1 = \mathbb{P}_1^a + \mathbb{P}_1^\perp$, $\mathbb{P}_1^a \ll \mathbb{P}_0$, *and* $\mathbb{P}_1^\perp \perp \mathbb{P}_0$;

(ii) $\mathbb{P}_1(f) \geqslant \mathbb{P}_1^a(f) = \mathrm{L}\mathbb{P}_0(f) = \mathbb{P}_0(\mathrm{L}f) = \mathbb{P}_1(f\mathbb{1}_{\{\mathrm{L}<\infty\}})$ *for all* $f \in \overline{\mathscr{X}}^+$;

(iii) $\mathbb{P}_0 \ll \mathbb{P}_1$ *if and only if* $\mathbb{P}_0(\mathrm{L}) = 1$ *if and only if* $\mathbb{P}_1(\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P}_1 = 0) = \mathbb{P}_1(\mathrm{L} = \infty) = 0$ *if and only if for all* $f \in \overline{\mathscr{X}}^+$ *holds* $\mathbb{P}_1(f) = \mathbb{P}_0(\mathrm{L}f)$. □

§04.07 **Reminder**. Consider a $\mathbb{R}^k$-valued statistic S defined on $(\mathcal{X}, \mathscr{X})$, i.e. $S \in \mathscr{X}^k$. If $\mathbb{P}_1 \ll \mathbb{P}_0$, then the probability measure $\mathbb{P}_1^S = \mathbb{P}_1 \circ S^{-1} \in \mathcal{W}(\mathscr{B}^k)$ induced by S under $\mathbb{P}_1$ can be calculated from the probability measure $\mathbb{P}_0^{(S,\mathrm{L})} = \mathbb{P}_0 \circ (S, \mathrm{L})^{-1}$ induced by the random vector (S, L) under $\mathbb{P}_0$ through the formula

$$\mathbb{P}_1(S \in B) = \mathbb{P}_1^S(\mathbb{1}_B) = \mathbb{P}_0\big(\mathbb{1}_B(S)\mathrm{L}\big) = \mathbb{P}_0^{(S,\mathrm{L})}\big(\mathbb{1}_B(\Pi_S)\Pi_\mathrm{L}\big) \quad \text{for all } B \in \mathscr{B}^k$$

using the coordinate maps $\Pi_\mathrm{L}(S, \mathrm{L}) = \mathrm{L}$ and $\Pi_S(S, \mathrm{L}) = S$. The formula, however, is only valid under the assumption $\mathbb{P}_1 \ll \mathbb{P}_0$, since a part of $\mathbb{P}_1$ orthogonal to $\mathbb{P}_0$ can't be recovered. □

Here and subsequently, let $\mathbb{P}_\Theta = (\mathbb{P}_\theta)_{\theta\in\Theta}$ with $\Theta \subseteq \mathbb{R}^k$ be a family of probability measures on a measurable space $(\mathcal{X}, \mathscr{X})$, and for each $\theta_o, \theta \in \Theta$ let $\mathrm{L}_{\theta_o}(\theta) := \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_{\theta_o}$ denote a likelihood ratio of $\mathbb{P}_\theta$ with respect to $\mathbb{P}_{\theta_o}$. Keep in mind, that $\mathrm{L}_{\theta_o}(\theta_o) = 1 (= \mathbb{1}_\mathcal{X})$.

§04.08 **Definition**. Let $s \geqslant 1$ and $\theta_o \in int(\Theta)$. The statistical model $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ (and the family $\mathbb{P}_\Theta$) is called *$\mathcal{L}_s(\theta_o)$-differentiable with derivative $\dot\ell_{\theta_o}$*, if $\dot\ell_{\theta_o} \in \mathcal{L}_s^k(\mathbb{P}_{\theta_o})$ and for all $\theta \to \theta_o$ hold

$$\|s(\mathrm{L}_{\theta_o}^{1/s}(\theta) - 1) - \langle\dot\ell_{\theta_o}, (\theta - \theta_o)\rangle\|_{\mathcal{L}_s(\mathbb{P}_{\theta_o})} = \mathrm{o}(\|\theta - \theta_o\|) \tag{04.03}$$

and $\mathbb{P}_\theta(\mathrm{L}_{\theta_o}(\theta) = \infty) = \mathrm{o}(\|\theta - \theta_o\|^s)$. □

§04.09 **Remark**. In case $s = 1$ the defining condition $\mathbb{P}_\theta(\mathrm{L}_{\theta_o}(\theta) = \infty) = \mathrm{o}(\|\theta - \theta_o\|)$ follows from (04.03) (Witting [1985], Hilfssatz 1.178, p164). We note that $\mathcal{L}_1(\theta_o)$-differentiability implies $\mathcal{L}_1(\mathbb{P}_{\theta_o})$-continuity of $\theta \mapsto \mathrm{L}_{\theta_o}(\theta)$ in θ_o, i.e., $\|\mathrm{L}_{\theta_o}(\theta) - \mathrm{L}_{\theta_o}(\theta_o)\|_{\mathcal{L}_1(\mathbb{P}_{\theta_o})} = \mathrm{o}(1)$ as $\theta \to \theta_o$. Since $\mathrm{L}_{\theta_o}(\theta)$ is unique up to $\mathbb{P}_\theta + \mathbb{P}_{\theta_o}$-a.e.-equivalence $\mathcal{L}_s(\theta_o)$-differentiability does not depend on the version $\mathrm{L}_{\theta_o}(\theta)$ of the likelihood ratio $\mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_{\theta_o}$. □

§04.10 **Lemma**. *If* $\mathbb{P}_\Theta$ *is* $\mathcal{L}_1(\theta_o)$*-differentiable with derivative* $\dot\ell_{\theta_o}$*, then it holds* $\mathbb{P}_{\theta_o}(\dot\ell_{\theta_o}) = 0$. *For any* $s \geqslant r \geqslant 1$ *if* $\mathbb{P}_\Theta$ *is* $\mathcal{L}_s(\theta_o)$*-differentiable with derivative* $\dot\ell_{\theta_o}$*, then* $\mathbb{P}_\Theta$ *is also* $\mathcal{L}_r(\theta_o)$*-differentiable with derivative* $\dot\ell_{\theta_o}$.

§04.11 **Proof** of Lemma §04.10. is given in the lecture. □

In order to avoid additional integrability conditions in Definition §04.08 the function $\theta \mapsto s(\mathrm{L}_\mu^{1/s}(\theta) - 1)$ is considered. The next assertion formulates differentiability under additional integrability conditions.

§04.12 **Lemma**. *Let $s \geqslant 1$ and $\theta_o \in \mathrm{int}(\Theta)$. The family $\mathbb{P}_\Theta$ is $\mathcal{L}_s(\theta_o)$-differentiable with derivative $\dot\ell_{\theta_o}$, if $\dot\ell_{\theta_o} \in \mathcal{L}_s^k(\mathbb{P}_{\theta_o})$, $\mathrm{L}_{\theta_o}(\theta) \in \mathcal{L}_s(\mathbb{P}_{\theta_o})$ for all $\theta \in U(\theta_o)$ and for all $\theta \to \theta_o$ hold*

$$\|(\mathrm{L}_{\theta_o}(\theta) - 1) - \langle \dot\ell_{\theta_o}, \theta - \theta_o\rangle\|_{\mathcal{L}_s(\mathbb{P}_{\theta_o})} = \mathrm{o}(\|\theta - \theta_o\|)$$

and $\mathbb{P}_\theta(\mathrm{L}_{\theta_o}(\theta) = \infty) = \mathrm{o}(\|\theta - \theta_o\|^s)$.

§04.13 **Proof** of Lemma §04.12. is given in the lecture. □

Let us assume in addition, that the family $\mathbb{P}_\Theta$ is dominated by $\mu \in \mathcal{M}_\sigma(\mathscr{X})$. For each $\theta \in \Theta$ denote by $\mathrm{L}_\mu(\theta) := \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mu \in \mathscr{X}^+$ a Radon-Nikodym density of $\mathbb{P}_\theta$ with respect to μ. Keeping Remark §04.05 in mind $\mathrm{L}_{*,\theta_o}(\theta) = \frac{\mathrm{L}_\mu(\theta)}{\mathrm{L}_\mu(\theta_o)}\mathbb{1}_{\{\mathrm{L}_\mu(\theta_o)\in\mathbb{R}^+_{\setminus 0}\}} + \infty\mathbb{1}_{\{\mathrm{L}_\mu(\theta_o)=0\}\cap\{\mathrm{L}_\mu(\theta)\in\mathbb{R}^+_{\setminus 0}\}}$ as in (04.02) is for each $\theta_o, \theta \in \Theta$ a version of the likelihood ratio $\mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_{\theta_o}$. We note that

$$\big\{\mathrm{L}_{*,\theta_o}(\theta) = \infty\big\} = \big\{\{\mathrm{L}_\mu(\theta_o) = 0\} \cap \{\mathrm{L}_\mu(\theta) \in \mathbb{R}^+_{\setminus 0}\}\big\} \subseteq \big\{\mathrm{L}_\mu(\theta_o) = 0\big\} =: \mathcal{N}_{\theta_o},$$

where $\mathbb{P}_{\theta_o}(\mathcal{N}_{\theta_o}) = 0$, and for all $\theta \in \Theta$ holds $\frac{\mathrm{L}_\mu(\theta)}{\mathrm{L}_\mu(\theta_o)}\mathbb{1}_{\mathcal{N}^c_{\theta_o}} = \mathrm{L}_{*,\theta_o}(\theta)\mathbb{1}_{\mathcal{N}^c_{\theta_o}} < \infty$ and $\mathbb{P}_\theta(\mathcal{N}_{\theta_o}) = \mathbb{P}_\theta(\mathcal{N}_{\theta_o} \cap \{\mathrm{L}_\mu(\theta) \in \mathbb{R}^+_{\setminus 0}\}) = \mathbb{P}_\theta(\mathrm{L}_{*,\theta_o}(\theta) = \infty) = \mathbb{P}_\theta(\mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_{\theta_o} = \infty)$. Decomposing the integral with respect to $\mathcal{X} = \mathcal{N}_{\theta_o} \biguplus \mathcal{N}^c_{\theta_o}$ it follows

$$\begin{aligned}
&\|2(\mathrm{L}_\mu^{1/2}(\theta) - \mathrm{L}_\mu^{1/2}(\theta_o)) - \langle \dot\ell_{\theta_o}, (\theta - \theta_o)\rangle \mathrm{L}_\mu^{1/2}(\theta_o)\|^2_{\mathcal{L}_2(\mu)} \\
&\quad = \|2(\mathrm{L}^{1/2}_{*,\theta_o}(\theta) - 1) - \langle \dot\ell_{\theta_o}, (\theta - \theta_o)\rangle\|^2_{\mathcal{L}_2(\mathbb{P}_{\theta_o})} + \|\mathbb{1}_{\mathcal{N}_{\theta_o}} 2\mathrm{L}_\mu^{1/2}(\theta)\|^2_{\mathcal{L}_2(\mu)} \\
&\quad = \|2(\mathrm{L}^{1/2}_{*,\theta_o}(\theta) - 1) - \langle \dot\ell_{\theta_o}, (\theta - \theta_o)\rangle\|^2_{\mathcal{L}_2(\mathbb{P}_{\theta_o})} + 4\mathbb{P}_\theta(\mathrm{L}_{*,\theta_o}(\theta) = \infty) \\
&\quad = \|2(\mathrm{L}^{1/2}_{\theta_o}(\theta) - 1) - \langle \dot\ell_{\theta_o}, (\theta - \theta_o)\rangle\|^2_{\mathcal{L}_2(\mathbb{P}_{\theta_o})} + 4\mathbb{P}_\theta(\mathrm{L}_{\theta_o}(\theta) = \infty). \qquad (04.04)
\end{aligned}$$

Keeping Remark §03.08 in mind for $\theta_o \in \mathrm{int}(\Theta)$ the family $\mathbb{P}_\Theta$ is *Hellinger-differentiable with derivative* $\dot\ell_{\theta_o}$, if $\dot\ell_{\theta_o} \in \mathcal{L}_2^k(\mathbb{P}_{\theta_o})$, hence $\dot\ell_{\theta_o}\mathrm{L}_\mu^{1/2}(\theta_o) \in \mathcal{L}_2^k(\mu)$, and for $\theta \to \theta_o$

$$\|\mathrm{L}_\mu^{1/2}(\theta) - \mathrm{L}_\mu^{1/2}(\theta_o) - \tfrac{1}{2}\langle \dot\ell_{\theta_o}, \theta - \theta_o\rangle \mathrm{L}_\mu^{1/2}(\theta_o)\|_{\mathcal{L}_2(\mu)} = \mathrm{o}(\|\theta - \theta_o\|).$$

Exploiting the identity (04.04) we obtain immediately the next property.

§04.14 **Property**. *Let $\mathbb{P}_\Theta \ll \mu \in \mathcal{M}_\sigma(\mathscr{X})$ and $\theta_o \in \mathrm{int}(\Theta)$. The family $\mathbb{P}_\Theta$ is Hellinger-differentiable with derivative $\dot\ell_{\theta_o}$ if and only if $\mathbb{P}_\Theta$ is $\mathcal{L}_2(\theta_o)$-differentiable with derivative $\dot\ell_{\theta_o}$.*

§04.15 **Proposition**. *Let $\mathbb{P}_\Theta \ll \mu \in \mathcal{M}_\sigma(\mathscr{X})$ with open $\Theta \subseteq \mathbb{R}^k$. If the likelihood $\mathrm{L}_\mu(\theta) := \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mu$, $\theta \in \Theta$, satisfies in addition the following conditions:*

(i) *for each $x \in \mathcal{X}$ the map $\theta \mapsto s(\theta, x) := \mathrm{L}_\mu^{1/2}(\theta, x)$ is continuously differentiable with derivative $\dot s_\theta := \frac{\partial}{\partial\theta} s$,*

(ii) *$\dot s_\theta \in \mathcal{L}_2(\mu)$ for all $\theta \in \Theta$, and hence $\mathcal{I}_\theta := 4\mu(\dot s_\theta \dot s_\theta^t) \in \mathbb{R}^{(k,k)}_\geqslant$,*

(iii) *the map $\theta \mapsto \mathcal{I}_\theta$ is continuous.*

Then $\mathbb{P}_\Theta$ is for all $\theta_0 \in \Theta$ Hellinger-differentiable with score function $\dot\ell_{\theta_o} = 2\frac{\dot s_{\theta_o}}{s(\theta_o)}\mathbb{1}_{\{s(\theta_o)\in\mathbb{R}^+_{\setminus 0}\}}$.

§04.16 **Proof** of Proposition §04.15. is given in the lecture. □

§04.17 **Example**. Consider a statistical location model $(\mathbb{R}, \mathscr{B}, \mathbb{P}_\mathbb{R})$ dominated by the Lebesgue measure $\lambda \in \mathcal{M}_\sigma(\mathscr{B})$ with likelihood for each $\theta \in \mathbb{R}$ given by $\mathrm{L}(\theta, x) = g(x - \theta)$, $x \in \mathbb{R}$, where g is a strictly positive density. If g is continuously differentiable with derivative $\dot g$ satisfying $\lambda(|\dot g|^2/g) < \infty$ then due to Proposition §04.15 the family $\mathbb{P}_\mathbb{R}$ is Hellinger-differentiable

with score function $\dot\ell_\theta = -\dot g(x-\theta)/g(x-\theta)$. Indeed, setting $s(\theta,x) := \sqrt{g(x-\theta)}$, we have $\dot s_\theta(x) = \frac{\partial}{\partial\theta}\sqrt{g(x-\theta)} = -\frac12 \dot g(x-\theta)/\sqrt{g(x-\theta)}$ which is continuous in θ and hence condition (i) is satisfied. Moreover conditions (ii) and (iii) hold true, since $\theta \mapsto \mathcal{I}_\theta = 4\lambda(\dot s_\theta)^2 = \lambda(|\dot g|^2/g) < \infty$ is constant and thus continuous. Applying Proposition §04.15 the family $\mathbb{P}_\mathbb{R}$ is Hellinger-differentiable with score function $\dot\ell_{\theta_o} = 2\frac{\dot s_{\theta_o}}{s(\theta_o)}\mathbb{1}_{\{s(\theta_o)\in\mathbb{R}^+_{\setminus 0}\}} = -\dot g(x-\theta_o)/g(x-\theta_o)$. □

§04|02 Contiguity

We introduce next an asymptotic version of absolute continuity. In this section we restrict our attention to probability measures $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, $n\in\mathbb{N}$, in short $(\mathbb{P}_0^n)_{n\in\mathbb{N}}, (\mathbb{P}_1^n)_{n\in\mathbb{N}} \in (\mathcal{W}(\mathscr{X}_n))_{n\in\mathbb{N}}$. We aim to obtain the limiting distribution of (test) statistics $S_n \in \mathscr{X}_n^k$, $n\in\mathbb{N}$, under $\mathbb{P}_1^n$ if its limiting distribution under $\mathbb{P}_0^n$ is known.

§04.18 **Definition**. Let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, $n\in\mathbb{N}$. The sequence $(\mathbb{P}_1^n)_{n\in\mathbb{N}}$ is called *contiguous* with respect to $(\mathbb{P}_0^n)_{n\in\mathbb{N}}$, symbolically $\mathbb{P}_1^n \lhd \mathbb{P}_0^n$, if for any $(B_n)_{n\in\mathbb{N}} \in (\mathscr{X}_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\mathbb{P}_0^n(B_n)=0$ holds $\lim_{n\to\infty}\mathbb{P}_1^n(B_n)=0$. The sequences $(\mathbb{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathbb{P}_0^n)_{n\in\mathbb{N}}$ are called *mutually contiguous*, symbolically $\mathbb{P}_0^n \lhd\rhd \mathbb{P}_1^n$, if both $\mathbb{P}_1^n \lhd \mathbb{P}_0^n$ and $\mathbb{P}_0^n \lhd \mathbb{P}_1^n$. □

§04.19 **Lemma**. *Let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, $n\in\mathbb{N}$.*

(i) $\mathbb{P}_1^n \lhd \mathbb{P}_0^n \Leftrightarrow$ *for all* $(S_n)_{n\in\mathbb{N}} \in (\mathscr{X}_n^k)_{n\in\mathbb{N}}$ *holds:* $S_n \xrightarrow{\mathbb{P}_0^n} 0 \Rightarrow S_n \xrightarrow{\mathbb{P}_1^n} 0$;

(ii) *For any statistic* $S_n : (\mathcal{X}_n, \mathscr{X}_n) \to (\mathcal{S}, \mathscr{S})$, $n\in\mathbb{N}$, *holds:* $\mathbb{P}_1^n \lhd \mathbb{P}_0^n \Rightarrow \mathbb{P}_1^n \circ S_n^{-1} \lhd \mathbb{P}_0^n \circ S_n^{-1}$;

(iii) *For any sub-sequence* $(n_k)_{k\in\mathbb{N}}$ *in* $\mathbb{N}$ *holds:* $\mathbb{P}_1^n \lhd \mathbb{P}_0^n \Rightarrow \mathbb{P}_1^{n_k} \lhd \mathbb{P}_0^{n_k}$;

(iv) $\mathbb{P}_1^n \lhd \mathbb{P}_0^n \Leftrightarrow$ *for any* $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ *exists* $\delta \in \mathbb{R}^+_{\setminus 0}$ *such that for all* $(B_n)_{n\in\mathbb{N}} \in (\mathscr{X}_n)_{n\in\mathbb{N}}$ *holds:* $\limsup_{n\to\infty}\mathbb{P}_0^n(B_n) < \delta \Rightarrow \limsup_{n\to\infty}\mathbb{P}_1^n(B_n) < \varepsilon$;

(v) *Let* $(S_n)_{n\in\mathbb{N}} \in (\mathscr{X}_n^k)_{n\in\mathbb{N}}$ *and* $\mathbb{P}_1^n \lhd \mathbb{P}_0^n$, *then:*

(v-a) $\mathbb{P}_0^n \circ S_n^{-1} \xrightarrow{d} \mathbb{P}_0$ *and* $\mathbb{P}_1^n \circ S_n^{-1} \xrightarrow{d} \mathbb{P}_1 \Rightarrow \mathbb{P}_1 \ll \mathbb{P}_0$;

(v-b) $(\mathbb{P}_0^n \circ S_n^{-1})_{n\in\mathbb{N}}$ *tight* $\Rightarrow (\mathbb{P}_1^n \circ S_n^{-1})_{n\in\mathbb{N}}$ *tight*.

§04.20 **Proof** of Lemma §04.19. is given in the lecture. □

§04.21 **Remark**. Next we characterise contiguity in terms of the asymptotic behaviour of the likelihood ratio $\mathrm{L}_n = \mathrm{d}\mathbb{P}_1^n/\mathrm{d}\mathbb{P}_0^n \in \overline{\mathscr{X}}_n^+$, $n\in\mathbb{N}$. First recall that $\mathbb{P}_1^n(\mathrm{L}_n < \infty) = \mathbb{P}_0^n(\mathrm{L}_n) \in [0,1]$ and $\mathbb{P}_0^n(\mathrm{L}_n = \infty) = \mathbb{P}_1^n(\mathrm{L}_n = 0) = 0$ for each $n\in\mathbb{N}$. Consequently, the probability measure $\mathbb{P}_0^n \circ \mathrm{L}_n^{-1} \in \mathcal{W}(\overline{\mathscr{B}})$ is concentrated in $\mathbb{R}^+$ meaning that $\mathbb{P}_0^n \circ \mathrm{L}_n^{-1}(\mathbb{R}^+) = \mathbb{P}_0^n(\mathrm{L}_n \in \mathbb{R}^+) = 1$ for each $n\in\mathbb{N}$. Moreover, $(\mathbb{P}_0^n \circ \mathrm{L}_n^{-1})_{n\in\mathbb{N}}$ is tight, since for any $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ and $c > 1/\varepsilon$ holds $\mathbb{P}_0^n(\mathrm{L}_n > c) \leqslant \frac1c \mathbb{P}_0^n(\mathrm{L}_n) \leqslant \frac1c < \varepsilon$ by Markov's inequality. However, $\mathbb{P}_1^n \circ \mathrm{L}_n^{-1}$ is generally not concentrated in $\mathbb{R}^+$, but under $\mathbb{P}_1^n \lhd \mathbb{P}_0^n$ holds $\mathbb{P}_1^n(\mathrm{L}_n = \infty) \to 0$ since $\mathbb{P}_0^n(\mathrm{L}_n = \infty) = 0$ for all $n\in\mathbb{N}$. Thereby, the limit distribution of $\mathbb{P}_1^n \circ \mathrm{L}_n^{-1}$ (if it exists) is concentrated in $\mathbb{R}^+$. □

Formally, we write $\mathrm{L}_n = \mathrm{L}_n \mathbb{1}_{\{\mathrm{L}_n\in\mathbb{R}^+\}} + \infty \mathbb{1}_{\{\mathrm{L}_n=\infty\}}$, where the second summand is negligible in the sense of Slutzky's lemma under contiguity $\mathbb{P}_1^n \lhd \mathbb{P}_0^n$.

§04.22 **Definition**. $(T_n)_{n\in\mathbb{N}} \in (\overline{\mathscr{X}}_n)_{n\in\mathbb{N}}$ converges in distribution to $\mathbb{P}^T \in \mathcal{W}(\mathscr{B})$ under $\mathbb{P}^n$, shortly $T_n \xrightarrow{d} \mathbb{P}^T$ *under* $\mathbb{P}^n$, if

$$\mathbb{P}^n \circ T_n^{-1} \xrightarrow{d} \mathbb{P}^T \quad :\Leftrightarrow \quad \mathbb{P}^n \circ (T_n \mathbb{1}_{\{T_n\in\mathbb{R}\}})^{-1} \xrightarrow{d} \mathbb{P}^T \quad \text{and} \quad \mathbb{P}^n(T_n \notin \mathbb{R}) \to 0. \tag{04.05}$$

We note that any family of probability measures on $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$ is tight, since $\overline{\mathbb{R}}$ is compact. A non trivial formulation of tightness for probability measures provides the next definition.

§04.23 **Definition**. A sequence $(\mathbb{P}^n)_{n\in\mathbb{N}} \in \mathcal{W}(\overline{\mathscr{B}})$ is called *asymptotically tight* if for all $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ exists $M \in \mathbb{R}^+_{\setminus 0}$ and $n_o \in \mathbb{N}$ such that for all $n > n_o$ holds $\mathbb{P}^n([-M, M]^c) < \varepsilon$. □

§04.24 **Remark**. Asymptotic tightness of $(\mathbb{P}^n)_{n\in\mathbb{N}} \in \mathcal{W}(\overline{\mathscr{B}})$ is equivalently characterised by: for any $(M_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ with $M_n \uparrow \infty$ holds $\mathbb{P}^n([-M_n, M_n]^c) \xrightarrow{n\to\infty} 0$. In particular, we have immediately $\mathbb{P}^n(\{-\infty, \infty\}) \xrightarrow{n\to\infty} 0$. The concept of asymptotic tightness and tightness as in Definition §20.21 coincide if $\mathbb{P}^n(\mathbb{R}) = 1$ for all $n \in \mathbb{N}$. Furthermore, it can be shown that the claim of Prohorov's theorem Property §20.24 holds also for families of asymptotically tight probability measures. □

§04.25 **Theorem**. *For each $n \in \mathbb{N}$ let $\mathbb{P}^n_0, \mathbb{P}^n_1 \in \mathcal{W}(\mathscr{X}_n)$, let $\mathrm{L}_n := \mathrm{d}\mathbb{P}^n_1/\mathrm{d}\mathbb{P}^n_0 \in \overline{\mathscr{X}_n}^+$ be a likelihood ratio of $\mathbb{P}^n_1$ with respect to $\mathbb{P}^n_0$ and let $\mathbb{P}^{\mathrm{L}}_0, \mathbb{P}^{\mathrm{L}}_1 \in \mathcal{W}(\mathscr{B})$. Then the following statements are equivalent:*

(a1) $\mathbb{P}^n_1 \lhd \mathbb{P}^n_0$;

(a2) $\mathbb{P}^n_0(\mathrm{L}_n) \xrightarrow{n\to\infty} 1$ *and for any $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ exists $M \in \mathbb{R}^+_{\setminus 0}$ with $\sup_{n\in\mathbb{N}} \mathbb{P}^n_0(\mathrm{L}_n \mathbb{1}_{\{\mathrm{L}_n > M\}}) < \varepsilon$, i.e. $(\mathbb{P}^n_0 \circ \mathrm{L}_n^{-1})_{n\in\mathbb{N}}$ is uniformly integrable;*

(a3) $(\mathbb{P}^n_1 \circ \mathrm{L}_n^{-1})_{n\in\mathbb{N}}$ *is asymptotically tight.*

If in addition $\mathrm{L}_n \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_0$ under $\mathbb{P}^n_0$, i.e. $\mathbb{P}^n_0 \circ \mathrm{L}_n^{-1} \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_0$, then the following statements are equivalent:

(b1) $\mathbb{P}^n_1 \lhd \mathbb{P}^n_0$;

(b2) $1 = \int_{\mathbb{R}} y\mathbb{P}^{\mathrm{L}}_0(dy) = \mathbb{P}^{\mathrm{L}}_0(\mathrm{id}_{\mathbb{R}}) = \mathbb{P}^{\mathrm{L}}_0(\mathrm{id}_{\mathbb{R}}\mathbb{1}_{\mathbb{R}})$;

(b3) $\mathrm{L}_n \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_1$ *under $\mathbb{P}^n_1$ with $\mathbb{P}^{\mathrm{L}}_1(B) = \mathbb{P}^{\mathrm{L}}_0(\mathrm{id}_{\mathbb{R}}\mathbb{1}_B) = \int_B y\mathbb{P}^{\mathrm{L}}_0(dy)$ for all $B \in \mathscr{B}$.*

§04.26 **Proof** of Theorem §04.25. is given in the lecture. □

Since $\mathbb{P}^n_0(\mathrm{L}_n) = \mathbb{P}^n_1(\mathrm{L}_n < \infty)$ it holds $\mathbb{P}^n_0(\mathrm{L}_n) \to 1 \Leftrightarrow \mathbb{P}^n_1(\mathrm{L}_n = \infty) \to 0$. Keeping (04.01) in mind the mass of the absolute continuous part of $\mathbb{P}^n_1$ with respect to $\mathbb{P}^n_0$ converges two 1, if and only if, the singular part vanishes.

§04.27 **Corollary**. *Under the notations of Theorem §04.25 the following statements are equivalent:*

(i) $\mathbb{P}^n_1 \lhd \mathbb{P}^n_0$;

(ii) *if $\mathbb{P}^{n_k}_0 \circ \mathrm{L}_{n_k}^{-1} \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_0 \in \mathcal{W}(\mathscr{B})$ along a sub-sequence $(n_k)_{k\in\mathbb{N}}$, then $\mathbb{P}^{\mathrm{L}}_0(\mathrm{id}_{\mathbb{R}}) = 1$;*

(iii) *if $\mathbb{P}^{n_k}_0 \circ \mathrm{L}_{n_k}^{-1} \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_0 \in \mathcal{W}(\mathscr{B})$ along a sub-sequence $(n_k)_{k\in\mathbb{N}}$, then $\mathbb{P}^{n_k}_1 \circ \mathrm{L}_{n_k}^{-1} \xrightarrow{d} \mathbb{P}^{\mathrm{L}}_1$, with $\mathbb{P}^{\mathrm{L}}_1(B) = \mathbb{P}^{\mathrm{L}}_0(\mathrm{id}_{\mathbb{R}}\mathbb{1}_B)$ for all $B \in \mathscr{B}$.*

§04.28 **Proof** of Corollary §04.27. is given in the lecture. □

We are particularly interested in mutual contiguity $(\mathbb{P}^n_0 \lhd\rhd \mathbb{P}^n_1)$ of $(\mathbb{P}^n_0)_{n\in\mathbb{N}}$ and $(\mathbb{P}^n_1)_{n\in\mathbb{N}}$, which can be characterised by applying Theorem §04.25 and its analogous formulation switching the roles of $\mathbb{P}^n_0$ and $\mathbb{P}^n_1$. However, for $n \in \mathbb{N}$ the transformation of a likelihood ratio $\mathrm{L}_n = \mathrm{d}\mathbb{P}^n_1/\mathrm{d}\mathbb{P}^n_0$ into a *log-likelihood ratio* (LLR) $\ell_n := \log \mathrm{L}_n = \log\left(\mathrm{d}\mathbb{P}^n_1/\mathrm{d}\mathbb{P}^n_0\right) \in \overline{\mathscr{X}}$ captures equally both orthogonal events $\{\mathrm{L}_n = 0\}$ and $\{\mathrm{L}_n = \infty\}$. Generally, ℓ_n takes the value $-\infty$ and $+\infty$ with positive $\mathbb{P}^n_0$- and $\mathbb{P}^n_1$-probability, respectively. In other words $\mathbb{P}^n_0 \circ \ell_n^{-1}$ and $\mathbb{P}^n_1 \circ \ell_n^{-1}$ is concentrated in $[-\infty, \infty)$ and $(-\infty, \infty]$, respectively, since by Definition §04.03 of L_n it holds

$$\mathbb{P}^n_0(\ell_n = \infty) = 0 \quad \text{and} \quad \mathbb{P}^n_1(\ell_n = -\infty) = 0 \quad \text{for all } n \in \mathbb{N}. \tag{04.06}$$

Thereby, similar to Remark §04.21 under mutual contiguity $\mathbb{P}^n_0 \lhd\rhd \mathbb{P}^n_1$ it follows

$$\mathbb{P}^n_1(\ell_n = \infty) \to 0 \quad \text{and} \quad \mathbb{P}^n_0(\ell_n = -\infty) \to 0 \quad \text{as } n \to \infty. \tag{04.07}$$

Consequently, the limit distribution of ℓ_n under both $\mathbb{P}_0^n$ and $\mathbb{P}_1^n$, if it exists, is concentrated in $\mathbb{R}$. Keeping Definition §04.22 in mind under mutual contiguity $\mathbb{P}_0^n \triangleleft\triangleright \mathbb{P}_1^n$ convergence in distribution of ℓ_n under $\mathbb{P}_0^n$ and $\mathbb{P}_1^n$ to $\mathbb{P}_0^\ell, \mathbb{P}_1^\ell \in \mathcal{W}(\mathscr{B})$, respectively, is equivalently characterised by

$$\begin{aligned} \mathbb{P}_0^n \circ \ell_n^{-1} \xrightarrow{d} \mathbb{P}_0^\ell \quad &\Leftrightarrow \quad \mathbb{P}_0^n \circ (\ell_n \mathbb{1}_{\{\ell_n > -\infty\}})^{-1} \xrightarrow{d} \mathbb{P}_0^\ell \quad \text{and} \\ \mathbb{P}_1^n \circ \ell_n^{-1} \xrightarrow{d} \mathbb{P}_1^\ell \quad &\Leftrightarrow \quad \mathbb{P}_1^n \circ (\ell_n \mathbb{1}_{\{\ell_n < \infty\}})^{-1} \xrightarrow{d} \mathbb{P}_1^\ell. \end{aligned} \tag{04.08}$$

If $\mathrm{L}_n^{-1} = \mathrm{d}\mathbb{P}_0^n/\mathrm{d}\mathbb{P}_1^n$ is a likelihood ratio of $\mathbb{P}_0^n$ with respect to $\mathbb{P}_1^n$, as for example in Remark §04.05, then making use of the identity $\log \mathrm{L}_n^{-1} = -\log \mathrm{L}_n = -\ell_n$ the convergence in distribution of ℓ_n under $\mathbb{P}_0^n$ respectively $\mathbb{P}_1^n$ implies immediately the corresponding convergence of $\log \mathrm{L}_n^{-1}$. Similar to Theorem §04.25 (b1)-(b3) the next result characterises mutual contiguity in terms of the log-likelihood ratio ℓ_n.

§04.29 **Theorem**. *For each $n \in \mathbb{N}$ let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, let $\ell_n := \log \mathrm{L}_n = \log(\mathrm{d}\mathbb{P}_1^n/\mathrm{d}\mathbb{P}_0^n) \in \overline{\mathscr{X}_n}$ be a log-likelihood ratio such that also $\mathrm{L}_n^{-1} = \mathrm{d}\mathbb{P}_0^n/\mathrm{d}\mathbb{P}_1^n \in \overline{\mathscr{X}_n}^+$ and let $\mathbb{P}_0^\ell, \mathbb{P}_1^\ell \in \mathcal{W}(\mathscr{B})$. If in addition $\ell_n \xrightarrow{d} \mathbb{P}_0^\ell$ under $\mathbb{P}_0^n$, i.e. $\mathbb{P}_0^n \circ \ell_n^{-1} \xrightarrow{d} \mathbb{P}_0^\ell$, then the following statements are equivalent:*

(b'1) $\mathbb{P}_1^n \triangleleft\triangleright \mathbb{P}_0^n$;

(b'2) $1 = \int_{\mathbb{R}} \exp(z)\mathbb{P}_0^\ell(dz) = \mathbb{P}_0^\ell(\exp) = \mathbb{P}_0^\ell(\exp \mathbb{1}_{\mathbb{R}})$

(b'3) $\ell_n \xrightarrow{d} \mathbb{P}_1^\ell$ *under* $\mathbb{P}_1^n$ *with* $\mathbb{P}_1^\ell(B) = \mathbb{P}_0^\ell(\exp \mathbb{1}_B) = \int_B \exp(z)\mathbb{P}_0^\ell(dz)$ *for all* $B \in \mathscr{B}$.

§04.30 **Proof** of Theorem §04.29. is given in the lecture. □

§04.31 **Remark**. Let $\mathfrak{f}_0^\ell$ and $\mathfrak{f}_1^\ell$ denote, respectively, a μ-density of $\mathbb{P}_0^\ell$ and $\mathbb{P}_1^\ell$ with respect to a measure $\mu \in \mathcal{M}_\sigma(\mathscr{B})$ dominating $\mathbb{P}_0^\ell$, and hence $\mathbb{P}_1^\ell$. The measure $\mathbb{P}_1^\ell$ in Theorem §04.29 (b'3) is equally defined by $\mathfrak{f}_1^\ell(z) = \exp(z)\mathfrak{f}_0^\ell(z)$ for μ-a.e. $z \in \mathbb{R}$. □

§04.32 **Corollary**. *Under the notations of Theorem §04.29 if $\mathbb{P}_0^n \circ \ell_n^{-1} \xrightarrow{d} \mathrm{N}_{(\mu,\sigma^2)}$ for $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ then the following statements are equivalent:*

(b"1) $\mathbb{P}_1^n \triangleleft\triangleright \mathbb{P}_0^n$;

(b"2) $\mu = -\sigma^2/2$

(b"3) $\ell_n \xrightarrow{d} \mathrm{N}_{(\sigma^2/2,\sigma^2)}$ *under* $\mathbb{P}_1^n$.

§04.33 **Proof** of Corollary §04.32. is given in the lecture. □

§04.34 **Example** *(Le Cam's first lemma)*. For $n \in \mathbb{N}$ let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$ and $\mathrm{L}_n := \mathrm{d}\mathbb{P}_1^n/\mathrm{d}\mathbb{P}_0^n \in \overline{\mathscr{X}_n}^+$. If $\ell_n := \log \mathrm{L}_n \xrightarrow{d} \mathrm{N}_{(-\sigma^2/2,\sigma^2)}$ under $\mathbb{P}_0^n$, then $\mathbb{P}_1^n \triangleleft\triangleright \mathbb{P}_0^n$ and $\ell_n \xrightarrow{d} \mathrm{N}_{(\sigma^2/2,\sigma^2)}$ under $\mathbb{P}_1^n$ due to Corollary §04.32. For $\sigma > 0$ from $\sigma^{-1}(\ell_n + \sigma^2/2) \xrightarrow{d} \mathrm{N}_{(0,1)}$ under $\mathbb{P}_0^n$ follows thus $\sigma^{-1}(\ell_n + \sigma^2/2) \xrightarrow{d} \mathrm{N}_{(\sigma,1)}$ under $\mathbb{P}_1^n$. In other words in this situation there is asymptotically a location shift by σ. □

For each $n \in \mathbb{N}$ let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, let $\mathrm{L}_n := \mathrm{d}\mathbb{P}_1^n/\mathrm{d}\mathbb{P}_0^n \in \overline{\mathscr{X}_n}^+$ be a likelihood ratio of $\mathbb{P}_1^n$ with respect to $\mathbb{P}_0^n$, let $\ell_n := \log \mathrm{L}_n$ and let $S_n \in \mathscr{X}_n^k$ be a $\mathbb{R}^k$-valued statistic defined on $(\mathfrak{X}_n, \mathscr{X}_n)$. We search conditions which allow to calculate the limiting distribution of (S_n, L_n) respectively (S_n, ℓ_n) under $\mathbb{P}_1^n$, from the limiting distribution of (S_n, L_n) respectively (S_n, ℓ_n) under $\mathbb{P}_0^n$. Keeping again (§04.22) in mind under mutual contiguity $\mathbb{P}_0^n \triangleleft\triangleright \mathbb{P}_1^n$ the joint convergence in distribution of $(S_n, \mathrm{L}_n) \xrightarrow{d} \mathbb{P}_1^{(S,\mathrm{L})} \in \mathcal{W}(\mathscr{B}^{k+1})$ under $\mathbb{P}_1^n$, $(S_n, \ell_n) \xrightarrow{d} \mathbb{P}_0^{(S,\ell)} \in \mathcal{W}(\mathscr{B}^{k+1})$

under $\mathbb{P}_0^n$ and $(S_n, \ell_n) \xrightarrow{d} \mathbb{P}_1^{(S,\ell)} \in \mathcal{W}(\mathscr{B}^{k+1})$ under $\mathbb{P}_1^n$, respectively, is equally characterised by

$$\begin{aligned}
&\mathbb{P}_1^n \circ (S_n, \mathrm{L}_n)^{-1} \xrightarrow{d} \mathbb{P}_1^{(S,\mathrm{L})} \quad \Leftrightarrow \quad \mathbb{P}_1^n \circ (S_n, \mathrm{L}_n \mathbb{1}_{\{\mathrm{L}_n < \infty\}})^{-1} \xrightarrow{d} \mathbb{P}_1^{(S,\mathrm{L})}, \\
&\mathbb{P}_0^n \circ (S_n, \ell_n)^{-1} \xrightarrow{d} \mathbb{P}_0^{(S,\ell)} \quad \Leftrightarrow \quad \mathbb{P}_0^n \circ (S_n, \ell_n \mathbb{1}_{\{\ell_n > -\infty\}})^{-1} \xrightarrow{d} \mathbb{P}_0^{(S,\ell)} \quad \text{and} \\
&\mathbb{P}_1^n \circ (S_n, \ell_n)^{-1} \xrightarrow{d} \mathbb{P}_1^{(S,\ell)} \quad \Leftrightarrow \quad \mathbb{P}_1^n \circ (S_n, \ell_n \mathbb{1}_{\{\ell_n < \infty\}})^{-1} \xrightarrow{d} \mathbb{P}_1^{(S,\ell)}. \qquad (04.09)
\end{aligned}$$

Denote by $\Pi_{\mathrm{L}} := \Pi_{k+1} \in \mathscr{B}^{k+1}$, i.e. $y = (y_i)_{i \in [\![k+1]\!]} \mapsto \Pi_{\mathrm{L}}(y) := y_{k+1}$ (respectively $\Pi_\ell := \Pi_{k+1} \in \mathscr{B}^{k+1}$) the coordinate map which allows us to write

$$\int_C y \mathbb{P}_1^{(S,\mathrm{L})}(ds, dy) = \int_{\mathbb{R}^{k+1}} \mathbb{1}_C(s,y) \Pi_{\mathrm{L}}(s,y) \mathbb{P}_1^{(S,\mathrm{L})}(ds, dy) = \mathbb{P}_1^{(S,\mathrm{L})}(\mathbb{1}_C \Pi_{\mathrm{L}}) \text{ for all } C \in \mathscr{B}^{k+1}.$$

§04.35 **Theorem.** *For each $n \in \mathbb{N}$ let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, let $\ell_n = \log \mathrm{L}_n = \log(\mathrm{d}\mathbb{P}_1^n / \mathrm{d}\mathbb{P}_0^n) \in \overline{\mathscr{X}_n}$ be a log-likelihood ratio, and let $S_n \in \mathscr{X}_n^k$ be a $\mathbb{R}^k$-valued statistic. Then, we have*

(i) *If $(S_n, \mathrm{L}_n) \xrightarrow{d} \mathbb{P}_0^{(S,\mathrm{L})} \in \mathcal{W}(\mathscr{B}^{k+1})$ under $\mathbb{P}_0^n$ and $\mathbb{P}_0^{(S,\mathrm{L})}(\Pi_{\mathrm{L}} \mathbb{1}_{\mathbb{R}^{k+1}}) = \mathbb{P}_0^{(S,\mathrm{L})}(\Pi_{\mathrm{L}}) = 1$, then $(S_n, \mathrm{L}_n) \xrightarrow{d} \mathbb{P}_1^{(S,\mathrm{L})}$ under $\mathbb{P}_1^n$ with $\mathbb{P}_1^{(S,\mathrm{L})}(C) := \mathbb{P}_0^{(S,\mathrm{L})}(\Pi_{\mathrm{L}} \mathbb{1}_C)$ for all $C \in \mathscr{B}^{k+1}$.*

(ii) *If $(S_n, \ell_n) \xrightarrow{d} \mathbb{P}_0^{(S,\ell)} \in \mathcal{W}(\mathscr{B}^{k+1})$ under $\mathbb{P}_0^n$ and $\mathbb{P}_0^{(S,\ell)}(\exp(\Pi_\ell) \mathbb{1}_{\mathbb{R}^{k+1}}) = \mathbb{P}_0^{(S,\ell)}(\exp(\Pi_\ell)) = 1$, then $(S_n, \ell_n) \xrightarrow{d} \mathbb{P}_1^{(S,\ell)}$ under $\mathbb{P}_1^n$ with $\mathbb{P}_1^{(S,\ell)}(C) := \mathbb{P}_0^{(S,\ell)}(\exp(\Pi_\ell) \mathbb{1}_C)$ for all $C \in \mathscr{B}^{k+1}$.*

§04.36 **Proof** of Theorem §04.35. is given in the lecture. □

§04.37 **Example** *(Le Cam's third lemma).* For each $n \in \mathbb{N}$ let $\mathbb{P}_0^n, \mathbb{P}_1^n \in \mathcal{W}(\mathscr{X}_n)$, let $\ell_n = \log \mathrm{L}_n = \log(\mathrm{d}\mathbb{P}_1^n / \mathrm{d}\mathbb{P}_0^n) \in \overline{\mathscr{X}_n}$ be a log-likelihood ratio, and let $S_n \in \mathscr{X}_n^k$ be a $\mathbb{R}^k$-valued statistic. Suppose that the limit distribution of (S_n, ℓ_n) under $\mathbb{P}_0^n$ is multivariate normal, that is

$$\mathbb{P}_0^n \circ (S_n, \ell_n)^{-1} \xrightarrow{d} \mathbb{P}_0^{(S,\ell)} = \mathrm{N}_{(v,M)} \quad \text{with} \quad v = \begin{pmatrix} \mu \\ -\frac{\sigma^2}{2} \end{pmatrix} \text{ and } M = \begin{pmatrix} \Sigma & \tau \\ \tau^t & \sigma^2 \end{pmatrix}. \qquad (04.10)$$

Then it holds $(S_n, \ell_n) \xrightarrow{d} \mathbb{P}_1^{(S,\ell)} = \mathrm{N}_{(v',M)}$ under $\mathbb{P}_1^n$ with $v' = (\mu + \tau, \sigma^2/2)^t$. Indeed, since $\mathbb{P}_0^{(S,\ell)}(\exp(\Pi_\ell)) = 1$ both assumptions of Theorem §04.35 (ii) are satisfied and hence it remains to calculate the limit distribution $\mathbb{P}_1^{(S,\ell)}(C) := \mathbb{P}_0^{(S,\ell)}(\exp(\Pi_\ell) \mathbb{1}_C)$ for all $C \in \mathscr{B}^{k+1}$. Suppose first $M > 0$, or equivalently $\Sigma > 0$ and $\sigma > 0$, then $\mathbb{P}_0^{(S,\ell)}$ has a density $\mathfrak{f}_0^{(S,\ell)}$ with respect to the Lebesgue-measure $\lambda^{k+1} \in \mathcal{M}_\sigma(\mathscr{B}^{k+1})$ and (see Remark §04.31) the Lebesgue-density $\mathfrak{f}_1^{(S,\ell)}$ of $\mathbb{P}_1^{(S,\ell)}$ satisfies $\mathfrak{f}_1^{(S,\ell)}(s,z) = \exp(z) \mathfrak{f}_0^{(S,\ell)}(s,z)$ for λ^{k+1}-a.e. $(s,z) \in \mathbb{R}^{k+1}$. Keeping the coordinate map Π_ℓ in mind we denote by $\mathfrak{f}_0^\ell$ and $\mathfrak{f}_1^\ell$ the marginal density of $\mathbb{P}_0^{(S,\ell)} \circ \Pi_\ell$ and $\mathbb{P}_1^{(S,\ell)} \circ \Pi_\ell$, respectively. Denoting by $\mathfrak{f}_0^{S|\ell=z}$ and $\mathfrak{f}_1^{S|\ell=z}$, respectively, a conditional density of S given $\ell = z$ under the joint distribution $\mathbb{P}_1^{(S,\ell)}$ and $\mathbb{P}_1^{(S,\ell)}$ (see Notation §21.11 (iv)) we have $\mathfrak{f}_1^{S|\ell=z}(s) \mathfrak{f}_1^\ell(z) = \exp(z) \mathfrak{f}_0^{S|\ell=z}(s) \mathfrak{f}_0^\ell(z)$ for λ^{k+1}-a.e. $(s,z) \in \mathbb{R}^{k+1}$. Exploiting Theorem §04.29 (b'3) it holds $\mathfrak{f}_1^\ell(z) = \exp(z) \mathfrak{f}_0^\ell(z)$ for λ-a.e. $z \in \mathbb{R}$ (see Remark §04.31). Consequently, it remains to verify that $\mathrm{N}_{(v,M)}$ and $\mathrm{N}_{(v',M)}$ have the same conditional distribution given $\ell = z$. Indeed, both are again multivariate normal (see Notation §21.11 (v)) with equal covariance matrix $\Sigma - \sigma^2 \tau \tau^t$ and conditional mean $\mathbb{P}_0^{S|\ell=z}(\mathrm{id}_{\mathbb{R}^k}) = \mu + \sigma^{-2}\tau(z + \sigma^2/2) = \mu + \tau + \sigma^{-2}\tau(z - \sigma^2/2) = \mathbb{P}_1^{S|\ell=z}(\mathrm{id}_{\mathbb{R}^k})$. The case of a positive semi-definite Σ and $\sigma^2 > 0$ follows by similar arguments when considering the projection onto the image of Σ. If $\sigma = 0$ the claim follows from Lemma §04.19 (i) together with Slutzky's lemma §20.10. In particular, note that $S_n \xrightarrow{d} \mathrm{N}_{(\mu,\Sigma)}$ under $\mathbb{P}_0^n$ and $S_n \xrightarrow{d} \mathrm{N}_{(\mu+\tau,\Sigma)}$ under $\mathbb{P}_1^n$ (see Reminder §04.07). □

§05 Local asymptotic normality (LAN)

§05.01 **Aim.** For each $n \in \mathbb{N}$ let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n = (\mathbb{P}_\theta^n)_{\theta\in\Theta})$ with $\Theta \subseteq \mathbb{R}^k$ be a statistical experiment. We aim to approximate $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)$ in a certain sense by a Gaussian location model after suitable reparametrisation.

§05.02 **Reminder.** Consider on $(\mathbb{R}^k, \mathscr{B}^k)$ the family $\mathrm{N}_{\mathbb{R}^k\times\{\Sigma\}} := (\mathrm{N}_{(h,\Sigma)})_{h\in\mathbb{R}^k}$ of multivariate normal distributions with common strictly positive definite covariance matrix $\Sigma \in \mathbb{R}_>^{(k,k)}$ and log-likelihood ratio $\log\big(\mathrm{dN}_{(h,\Sigma)}/\mathrm{dN}_{(0,\Sigma)}\big)(z) = \langle\Sigma^{-1}h, z\rangle - \frac{1}{2}\langle\Sigma^{-1}h, h\rangle$, $z \in \mathbb{R}^k$. Noting that for each $h \in \mathbb{R}^k$ the likelihood $\mathrm{L}(h) = \mathrm{dN}_{(h,\Sigma)}/\mathrm{d}\lambda^k$ of $\mathrm{N}_{(h,\Sigma)}$ with respect to the Lebesgue measure λ^k on $\mathbb{R}^k$ satisfies $\mathrm{L}(h, x) = \mathrm{L}(0, x - h)$ for all $x \in \mathbb{R}^k$ the statistical experiment $(\mathbb{R}^k, \mathscr{B}^k, \mathrm{N}_{\mathbb{R}^k\times\{\Sigma\}})$ is called a *Gaussian location model*. □

Consider a localised reparametrisation centred around a parameter value $\theta_o \in int(\Theta)$ which is in the sequel regarded as fixed.

§05.03 **Definition.** Consider a sequence of statistical experiments $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)$, $n \in \mathbb{N}$, with common parameter set $\Theta \subseteq \mathbb{R}^k$. Given a *localising rate* $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n = o(1)$ for each $n \in \mathbb{N}$ define a *local parameter set* $\Theta_o^n := \{\delta_n^{-1}(\theta - \theta_o) : \theta \in \Theta\} \subseteq \mathbb{R}^k$. For each $\theta \in \Theta$ and associated *local parameter* $h = \delta_n^{-1}(\theta - \theta_o) \in \Theta_o^n$ rewriting $\mathbb{P}_\theta^n$ as $\mathbb{P}_{\theta_o+\delta_n h}^n$ we obtain a sequence of *localised statistical experiment* $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_{\delta_n\Theta_o^n+\theta_o}^n := (\mathbb{P}_{\theta_o+\delta_n h}^n)_{h\in\Theta_o^n})$, $n \in \mathbb{N}$. □

§05.04 **Remark.** In the sequel we eventually take the local parameter set Θ_o^n equal to $\mathbb{R}^k$ which is not correct if the parameter set Θ is a strict subset of $\mathbb{R}^k$. However, if $\theta_o \in int(\Theta)$ is an inner point of Θ, which is assumed throughout this section, then for each $h \in \mathbb{R}^k$ the parameter $\theta = \theta_o + \delta_n h$ belongs to Θ for every sufficiently large n. In other words, the local parameter set Θ_o^n converges to the whole of $\mathbb{R}^k$ as $n \to \infty$, i.e., $\cup_{n\in\mathbb{N}}\Theta_o^n = \mathbb{R}^k$. Thereby, we tactically may either define the probability measure $\mathbb{P}_{\theta_o+\delta_n h}$ arbitrarily if $\theta_o + \delta_n h$ does not belong to Θ, or assume that n is sufficiently large. □

§05.05 **Aim.** We show, for large n, that the localised statistical experiment $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_{\delta_n\mathbb{R}^k+\theta_o}^n)$ and a Gaussian location model $(\mathbb{R}^k, \mathscr{B}^k, \mathrm{N}_{\mathbb{R}^k\times\{\mathcal{I}_{\theta_o}^{-1}\}})$ are similar in statistical properties whenever the original experiments, i.e., $\theta \mapsto \mathbb{P}_\theta$, are "smooth".

§05.06 **Heuristics.** Consider a statistical experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ dominated by $\mu \in \mathcal{M}_\sigma(\mathscr{X})$, i.e., $\mathbb{P}_\Theta \ll \mu$, with $\Theta \subseteq \mathbb{R}$, positive real likelihood $\mathrm{L}(\theta) = \mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mu \in \mathscr{X}^+$ and log-likelihood $\ell = \log \mathrm{L}$. Assume that for all $x \in \mathcal{X}$, the map $\theta \mapsto \ell(\theta, x)$ is twice differentiable with derivatives $\dot\ell_\theta := \frac{\partial}{\partial\theta}\ell$ and $\ddot\ell_\theta := \frac{\partial^2}{\partial^2\theta}\ell$. A *Taylor expansion* of the log-likelihood ratio leads to $\ell(\theta + h, x) - \ell(\theta, x) = h\dot\ell_\theta(x) + \frac{1}{2}h^2\ddot\ell_\theta(x) + \mathrm{o}_x(h^2)$ where the remainder term depends on x. Considering a product experiment $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ eventually it holds $\log(\mathrm{d}\mathbb{P}_{\theta+h/\sqrt{n}}^{\otimes n}/\mathrm{d}\mathbb{P}_\theta^{\otimes n}) = h\sqrt{n}\widehat{\mathbb{P}}_n\big(\dot\ell_\theta\big) + \frac{1}{2}h^2\widehat{\mathbb{P}}_n\big(\ddot\ell_\theta\big) + R_n$ where the score $\dot\ell$ has mean zero, i.e., $\mathbb{P}_\theta\big(\dot\ell_\theta\big) = 0$, and the Fisher information $\mathcal{I}_\theta$ equals $-\mathbb{P}_\theta\big(\ddot\ell_\theta\big) = \mathbb{P}_\theta\big(|\dot\ell_\theta|^2\big)$. Setting $\mathcal{Z}_\theta^n := \sqrt{n}\widehat{\mathbb{P}}_n\big(\dot\ell_\theta\big)$ from the central limit theorem §20.13 follows $\mathcal{Z}_\theta^n \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_\theta)}$ under $\mathbb{P}_\theta^{\otimes n}$ while due to the law of large numbers §20.06 it holds $\widehat{\mathbb{P}}_n\ddot\ell_\theta = -\mathcal{I}_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$. If in addition the remainder term is negligible, i.e., $R_n = \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$, then the log-likelihood ratio permits an expansion

$$\log(\mathrm{d}\mathbb{P}_{\theta+h/\sqrt{n}}^{\otimes n}/\mathrm{d}\mathbb{P}_\theta^{\otimes n}) = h\mathcal{Z}_\theta^n - \tfrac{1}{2}h^2\mathcal{I}_\theta + \mathrm{o}_{\mathbb{P}_\theta^{\otimes n}}(1)$$

which in the limit equals the log-likelihood ratio in a Gaussian location model. □

§05.07 **Definition.** A sequence of statistical experiments $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)_{n\in\mathbb{N}}$ with $\Theta \subseteq \mathbb{R}^k$ is called *local asymptotic normal (LAN)* in $\theta_o \in int(\Theta)$, if there is a localising rate $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n = o(1)$, a sequence of statistics $(\mathcal{Z}_{\theta_o}^n)_{n\in\mathbb{N}} \in (\mathscr{X}_n^k)_{n\in\mathbb{N}}$ and a matrix $\mathcal{I}_{\theta_o} \in \mathbb{R}^{(k,k)}$ such that for every $h \in \mathbb{R}^k$ the following three statements hold true:

(a) $\theta_o + \delta_n h \in \Theta$ for all sufficiently large n, i.e., $n \geqslant n_o(h)$;

(b) $\mathcal{Z}_{\theta_o}^n \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_{\theta_o})}$ under $\mathbb{P}_{\theta_o}^n$, i.e., $\mathbb{P}_{\theta_o}^n \circ (\mathcal{Z}_{\theta_o}^n)^{-1} \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_{\theta_o})}$;

(c) $\log(\mathrm{d}\mathbb{P}_{\theta_o+\delta_n h}^n / \mathrm{d}\mathbb{P}_{\theta_o}^n) = \langle \mathcal{Z}_{\theta_o}^n, h\rangle - \frac{1}{2}\langle \mathcal{I}_{\theta_o} h, h\rangle + R_{n,h}$ where $R_{n,h} = \mathrm{o}_{\mathbb{P}_{\theta_o}^n}(1)$.

The matrix $\mathcal{I}_{\theta_o}$ and the sequence of statistics $(\mathcal{Z}_{\theta_o}^n)_{n\in\mathbb{N}}$ is called, respectively, *Fisher information* at θ_o and *central sequence*. □

§05.08 **Comment.** If we assume in addition a strictly positive definite matrix $\mathcal{I}_{\theta_o} \in \mathbb{R}_>^{(k,k)}$ with inverse $\mathcal{I}_{\theta_o}^{-1}$ the sequence of statistics $(\widetilde{\mathcal{Z}}_{\theta_o}^n := \mathcal{I}_{\theta_o}^{-1}\mathcal{Z}_{\theta_o}^n)_{n\in\mathbb{N}} \in (\mathscr{X}_n^k)_{n\in\mathbb{N}}$ is equally a central sequence satisfying $\widetilde{\mathcal{Z}}_{\theta_o}^n \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_{\theta_o}^{-1})}$ under $\mathbb{P}_\theta^n$ and $\log(\mathrm{d}\mathbb{P}_{\theta_o+\delta_n h}^n / \mathrm{d}\mathbb{P}_{\theta_o}^n) = \langle \mathcal{I}_{\theta_o} h, \widetilde{\mathcal{Z}}_{\theta_o}^n\rangle - \frac{1}{2}\langle \mathcal{I}_{\theta_o} h, h\rangle + \mathrm{o}_{\mathbb{P}_{\theta_o}^n}(1)$. In other words the likelihood ratio $\mathrm{d}\mathbb{P}_{\theta_o+\delta_n h}^n / \mathrm{d}\mathbb{P}_{\theta_o}^n$ equals approximately the likelihood ratio $\mathrm{dN}_{(h,\mathcal{I}_{\theta_o}^{-1})} / \mathrm{dN}_{(0,\mathcal{I}_{\theta_o}^{-1})}$ as in the Reminder §05.02. Consequently, the localised statistical model $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_{\delta_n\Theta_o^n+\theta_o}^n)$ is similar to a Gaussian location model $(\mathbb{R}^k, \mathscr{B}^{\otimes k}, \mathrm{N}_{\mathbb{R}^k\times\{\mathcal{I}_{\theta_o}^{-1}\}})$ in the sense of Definition §05.07. □

§05.09 **Definition.** A LAN sequence of statistical experiments is called *uniformly local asymptotic normal (ULAN)* in $\theta_o \in \Theta$, if the condition (c) in Definition §05.07 is replaced by

(c') for $h_n \to h$ it holds $\log(\mathrm{d}\mathbb{P}_{\theta_o+\delta_n h_n}^n / \mathrm{d}\mathbb{P}_{\theta_o}^n) = \langle \mathcal{Z}_{\theta_o}^n, h\rangle - \frac{1}{2}\langle \mathcal{I}_{\theta_o} h, h\rangle + \mathrm{o}_{\mathbb{P}_{\theta_o}^n}(1)$. □

§05.10 **Theorem.** *Let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}_\Theta^n)_{n\in\mathbb{N}}$ be LAN in $\theta_o \in \Theta \subseteq \mathbb{R}^k$ with localising rate $(\delta_n)_{n\in\mathbb{N}}$, central sequence $(\mathcal{Z}_{\theta_o}^n)_{n\in\mathbb{N}}$ and Fisher information matrix $\mathcal{I}_{\theta_o} \in \mathbb{R}^{(k,k)}$. Then for any $h, h' \in \mathbb{R}^k$ the following statements hold true:*

(i) *$(\mathbb{P}_{\theta_o+\delta_n h}^n)_{n\in\mathbb{N}}$ and $(\mathbb{P}_{\theta_o+\delta_n h'}^n)_{n\in\mathbb{N}}$ are mutually contiguous, i.e., $\mathbb{P}_{\theta_o+\delta_n h}^n \lhd\rhd \mathbb{P}_{\theta_o+\delta_n h'}^n$;*

(ii) *$\mathcal{Z}_{\theta_o}^n \xrightarrow{d} \mathrm{N}_{(\mathcal{I}_{\theta_o} h, \mathcal{I}_{\theta_o})}$ under $\mathbb{P}_{\theta_o+\delta_n h}^n$.*

If the sequence of statistical experiments is ULAN, then for any $h_n \to h$ and $h'_n \to h'$ in $\mathbb{R}^k$ the following statements hold true:

(i') *$(\mathbb{P}_{\theta_o+\delta_n h_n}^n)_{n\in\mathbb{N}}$ and $(\mathbb{P}_{\theta_o+\delta_n h'_n}^n)_{n\in\mathbb{N}}$ are mutually contiguous, i.e., $\mathbb{P}_{\theta_o+\delta_n h_n}^n \lhd\rhd \mathbb{P}_{\theta_o+\delta_n h'_n}^n$;*

(ii') *$\mathcal{Z}_{\theta_o}^n \xrightarrow{d} \mathrm{N}_{(\mathcal{I}_{\theta_o} h, \mathcal{I}_{\theta_o})}$ under $\mathbb{P}_{\theta_o+\delta_n h_n}^n$.*

§05.11 **Proof** of Theorem §05.10. is given in the lecture. □

§05.12 **Theorem.** *Let $\mathbb{P}_\Theta \ll \mu \in \mathcal{M}_\sigma(\mathscr{X})$ with open $\Theta \subseteq \mathbb{R}^k$ be Hellinger-differentiable in $\theta_o \in \Theta$ with derivative $\dot{\ell}_{\theta_o}$ and Fisher information matrix $\mathcal{I}_{\theta_o} = \mathbb{P}_{\theta_o}(\dot{\ell}_{\theta_o}\dot{\ell}_{\theta_o}^t) \in \mathbb{R}_\geqslant^{(k,k)}$. Then the sequence of product experiments $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ is ULAN in θ_o with localising rate $\delta_n := n^{-1/2}$ and central sequence $\mathcal{Z}_{\theta_o}^n := \sqrt{n}\,\widehat{\mathbb{P}}_n(\dot{\ell}_{\theta_o})$, $n \in \mathbb{N}$, that is,*

(i) *$\sqrt{n}\,\widehat{\mathbb{P}}_n(\dot{\ell}_{\theta_o}) \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I}_{\theta_o})}$ under $\mathbb{P}_{\theta_o}^{\otimes n}$ and*

(ii) *for $h_n \to h$ it holds $\log(\mathrm{d}\mathbb{P}_{\theta_o+h_n/\sqrt{n}}^{\otimes n} / \mathrm{d}\mathbb{P}_{\theta_o}^{\otimes n}) = \langle \mathcal{Z}_{\theta_o}^n, h\rangle - \frac{1}{2}\langle \mathcal{I}_{\theta_o} h, h\rangle + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$.*

§05.13 **Proof** of Theorem §05.12. is given in the lecture. □

§05.14 **Corollary.** *Under the assumptions of Theorem §05.12 consider for each $n \in \mathbb{N}$ a statistical product experiment $(\mathcal{X}^n, \mathscr{X}^{\otimes n}, \mathbb{P}_\Theta^{\otimes n})$ and an estimator $\widehat{\gamma}_n \in (\mathscr{X}^{\otimes n})^p$ of a parameter of interest $\gamma : \Theta \to \mathbb{R}^p$ allowing an expansion $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta_o)) = \sqrt{n}\,\widehat{\mathbb{P}}_n(\psi_{\theta_o}) + o_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ for some function $\psi_{\theta_o} \in$*

$\mathcal{L}_2^p(\mathbb{P}_{\theta_o})$ with $\mathbb{P}_{\theta_o}(\psi_{\theta_o}) = 0$. Then, $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta_o)) \xrightarrow{d} \mathrm{N}_{(0,\Sigma_o)}$ under $\mathbb{P}_{\theta_o}^{\otimes n}$ with $\Sigma_o := \mathbb{P}_{\theta_o}(\psi_{\theta_o}\psi_{\theta_o}^t)$ and for each $h \in \mathbb{R}^k$ holds $\sqrt{n}(\widehat{\gamma}_n - \gamma(\theta_o)) \xrightarrow{d} \mathrm{N}_{(\tau_h,\Sigma_o)}$ under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$ with $\tau_h := \mathbb{P}_{\theta_o}(\psi_{\theta_o}\dot{\ell}_{\theta_o}^t)h$.

§05.15 **Proof** of Corollary §05.14. is given in the lecture. □

§05.16 **Example** *(Example §03.06 continued).* Under the assumptions of Theorem §05.12 let $\gamma : \theta \to \mathbb{R}^p$ be a parameter of interest. Consider $\mathrm{m}(\gamma) \in \mathcal{L}_1(\mathbb{P}_\theta)$ for all $\gamma \in \mathbb{R}^p$, a criterion process $\widehat{\mathrm{M}}_n(\gamma) = \widehat{\mathbb{P}}_n\big(\mathrm{m}(\gamma)\big)$, a criterion function $\mathrm{M}(\theta,\gamma) = \mathbb{P}_\theta\big(\mathrm{m}(\gamma)\big)$ and a M-estimator $\widehat{\gamma}_n \in \arg\inf_{\gamma\in\Gamma}\{\widehat{\mathrm{M}}_n(\gamma)\}$ of $\{\gamma_o := \gamma(\theta_o)\} = \arg\inf_{\gamma\in\Gamma}\{\mathrm{M}(\theta_o,\gamma)\}$. Under regularity conditions as in Example §03.06 we have $\sqrt{n}(\widehat{\gamma}_n - \gamma_o) = \sqrt{n}\widehat{\mathbb{P}}_n\big(\psi_{\theta_o}\big) + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ with $\psi_{\theta_o} := -\ddot{\mathrm{M}}_o^{-1}\dot{\mathrm{m}}(\gamma_o)$ assuming a regular matrix $\ddot{\mathrm{M}}_o := \mathbb{P}_{\theta_o}\big(\ddot{\mathrm{m}}(\gamma_o)\big)$. Consequently, setting $\Sigma_o = \mathbb{P}_{\theta_o}(\psi_{\theta_o}\psi_{\theta_o}^t) = \ddot{\mathrm{M}}_o^{-1}\mathbb{P}_{\theta_o}(\dot{\mathrm{m}}(\gamma_o)\dot{\mathrm{m}}(\gamma_o)^t)\ddot{\mathrm{M}}_o^{-1}$ from Corollary §05.14 it follows

$$\sqrt{n}(\widehat{\gamma}_n - \gamma_o) \xrightarrow{d} \mathrm{N}_{(\tau_h,\Sigma_o)} \quad \text{under} \quad \mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n} \text{ with } \tau_h = -\ddot{\mathrm{M}}_o^{-1}\mathbb{P}_{\theta_o}(\dot{\mathrm{m}}(\gamma_o)\dot{\ell}_{\theta_o}^t)h.$$

In the particular case of a MLE $\widehat{\theta}_n$ of θ, i.e., $(\gamma = \mathrm{id}_{\mathbb{R}^k})$, as in Example §03.07 setting $\mathrm{m} := -\log(\mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_o)$ we have $\dot{\mathrm{m}}(\theta_o) = -\dot{\ell}_{\theta_o}$, $\mathcal{I}_{\theta_o} = \mathbb{P}_{\theta_o}\big(\dot{\mathrm{m}}(\theta_o)\dot{\mathrm{m}}(\theta_o)^t\big) = \mathbb{P}_{\theta_o}\big(\ddot{\mathrm{m}}(\theta_o)\big) = \ddot{\mathrm{M}}_o$ and thus $\Sigma_o = \ddot{\mathrm{M}}_o^{-1}\mathbb{P}_{\theta_o}(\dot{\mathrm{m}}(\gamma_o)\dot{\mathrm{m}}(\gamma_o)^t)\ddot{\mathrm{M}}_o^{-1} = \mathcal{I}_{\theta_o}^{-1}$ and $\tau_h := -\ddot{\mathrm{M}}_o^{-1}\mathbb{P}_{\theta_o}(\dot{\mathrm{m}}(\theta_o)\dot{\ell}_{\theta_o}^t)h = h$. Therewith, $\sqrt{n}(\widehat{\theta}_n - \theta_o) \xrightarrow{d} \mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}$ under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$. □

§05.17 **Remark.** Supposing $\sqrt{n}(\widehat{\theta}_n - \theta_o) = \sqrt{n}\,\widehat{\mathbb{P}}_n(\psi_{\theta_o}) + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ let us further assume a transformation $A : \Theta \to \mathbb{R}^p$ that is "smooth", and hence by employing the delta method §20.16, for instance satisfies $\sqrt{n}(A(\widehat{\theta}_n) - A(\theta_o)) = \dot{A}_{\theta_o}\sqrt{n}\,\widehat{\mathbb{P}}_n(\psi_{\theta_o}) + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$. Consequently, it follows $\sqrt{n}(A(\widehat{\theta}_n) - A(\theta_o)) \xrightarrow{d} \mathrm{N}_{(\tau_h,\Sigma_o)}$ under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$ with $\tau_h = \dot{A}_{\theta_o}\mathbb{P}_{\theta_o}(\psi_{\theta_o}\dot{\ell}_{\theta_o}^t)h$ and $\Sigma_o = \dot{A}_{\theta_o}\mathbb{P}_{\theta_o}(\psi_{\theta_o}\psi_{\theta_o}^t)\dot{A}_{\theta_o}^t$. In the special case of a MLE we have $\sqrt{n}(A(\widehat{\theta}_n) - A(\theta_o)) \xrightarrow{d} \mathrm{N}_{(\dot{A}_{\theta_o}h,\dot{A}_{\theta_o}\mathcal{I}_{\theta_o}^{-1}\dot{A}_{\theta_o}^t)}$ under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$. □

§06 Asymptotic relative efficiency

§06.01 **Heuristics** *(§03.09 and §03.10 continued).* Under the conditions of Corollary §05.14 consider the statistical testing task $H_0 : A(\theta_o) = 0$ against the alternative $H_1 : A(\theta_o) \neq 0$ for some transformation $A : \Theta \to \mathbb{R}^p$ satisfying $\sqrt{n}(A(\widehat{\theta}_n) - A(\theta_o)) = \dot{A}_{\theta_o}\sqrt{n}\,\widehat{\mathbb{P}}_n(\psi_{\theta_o}) + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$. As in §03.09 let $\widehat{W}_n := nA(\widehat{\theta}_n)^t\widehat{\Sigma}_n^{-1}A(\widehat{\theta}_n)$ where $\widehat{\Sigma}_n = \Sigma + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ is a consistent estimator of $\Sigma = \dot{A}_{\theta_o}\mathbb{P}_{\theta_o}(\psi_{\theta_o}\psi_{\theta_o}^t)\dot{A}_{\theta_o}^t$, then a *Wald test* is given by $\varphi_n := \mathbb{1}_{\{\widehat{W}_n > \chi^2_{p,1-\alpha}\}}$. Thereby, under H_0, i.e. $A(\theta_o) = 0$, we have $\sqrt{n}A(\widehat{\theta}_n) = \dot{A}_{\theta_o}\sqrt{n}\,\widehat{\mathbb{P}}_n(\psi_{\theta_o}) + \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ and $\widehat{W}_n \xrightarrow{d} \chi^2_p$ under $\mathbb{P}_{\theta_o}^{\otimes n}$ which in turn implies $\mathbb{P}_{\theta_o}^{\otimes n}(\varphi_n = 1) \xrightarrow{n\to\infty} \chi^2_p((\chi^2_{p,1-\alpha},\infty)) = \alpha$. In other words, the Wald test is asymptotically a level α test. For each $\theta \in \Theta$ let us denote $\beta_{\varphi_n}(\theta) := \mathbb{P}_\theta^{\otimes n}(\varphi_n) = \mathbb{P}_\theta^{\otimes n}(\varphi_n = 1) = \mathbb{P}_\theta^{\otimes n}(\widehat{W}_n > \chi^2_{p,1-\alpha})$ which equals the power of the Wald test φ_n under H_1, i.e. $\theta \in \Theta$ with $A(\theta) \neq 0$. In the sequel we consider local alternatives of the form $\theta = \theta_o + h/\sqrt{n}$ and thus we are interested in $\beta_{\varphi_n}(\theta_o+h/\sqrt{n}) = \mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}(\widehat{W}_n > \chi^2_{p,1-\alpha})$. Keeping Remark §05.17 under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$ we have $\sqrt{n}A(\widehat{\theta}_n) \xrightarrow{d} \mathrm{N}_{(\dot{A}_{\theta_o}\mathbb{P}_{\theta_o}(\psi_{\theta_o}\dot{\ell}_{\theta_o}^t)h,\Sigma)}$, assuming additionally $\Sigma > 0$ also $\Sigma^{-1/2}\sqrt{n}A(\widehat{\theta}_n) \xrightarrow{d} \mathrm{N}_{(a_h,\mathrm{Id}_p)}$ with $a_h := \Sigma^{-1/2}\dot{A}_{\theta_o}\mathbb{P}_{\theta_o}(\psi_{\theta_o}\dot{\ell}_{\theta_o})^t h$, and hence, $nA(\widehat{\theta}_n)^t\Sigma^{-1}A(\widehat{\theta}_n) \xrightarrow{d} \chi^2_p(\|a_h\|^2)$. Here $\chi^2_p(c)$ denotes a non-central χ^2-distribution with p degrees of freedom and non-centrality parameter $c \in \mathbb{R}^+$. Moreover, $\widehat{W}_n - nA(\widehat{\theta}_n)^t\Sigma^{-1}A(\widehat{\theta}_n) = \mathrm{o}_{\mathbb{P}_{\theta_o}^{\otimes n}}(1)$ and thus $\widehat{W}_n - nA(\widehat{\theta}_n)^t\Sigma^{-1}A(\widehat{\theta}_n) = \mathrm{o}_{\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}}(1)$ due to Lemma §04.19 (ii) by employing that $\mathbb{P}_{\theta_o}^{\otimes n} \lhd\rhd \mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$ are mutually contiguous. Consequently, $\widehat{W}_n \xrightarrow{d} \chi^2_p(\|a_h\|^2)$ under $\mathbb{P}_{\theta_o+h/\sqrt{n}}^{\otimes n}$ and thus $\beta_{\varphi_n}(\theta_o + h/\sqrt{n}) \xrightarrow{n\to\infty} \chi^2_p(\|a_h\|^2)\big((\chi^2_{p,1-\alpha},\infty)\big)$. Note that a_h simplifies to $h^t\dot{A}_{\theta_o}^t(\dot{A}_{\theta_o}\mathcal{I}_{\theta_o}^{-1}\dot{A}_{\theta_o}^t)^{-1}\dot{A}_{\theta_o}h$ in the particular case of a MLE $\widehat{\theta}_n$. □

§06.02 **Reminder** *(Gauß test).* In a Gaussian location model, i.e. $Y \curlywedge \mathrm{N}_{\mathbb{R}^k \times \{\mathcal{I}_{\theta_o}^{-1}\}}$ with $\mathcal{I}_{\theta_o} \in \mathbb{R}_{>}^{(k,k)}$, consider the binary testing task $H_0 : \{\mathrm{N}_{(0,\mathcal{I}_{\theta_o}^{-1})}\}$ against the alternative $H_1 : \{\mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}\}$ for some $h \in \mathbb{R}^k$. In this situation the log-likelihood ratio $\ell_h = \log(\mathrm{d}\mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}/\mathrm{d}\mathrm{N}_{(0,\mathcal{I}_{\theta_o}^{-1})})$ satisfies $\ell_h(y) = \langle \mathcal{I}_{\theta_o} y, h\rangle - \frac{1}{2}\sigma_h^2$ for all $y \in \mathbb{R}^k$ with $\sigma_h^2 := \langle \mathcal{I}_{\theta_o} h, h\rangle$. Consequently, $\ell_h \sim \mathrm{N}_{(-\sigma_h^2/2,\sigma_h^2)}$ under $\mathrm{N}_{(0,\mathcal{I}_{\theta_o}^{-1})}$, i.e. under the hypothesis H_0, and $\ell_h \sim \mathrm{N}_{(\sigma_h^2/2,\sigma_h^2)}$ under $\mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}$, i.e. under the alternative H_1. For $\alpha \in (0,1)$ let $c_{h,1-\alpha} \in \mathbb{R}$ satisfy $\mathrm{N}_{(-\sigma_h^2/2,\sigma_h^2)}\big((c_{h,1-\alpha},\infty)\big) = \alpha$ and thus $\mathrm{N}_{(0,\mathcal{I}_{\theta_o}^{-1})}(\ell_h > c_{h,1-\alpha}) = \mathrm{N}_{(-\sigma_h^2/2,\sigma_h^2)}\big((c_{h,1-\alpha},\infty)\big) = \alpha$. Keeping in mind that any most powerful level-α test has Neyman-Pearson form and the *Gauß test* $\varphi^\star := \mathbb{1}_{\{\ell_h > c_{h,1-\alpha}\}}$ is a Neyman-Pearson level-α test. Its power given by $\beta_{\varphi^\star}(h) := \mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}(\varphi^\star = 1) = \mathrm{N}_{(h,\mathcal{I}_{\theta_o}^{-1})}(\ell_h > c_{h,1-\alpha}) = \mathrm{N}_{(\sigma_h^2/2,\sigma_h^2)}\big((c_{h,1-\alpha},\infty)\big)$ is maximal in the class of all level-α tests, i.e., for any level-α test φ holds $\beta_\varphi(h) \leqslant \beta_{\varphi^\star}(h)$. In other words, $\varphi^\star$ is a most powerful level-α test (Statistik 1, Satz §18.16, p.56). □

§06.03 **Example** *(Neyman-Pearson test).* Assume local asymptotic normality as in Definition §05.07 where $\ell_{h,n} := \log(\mathrm{d}\mathbb{P}^n_{\theta_o+\delta_n h}/\mathrm{d}\mathbb{P}^n_{\theta_o}) \xrightarrow{d} \mathrm{N}_{(-\sigma_h^2/2,\sigma_h^2)}$ under $\mathbb{P}^n_{\theta_o}$ with $\sigma_h^2 := \langle \mathcal{I}_{\theta_o} h, h\rangle$ for $h \in \mathbb{R}^k$. Hence by Le Cam's first lemma (Example §04.34) mutual contiguity $\mathbb{P}^n_{\theta_o+\delta_n h} \triangleleft\triangleright \mathbb{P}^n_{\theta_o}$ and $\ell_{h,n} \xrightarrow{d} \mathrm{N}_{(\sigma_h^2/2,\sigma_h^2)}$ under $\mathbb{P}^n_{\theta_o+\delta_n h}$ hold. Consider the binary testing task of the hypothesis $H_0 : \{\mathbb{P}^n_{\theta_o}\}$ against a local alternative $H_1 : \{\mathbb{P}^n_{\theta_o+\delta_n h}\}$. In this situation $\varphi_n^\star = \mathbb{1}_{\{\ell_{h,n} > c_{h,n,1-\alpha}\}}$ is a *Neyman-Pearson test*, which is a most powerful level-α test, if $\mathbb{P}^n_{\theta_o}(\varphi_n^\star = 1) = \alpha$. Keeping its power function $\beta_{\varphi_n^\star}(\theta) = \mathbb{P}^n_\theta(\varphi_n^\star) = \mathbb{P}^n_\theta(\varphi_n^\star = 1) = \mathbb{P}^n_\theta(\ell_{h,n} > c_{h,n,1-\alpha})$ evaluated at θ in mind the value $\beta_{\varphi_n^\star}(\theta_o + \delta_n h)$ equals the maximal size of the power in the class of all level-α tests. Considering $c_{h,1-\alpha} \in \mathbb{R}$ as in Reminder §06.02 under local asymptotic normality it follows $\alpha = \mathbb{P}^n_{\theta_o}(\varphi_n^\star) = \mathbb{P}^n_{\theta_o}(\ell_{h,n} > c_{h,n,1-\alpha}) \xrightarrow{n\to\infty} \mathrm{N}_{(-\sigma_h^2/2,\sigma_h^2)}\big((c_{h,1-\alpha},\infty)\big) = \alpha$ which implies $c_{h,n,1-\alpha} \xrightarrow{n\to\infty} c_{h,1-\alpha}$, and in addition $\beta_{\varphi_n^\star}(\theta_o+\delta_n h) = \mathbb{P}^n_{\theta_o+\delta_n h}(\varphi_n^\star) = \mathbb{P}^n_{\theta_o+\delta_n h}(\ell_{h,n} > c_{h,n,1-\alpha}) \xrightarrow{n\to\infty} \mathrm{N}_{(\sigma_h^2/2,\sigma_h^2)}\big((c_{h,1-\alpha},\infty)\big) = \beta_{\varphi^\star}(h)$ with Neyman-Pearson test $\varphi^\star$ in a Gaussian location model as in Reminder §06.02. □

§06.04 **Theorem.** *Let $\Theta \subseteq \mathbb{R}$. Consider a one-sided test task $H_0 : (-\infty, \theta_o]$ against $H_1 : (\theta_o, \infty)$. Suppose that $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}^n_\Theta)$ is LAN in $\theta_o \in \Theta$ with localising sequence $(\delta_n)_{n\in\mathbb{N}}$, central sequence $(\mathcal{Z}^n_{\theta_o})_{n\in\mathbb{N}} \in (\mathscr{X}_n)_{n\in\mathbb{N}}$ and strictly positive Fisher information $\mathcal{I}_{\theta_o} \in \mathbb{R}^+_{\setminus 0}$.*

(i) *Given a sequence $(T_n)_{n\in\mathbb{N}} \in (\mathscr{X}_n)_{n\in\mathbb{N}}$ of test statistics satisfying $(T_n, \mathcal{Z}^n_{\theta_o}) \xrightarrow{d} \mathrm{N}_{(0,M)}$ with $M = \big((\sigma^2,\rho)^t, (\rho,\mathcal{I}_{\theta_o})^t\big)$ consider the randomised test $\varphi_n := \mathbb{1}_{\{T_n > c_n\}} + \gamma_n \mathbb{1}_{\{T_n = c_n\}}$ with $\gamma_n \in [0,1]$ and $c_n \in \mathbb{R}$ such that $\beta_{\varphi_n}(\theta_o) = \mathbb{P}^n_{\theta_o}(\varphi_n) = \mathbb{P}^n_{\theta_o}(T_n > c_n) + \gamma_n \mathbb{P}^n_{\theta_o}(T_n = c_n) = \alpha_n \xrightarrow{n\to\infty} \alpha$. Choosing $z_{1-\alpha} \in \mathbb{R}$ with $1 - \mathbb{F}_{[0,1]}(z_{1-\alpha}) := \mathrm{N}_{(0,1)}\big((z_{1-\alpha},\infty)\big) = \alpha$ we have*

$$\beta_{\varphi_n}(\theta_o + \delta_n h) = \mathbb{P}^n_{\theta_o+\delta_n h}(\varphi_n) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\rho/\sigma).$$

(ii) *In case $T_n = \mathcal{Z}^n_{\theta_o}$ consider $\varphi_n^\star = \mathbb{1}_{\{\mathcal{Z}^n_{\theta_o} > z_{1-\alpha}\mathcal{I}_{\theta_o}^{1/2}\}}$, i.e. $\gamma_n = 0$ and $c_n = z_{1-\alpha}\mathcal{I}_{\theta_o}^{1/2}$. Then $\beta_{\varphi_n^\star}(\theta_o) = \mathbb{P}^n_{\theta_o}(\varphi_n^\star) = \mathbb{P}^n_{\theta_o}(\mathcal{I}_{\theta_o}^{-1/2}\mathcal{Z}^n_{\theta_o} > z_{1-\alpha}) \xrightarrow{n\to\infty} 1 - \mathbb{F}_{[0,1]}(z_{1-\alpha}) = \alpha$ and*

$$\beta_{\varphi_n^\star}(\theta_o + \delta_n h) = \mathbb{P}^n_{\theta_o+\delta_n h}(\varphi_n^\star) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\mathcal{I}_{\theta_o}^{1/2}).$$

§06.05 **Proof** of Theorem §06.04. is given in the lecture. □

§06.06 **Remark.**

(a) By using Theorem §04.35 directly it could be possible to calculate an asymptotic power of a test if $\log(\mathrm{d}\mathbb{P}^n_{\theta_o+\delta_n h}/\mathrm{d}\mathbb{P}^n_{\theta_o}) \xrightarrow{d} \mathbb{P}$ under $\mathbb{P}^n_{\theta_o}$ where $\mathbb{P}$ equals not necessarily $\mathrm{N}_{(0,1)}$.

(b) Let $(Y_1, Y_2) \sim \mathrm{N}_{(0,M)}$ with $M = \big((\sigma^2,\rho)^t, (\rho,\mathcal{I}_{\theta_o})^t\big)$ as in Theorem §06.04 (i), then $\rho^2 = |\operatorname{Cov}(Y_1,Y_2)|^2 \leqslant \mathfrak{v}\mathrm{ar}_m(Y_1)\mathfrak{v}\mathrm{ar}_m(Y_2) = \sigma^2\mathcal{I}_{\theta_o}$. Consequently, the test $\varphi_n^\star$ given in (ii) maximises the asymptotic power when considering only a randomised test φ_n as given in part (i) of Theorem §06.04. □

§06.07 **Theorem**. *Let the assumptions of Theorem §06.04 be satisfied. Any test φ_n of the one-sided testing task $H_0 : (\infty, \theta_o]$ against $H_1 : (\theta_o, \infty)$ with $\beta_{\varphi_n}(\theta_o) := \mathbb{P}^n_{\theta_o}(\varphi_n) = \alpha_n \xrightarrow{n\to\infty} \alpha$ fulfils*

(i) $\limsup_{n\to\infty} \beta_{\varphi_n}(\theta_o + \delta_n h) \leqslant \mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\sqrt{\mathcal{I}_{\theta_o}})$ *for all* $h \in \mathbb{R}^+_{\setminus 0}$;

(ii) $\liminf_{n\to\infty} \beta_{\varphi_n}(\theta_o - \delta_n h) \geqslant \mathbb{F}_{[0,1]}(-z_{1-\alpha} - h\sqrt{\mathcal{I}_{\theta_o}})$ *for all* $h \in \mathbb{R}^+_{\setminus 0}$.

§06.08 **Proof** of Theorem §06.07. is given in the lecture. □

§06.09 **Remark**. Keeping Theorem §06.07 in mind we call the test (sequence) $(\varphi^\star_n)_{n\in\mathbb{N}}$ given in Theorem §06.04 (ii) asymptotically uniformly most powerful level-α test (sequence) in the class of all asymptotic level-α test (sequences). Its asymptotic power function equals $\mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\sqrt{\mathcal{I}_{\theta_o}})$ which is the power function of the uniformly most powerful test of $H_0 : (-\infty, 0]$ against $H_1 : (0, \infty)$ in the limit Gaussian location experiment $(\mathbb{R}, \mathscr{B}, \mathrm{N}_{\mathbb{R}\times\{\mathcal{I}^{-1}_{\theta_o}\}})$. □

§06.10 **Asymptotic relative efficiency**. Let $(\mathcal{X}_n, \mathscr{X}_n, \mathbb{P}^n_\Theta)_{n\in\mathbb{N}}$ be LAN with localising rate $\delta_n := n^{-1/2}$. Consider a test φ^a_n satisfying the conditions of Theorem §06.04 (i) and hence, admitting an asymptotic power function such that $\beta_{\varphi^a_n}(\theta_o + h/\sqrt{n}) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\rho_a/\sigma_a)$. Thereby, choosing $\eta = h/\sqrt{n}$ the approximation $\beta_{\varphi^a_n}(\theta_o + \eta) \approx \mathbb{F}_{[0,1]}(-z_{1-\alpha} + \eta\sqrt{n}\rho_a/\sigma_a)$ is reasonable. In analogy, if φ^b_n is another test satisfying the conditions of Theorem §06.04 (i) and admitting $\beta_{\varphi^b_n}(\theta_o + \eta) \approx \mathbb{F}_{[0,1]}(-z_{1-\alpha} + \eta\sqrt{n}\rho_b/\sigma_b)$. Roughly speaking, this means, that at $\theta_o + \eta$ the power of the test $\varphi^a_{n_a}$ and $\varphi^b_{n_b}$ with sample size n_a and n_b, respectively, is approximately equal if $n_a\rho^2_a/\sigma^2_a = n_b\rho^2_b/\sigma^2_b$. The quantity $\mathrm{are}(\varphi^a_{n_a}, \varphi^b_{n_b}) = (n_a/n_b) = (\rho^2_b\sigma^2_a)/(\rho^2_a\sigma^2_b)$ is called *asymptotic relative efficiency*. Meaning, that a sample of size $n_a = \mathrm{are}(\varphi^a_{n_a}, \varphi^b_{n_b})\, n_b$ is needed for the test $\varphi^a_{n_a}$ to attain at $\theta_o + \eta$ approximately the same power $\mathbb{F}_{[0,1]}(-z_{1-\alpha} + \eta\sqrt{n_a}\rho_a/\sigma_a) = \mathbb{F}_{[0,1]}(-z_{1-\alpha} + \eta\sqrt{n_b}\rho_b/\sigma_b)$ as the test $\varphi^b_{n_b}$ with sample size n_b. A comparison with the test $\varphi^\star_n$ as in Theorem §06.04 (ii) allows analogously to introduce a notion of *asymptotic absolute efficiency*. □

§07 Rank tests

§07.01 **Reminder**. Consider on the sample space $(\mathbb{R}^n, \mathscr{B}^n)$ the statistic $T : \mathbb{R}^n \to \mathbb{R}^n$ with $x \mapsto T(x) = (T_i(x))_{i\in[\![n]\!]}$ and $T_i(x) := \min\{c \in \mathbb{R} : \sum_{j\in[\![n]\!]} \mathbb{1}_{\{x_j\leqslant c\}} \geqslant i\}$, $i \in [\![n]\!]$. Since $T_1(x) \leqslant T_2(x) \leqslant \cdots \leqslant T_n(x)$ for all $x \in \mathbb{R}^n$ the statistic T (and any other statistic with this property) is called an *order statistic*. Denote by $\mathcal{S}_n$ the symmetric group of order n, i.e. the set of all permutations of the set $[\![n]\!]$. We identify as usual a vector $s = (s_i)_{i\in[\![n]\!]} \in [\![n]\!]^n$ with the map $s : [\![n]\!] \to [\![n]\!]$, $i \mapsto s(i) := s_i$, and hence $\mathcal{S}_n \subseteq [\![n]\!]^n$. Let $s^- \in \mathcal{S}_n$ denote the inverse permutation of $s \in \mathcal{S}_n$, i.e. $\mathrm{id}_{\mathcal{S}_n} = s \circ s^- = s^- \circ s$. For a permutation $s = (s_i)_{i\in[\![n]\!]} \in \mathcal{S}_n$ and a vector $x = (x_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$ we write shortly $x_s := (x_{s_i})_{i\in[\![n]\!]}$. A Borel-measurable map $S := (S_i)_{i\in[\![n]\!]} : \mathbb{R}^n \to \mathcal{S}_n$, i.e. $S^{-1}(\{s\}) \in \mathscr{B}^n$ for all $s \in \mathcal{S}_n$, is called a *random permutation* on $(\mathbb{R}^n, \mathscr{B}^n)$. The associated map $S^- : \mathbb{R}^n \to \mathcal{S}_n$ satisfying $\mathrm{id}_{\mathcal{S}_n} = S^-(x) \circ S(x) = S(x) \circ S^-(x)$ for all $x \in \mathcal{X}$ is trivially again Borel-measurable, and hence called random inverse permutation of S. Moreover the statistic $X_S : \mathbb{R}^n \to \mathbb{R}^n$ with $x \mapsto X_S(x) := (x_{S_i(x)})_{i\in[\![n]\!]} = x_{S(x)} = \sum_{s\in\mathcal{S}_n} x_s \mathbb{1}_{\{s\}}(S(x))$ (a finite sum of Borel-measurable functions $x \mapsto x_s \mathbb{1}_{S^{-1}(s)}(x)$) is called a *random arrangement*. □

§07.02 **Definition**. A random permutation $O = (O_i)_{i\in[\![n]\!]}$ on $(\mathbb{R}^n, \mathscr{B}^n)$ is called *order permutation*, if the associated random arrangement $X_O : \mathbb{R}^n \to \mathbb{R}^n$ with $x \mapsto x_{O(x)}$ is an order statistic, i.e. $x_{O_1(x)} \leqslant x_{O_2(x)} \leqslant \cdots \leqslant x_{O_n(x)}$ for all $x \in \mathbb{R}^n$. A random permutation $R = (R_i)_{i\in[\![n]\!]}$ on $(\mathbb{R}^n, \mathscr{B}^n)$ is called *rank permutation*, if its random inverse permutation $O := R^-$ is an order permutation. For $i \in [\![n]\!]$ the i-th component $R_i(x)$ of $R(x)$ is called the *rank* of the i-th component of $x \in \mathbb{R}^n$. □

§07.03 **Comment.** An order permutation O is uniquely determined on the Borel-set $\{x_i \neq x_j\} := \{(x_i)_{i\in[\![n]\!]} \in \mathbb{R}^n : x_i \neq x_j, \forall\, j \in [\![n]\!]\backslash\{i\}, \forall\, i \in [\![n]\!]\}$ only. However, for $x \in \mathbb{R}^n$, the permutation $o := O(x) \in \mathcal{S}_n$ and $i \in [\![n]\!]$ the value at the i-th position in the ordered vector x_o equals the value at the o_i-th position in the original vector x. Conversely, for the permutation $r := R(x) \in \mathcal{S}_n$ of the rank permutation $R := O^-$ the value at the r_i-th position in the ordered vector x_o equals the value at the i-th position in the original vector x. □

§07.04 **Remark.** The map $R^* = (R_i^*)_{i\in[\![n]\!]} : \mathbb{R}^n \to \mathcal{S}_n$ with $x \mapsto R_i^*(x) := \sum_{j\in[\![i]\!]} \mathbb{1}_{\{x_i=x_j\}} + \sum_{j\in[\![n]\!]} \mathbb{1}_{\{x_i>x_j\}}$ for each $i \in [\![n]\!]$ is a rank permutation. Indeed, for each $x \in \mathbb{R}^n$ we have $r := R^*(x) \in \mathcal{S}_n$ ($r : [\![n]\!] \to [\![n]\!]$ is injective and hence bijective) and its inverse permutation $o := r^-$ satisfies $x_{o_1} \leqslant x_{o_2} \leqslant \cdots \leqslant x_{o_n}$. Furthermore, each component of R^* is $\mathscr{B}$-$2^{[\![n]\!]}$-measurable, and hence R^* is a rank permutation. On the Borel-set $\{x_i \neq x_j\}$ each rank permutation $R = (R_i)_{i\in[\![n]\!]}$ is uniquely determined by $R_i(x) = \sum_{j\in[\![n]\!]} \mathbb{1}_{\{x_j\leqslant x_i\}} = R_i^*(x)$, $i \in [\![n]\!]$. For each $y \in \mathbb{R}$ define $\widehat{\mathbb{F}}_n(y) := \widehat{\mathbb{P}}_n(\mathbb{1}_{(-\infty,y]})$ with $\widehat{\mathbb{F}}_n(y,x) := \frac{1}{n}\sum_{j\in[\![n]\!]} \mathbb{1}_{\{x_j\leqslant y\}} \in [0,1]$ for all $x \in \mathbb{R}^n$. $\widehat{\mathbb{F}}_n$ is called *empirical cumulative distribution function*. If in addition $r := R(x)$ and $o := r^-$ for $x \in \{x_i \neq x_j\}$ then $i = n\widehat{\mathbb{F}}_n(x_{o_i}, x)$ and $r_i = n\widehat{\mathbb{F}}_n(x_i, x)$ for each $i \in [\![n]\!]$. □

§07.05 **Comment.** We assume a product probability measure $\mathbb{P}^n = \bigotimes_{j\in[\![n]\!]} \mathbb{P}_j$ on the sample space $(\mathbb{R}^n, \mathscr{B}^n)$ where for each $j \in [\![n]\!]$ the marginal probability measure $\mathbb{P}_j \in \mathcal{W}(\mathscr{B})$ admits a Lebesgue density $\mathfrak{f}_j = \mathrm{d}\mathbb{P}_j/\mathrm{d}\lambda$ and hence $\mathbb{P}^n \ll \lambda^n \in \mathcal{M}_\sigma(\mathscr{B}^n)$ with Lebesgue density $\mathrm{d}\mathbb{P}^n/\mathrm{d}\lambda^n = \prod_{j\in[\![n]\!]} \mathfrak{f}_j$. Noting that the complement $\{x_i = x_j\} := \{x_i \neq x_j\}^c$ of the Borel-set $\{x_i \neq x_j\}$ is a λ^n null set, and hence it is also a $\mathbb{P}^n$ null set. Thereby, each rank permutation R on $(\mathbb{R}^n, \mathscr{B}^n)$ with corresponding order permutation $O := R^-$ satisfies $x_{O_1(x)} < x_{O_2(x)} < \cdots < x_{O_n(x)}$ for $\mathbb{P}^n$-a.e. $x \in \mathbb{R}^n$. Moreover, for $\mathbb{P}^n$-a.e. $x \in \mathbb{R}^n$ the vector of ranks $R(x)$ (and the rang permutation R) is determined by $R_i(x) = \sum_{j\in[\![n]\!]} \mathbb{1}_{\{x_j\leqslant x_i\}} = n\widehat{\mathbb{F}}_n(x_i, x)$, $i \in [\![n]\!]$. □

§07.06 **Lemma.** *Consider a product probability measure $\mathbb{P}^{\otimes n}$ on $(\mathbb{R}^n, \mathscr{B}^n)$ with identical marginal distribution $\mathbb{P} \in \mathcal{W}(\mathscr{B})$, cumulative distribution function $\mathbb{F}(y) := \mathbb{P}(\mathbb{1}_{(-\infty,y]})$, $y \in \mathbb{R}$, and Lebesgue density $\mathfrak{f} = \mathrm{d}\mathbb{P}/\mathrm{d}\lambda$. Let R and X_O with $O = R^-$ be a rang permutation on $(\mathbb{R}^n, \mathscr{B}^n)$ and the corresponding order statistic, respectively.*

(i) *R is under $\mathbb{P}^{\otimes n}$ uniformly distributed on the symmetric group $\mathcal{S}_n$, precisely, $(\mathbb{P}^{\otimes n})^R(\{s\}) = (\mathbb{P}^{\otimes n} \circ R^{-1})(\{s\}) = \mathbb{P}^{\otimes n}(R = s) = \frac{1}{n!}$, $s \in \mathcal{S}_n$, in short $R \sim (\mathbb{P}^{\otimes n})^R = \mathcal{U}_{\mathcal{S}_n}$.*

(ii) *R and X_O are independent under $\mathbb{P}^{\otimes n}$.*

(iii) *The distribution of X_O admits under $\mathbb{P}^{\otimes n}$ a Lebesgue density $\mathfrak{f}_{X_O}(x) = n!\mathbb{1}_B(x)\prod_{i\in[\![n]\!]} \mathfrak{f}(x_i)$, $x \in \mathbb{R}^n$, with $B := \{(x_i)_{i\in[\![n]\!]} \in \mathbb{R}^n, x_1 < \ldots < x_n\}$.*

(iv) *For each $i \in [\![n]\!]$ the distribution of the i-th component of X_O admits under $\mathbb{P}^{\otimes n}$ a Lebesgue density $\mathfrak{f}_i(x) = i\binom{n}{i}|\mathbb{F}(x)|^{i-1}|1-\mathbb{F}(x)|^{n-i}\mathfrak{f}(x)$, $x \in \mathbb{R}$.*

§07.07 **Proof** of Lemma §07.06. is given in the lecture. □

§07.08 **Definition.** Let $\mathbb{P}_o$ and $\mathbb{P}$ be probability measures on $(\mathbb{R}, \mathscr{B})$. We say $\mathbb{P}_o$ is *stochastically smaller* than $\mathbb{P}$, or $\mathbb{P}_o \preceq \mathbb{P}$ for short, if $\mathbb{P}_o((c,\infty)) \leqslant \mathbb{P}((c,\infty))$ for all $c \in \mathbb{R}$. If in addition $\mathbb{P}_o \neq \mathbb{P}$, then we write $\mathbb{P}_o \prec \mathbb{P}$. □

§07.09 **Remark.** Roughly speaking, $\mathbb{P}_o \preceq \mathbb{P}$ says that realisations of $\mathbb{P}_o$ are typically smaller than realisations of $\mathbb{P}$. □

§07.10 **Example.** For $\sigma \in \mathbb{R}^+$ consider on $(\mathbb{R}, \mathscr{B})$ a Gaussian location family $\mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}$. Then for all $a, b \in \mathbb{R}$ holds $\mathrm{N}_{(a,\sigma^2)} \prec \mathrm{N}_{(b,\sigma^2)}$ if and only if $a \lneqq b$. More generally, given a location family $\mathbb{P}_{\mathbb{R}}$ on

$(\mathbb{R}, \mathscr{B})$ as introduced in Example §04.17 with likelihood function $\mathrm{L}(\theta, x) = g(x-\theta)$, $\theta, x \in \mathbb{R}$, for some strictly positive Lebesgue-density g on $\mathbb{R}$. Then for all $a, b \in \mathbb{R}$ holds $\mathbb{P}_a \prec \mathbb{P}_b$ if and only if $a \lneqq b$. □

§07.11 **Heuristics.** Given a sample from each distribution $\mathbb{P}_o, \mathbb{P} \in \mathcal{W}(\mathscr{B})$ we consider the testing task $H_o : \mathbb{P} = \mathbb{P}_o$ against the alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$. Loosely speaking, this means, that we aim to reject the null hypothesis if realisations of $\mathbb{P}_o$ are *significantly* smaller than realisation of $\mathbb{P}$. More precisely, we assume a sample of $n = m + l$ independent real random variables $(X_i)_{i\in[\![n]\!]}$ where the first m and the last l have as common marginal distribution $\mathbb{P}_o$ and $\mathbb{P}$, respectively. In other words $X = (X_i)_{i\in[\![n]\!]}$ takes its values in the pooled sample space $(\mathbb{R}^n, \mathscr{B}^n)$. Considering a rank permutation R on $(\mathbb{R}^n, \mathscr{B}^n)$ and an observation $x \in \mathbb{R}^n$ it seems reasonable to reject the hypothesis if the sum of ranks within the first group of m random variables, i.e. $W_o(x) := \sum_{i\in[\![m]\!]} R_i(x)$, takes *sufficiently* smaller values then the sum of ranks within the second group of l random variables, i.e. $W(x) := \sum_{i\in[\![l]\!]} R_{i+m}(x)$ where obviously $W_o(x) + W(x) = \sum_{i\in[\![n]\!]} R_i(x) = \sum_{i\in[\![n]\!]} i = \frac{n(n+1)}{2}$. □

§07.12 **Lemma.** *For $m, l \in \mathbb{N}$ and $n := m + l$ let $R = (R_i)_{i\in[\![n]\!]}$ be a rang permutation on $(\mathbb{R}^n, \mathscr{B}^n)$, $W_o := \sum_{i\in[\![m]\!]} R_i$, $W := \sum_{i\in[\![l]\!]} R_{i+m}$ and $U_{ml} : \mathbb{R}^n \to [\![0, ml]\!]$ with $x \mapsto U_{ml}(x) := \sum_{i\in[\![m]\!]} \sum_{j\in[\![l]\!]} \mathbb{1}_{\{x_i > x_{j+m}\}}$. Then for each $x \in \{x_i \neq x_j\}$ it holds $W_o(x) = U_{ml}(x) + \frac{m(m+1)}{2}$ and consequently $W(x) = ml - U_{ml}(x) + \frac{l(l+1)}{2}$.*

§07.13 **Proof** of Lemma §07.12. is given in the lecture. □

§07.14 **Comment.** Keeping Lemma §07.12 in mind, we use the test statistic W_o or equivalently U_{ml} to reject the hypothesis $H_0 : \mathbb{P} = \mathbb{P}_o$ against the alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$, if $U_{ml} < c$ or equivalently $W_o < c + \frac{m(m+1)}{2}$ for a certain threshold $c \in (0, ml]$. The test is called (one-sided) Mann-Whitney U-test or Wilcoxon two-sample rank sum test[1]. The critical value has to be chosen according to a pre-specified level $\alpha \in (0, 1)$ which under the null hypothesis necessitates the knowledge of the distribution of U_{ml} or an asymptotic approximation. Interestingly the next proposition shows that under the null hypothesis the distribution of U_{ml} is *distribution free* in the following sense: If $\mathbb{P}_o = \mathbb{P}$ and $\mathbb{P}$ admits a Lebesgue density, then the distribution of U_{ml} is determined and it is independent of the underlying distribution $\mathbb{P}$. □

§07.15 **Proposition.** *For $m, l \in \mathbb{N}$ and $n := m + l$ let $\mathbb{P}^{\otimes n} \in \mathcal{W}(\mathscr{B}^n)$ with identical marginal distribution $\mathbb{P} \ll \lambda$. For all $k \in [\![0, ml]\!]$ it holds $\mathbb{P}^{\otimes n}(U_{ml} = k) = N(k; m, l)/\binom{n}{k}$ where $N(k; m, l)$ denotes the number of all partitions $\sum_{i\in[\![m]\!]} k_i = k$ of k in m increasingly ordered numbers $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_m$ taking from the set $[\![0, l]\!]$. In particular, it holds $\mathbb{P}^{\otimes n}(U_{ml} = k) = \mathbb{P}^{\otimes n}(U_{ml} = ml - k)$.*

§07.16 **Proof** of Proposition §07.15. is given in the lecture. □

§07.17 **Remark.** For small values of k the partition number $N(k; m, l)$ can be calculated by combinatorical means and there exists tables gathering certain quantiles of the U_{ml}-distribution. However, for large values of k the exact calculation of quantiles of the U_{ml}-distribution may be avoided by using an appropriate asymptotic approximation. In the sequel we let m and l and thus $n = m + l$ tend to infinity, which formally means that we consider sequences $(m_n)_{n\in\mathbb{N}}$ and $(l_n)_{n\in\mathbb{N}}$ satisfying $m_n + l_n = n$ for any $n \in \mathbb{N}$. Here and subsequently we assume that $m_n/n \xrightarrow{n\to\infty} \gamma \in (0, 1)$ and hence $l_n/n \xrightarrow{n\to\infty} 1 - \gamma$. For sake of presentation, however, we drop the additional index n and write shortly $n = m + l$ with $m/n \xrightarrow{n\to\infty} \gamma$ and hence $l/n \xrightarrow{n\to\infty} 1 - \gamma$. □

[1] The version based on W_o has been proposed by Wilcoxon [1945], while the U_{ml}-version has been independently be introduced by Mann and Whitney [1947].

§07.18 **Theorem.** *For $m, l \in \mathbb{N}$ and $n := m + l$ let $\mathbb{P}^{\otimes n} \in \mathcal{W}(\mathscr{B}^n)$ with identical marginal distribution $\mathbb{P} \ll \lambda$, and hence continuous cumulative distribution function $\mathbb{F}$. Consider $U_{ml} : \mathbb{R}^n \to [\![0, ml]\!]$ and $T_{ml} : \mathbb{R}^n \to \mathbb{R}$ with $x \mapsto U_{ml}(x) := \sum_{i\in[\![m]\!]} \sum_{j\in[\![l]\!]} \mathbb{1}_{\{x_i > x_{j+m}\}}$ and*

$$x \mapsto T_{ml}(x) := l \sum_{i\in[\![m]\!]} \mathbb{F}(x_i) - m \sum_{i\in[\![l]\!]} \mathbb{F}(x_{i+m}) = l \sum_{i\in[\![m]\!]} (\mathbb{F}(x_i) - 1/2) - m \sum_{i\in[\![l]\!]} (\mathbb{F}(x_{i+m}) - 1/2).$$

Define further $v_{ml} := ml(n+1)/12$, $T^\star_{ml} := T_{ml}/\sqrt{v_{ml}}$ and $U^\star_{ml} := (U_{ml} - ml/2)/\sqrt{v_{ml}}$. If in addition $m/n \to \gamma \in (0,1)$ then $U^\star_{ml} - T^\star_{ml} = \mathrm{o}_{\mathbb{P}^{\otimes n}}(1)$ and $T^\star_{ml} \xrightarrow{d} \mathrm{N}_{(0,1)}$ under $\mathbb{P}^{\otimes n}$, and thus $U^\star_{ml} \xrightarrow{d} \mathrm{N}_{(0,1)}$ under $\mathbb{P}^{\otimes n}$.

§07.19 **Proof** of Theorem §07.18. is given in the lecture. □

§07.20 **Remark.** Considering two independent samples $(X_i)_{i\in[\![m]\!]} \sim \mathbb{P}_o^{\otimes m}$ and $(X_{i+m})_{i\in[\![l]\!]} \sim \mathbb{P}^{\otimes l}$ set $n := m + l$ and $X := (X_i)_{i\in[\![n]\!]}$. Keeping Theorem §07.18 in mind we reject the null hypothesis $H_o : \mathbb{P}_o = \mathbb{P}$ against the alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$, if $U_{ml}(X) < ml/2 + z_\alpha\sqrt{v_{ml}}$ with $\mathbb{F}_{[0,1]}(z_\alpha) = \alpha \in (0,1)$. This test is asymptotically a level-α test due to Theorem §07.18 by exploiting that under the null $\mathbb{P}^{\otimes n}(U_{ml} < ml/2 + z_\alpha\sqrt{v_{ml}}) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(z_\alpha) = \alpha$ for $m/n \xrightarrow{n\to\infty} \gamma \in (0,1)$. Note that we reject similarly the null hypothesis $H_o : \mathbb{P}_o = \mathbb{P}$ against the alternative $H_1 : \mathbb{P} \prec \mathbb{P}_o$ if $U_{ml} > ml/2 + z_{1-\alpha}\sqrt{v_{ml}}$. Next we study the (asymptotic) size of the power of the rank test under local alternatives where we use that under the assumptions of Theorem §07.18 it holds

$$\begin{aligned} U^\star_{ml} = \frac{U_{ml} - ml/2}{\sqrt{v_{ml}}} &= \sqrt{\tfrac{l}{n+1}}\tfrac{1}{\sqrt{m}} \sum_{i\in[\![m]\!]} \tfrac{\mathbb{F}(X_i)-1/2}{\sqrt{1/12}} - \sqrt{\tfrac{m}{n+1}}\tfrac{1}{\sqrt{l}} \sum_{i\in[\![l]\!]} \tfrac{\mathbb{F}(X_{i+m})-1/2}{\sqrt{1/12}} + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1) \\ &= \sqrt{1-\gamma}\,\sqrt{m}\,\widehat{\mathbb{P}}_m(g) - \sqrt{\gamma}\,\sqrt{l}\,\widehat{\mathbb{P}}_l(g) + \mathrm{o}_{\mathbb{P}^{\otimes n}}(1) \quad (07.01) \end{aligned}$$

setting $g := \sqrt{12}(\mathbb{F} - 1/2)$, $\widehat{\mathbb{P}}_m(g) := \frac{1}{m}\sum_{i\in[\![m]\!]} g(X_i)$ and $\widehat{\mathbb{P}}_l(g) := \frac{1}{l}\sum_{i\in[\![l]\!]} g(X_{i+m})$ where $\widehat{\mathbb{P}}_m(g)$ and $\widehat{\mathbb{P}}_l(g)$ are independent, $\mathbb{P}(g) = 0$, and $\mathbb{P}(g^2) = 1$ by construction. □

§08 Asymptotic power of rank tests

§08.01 **Motivation.** Considering the test of the hypothesis $H_0 : \mathbb{P}_o = \mathbb{P}$ against the alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$ we restrict our attention to the special case that $\mathbb{P}_o$ and $\mathbb{P}$ belong to a location family $\mathbb{P}_\mathbb{R}$ as introduced in Example §04.17. Precisely, we assume that the family $\mathbb{P}_\mathbb{R}$ of probability measures on $(\mathbb{R}, \mathscr{B})$ is dominated by the Lebesgue measure. For each $\theta \in \mathbb{R}$, $\mathbb{P}_\theta$ admits a likelihood function given by $\mathrm{L}(\theta, x) = g(x-\theta)$, $x \in \mathbb{R}$, where g is a continuous and strictly positive density on $\mathbb{R}$. Recall that in this context $\mathbb{P}_a \prec \mathbb{P}_b$ holds if and only if $a \lneqq b$ (see Example §07.10). Observe further that we can assume that $\mathbb{P}_o = \mathbb{P}_0$ (possibly after a reparametrisation). For $m, l \in \mathbb{N}$ and $n = m + l$ supposing independent random variables $(X_i)_{i\in[\![n]\!]}$ with $(X_i)_{i\in[\![m]\!]} \wr \mathbb{P}_\mathbb{R}^{\otimes m}$ and $(X_{i+m})_{i\in[\![l]\!]} \sim \mathbb{P}_0^{\otimes l}$ their joint distribution belongs to the two sample location family $\mathbb{P}_\mathbb{R}^{m,l} := \big(\mathbb{P}_\theta^{m,l} := \mathbb{P}_\theta^{\otimes m} \otimes \mathbb{P}_0^{\otimes l}\big)_{\theta\in\mathbb{R}}$. Summarising, based on the statistical two sample location experiment $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P}_\mathbb{R}^{m,l})$ the aim is to test the hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : 0 < \theta$. □

§08.02 **Regular location model.** A location family $\mathbb{P}_\mathbb{R}$ of probability measures on $(\mathbb{R}, \mathscr{B})$ dominated by the Lebesgue measure $\lambda \in \mathcal{M}_\sigma(\mathscr{B})$ with likelihood for each $\theta \in \mathbb{R}$ and a strictly positive density $g \in \mathscr{B}^+$ given by $\mathrm{L}(\theta, x) = g(x-\theta)$, $x \in \mathbb{R}$, is called *regular* if the density g is in addition continuously differentiable with derivative $\dot{g}$ satisfying $\lambda(|\dot{g}|^2/g) < \infty$. □

§08.03 **Reminder**. A regular location family $\mathbb{P}_{\mathbb{R}}$ is Hellinger-differentiable in each $\theta \in \mathbb{R}$ with score function $\dot{\ell}_\theta = -\dot{g}(x-\theta)/g(x-\theta)$ and Fisher information $\mathcal{I} := \mathcal{I}_\theta = \lambda(|\dot{g}|^2/g)$ (see Example §04.17). Due to Theorem §05.12 the statistical product experiment $(\mathbb{R}^m, \mathscr{B}^m, \mathbb{P}_{\mathbb{R}}^{\otimes m})$ is ULAN in $\theta_o = 0$ with localising rate $(\delta_m := m^{-1/2})_{m\in\mathbb{N}}$ and central sequence $(\mathcal{Z}_0^m := -\sqrt{m}\widehat{\mathbb{P}}_m(\dot{g}/g))_{m\in\mathbb{N}}$. Precisely, for any sequence $h_m \to h$ as $m \to \infty$ it holds

$$\log(\mathrm{d}\mathbb{P}_{h_m/\sqrt{m}}^{\otimes m}/\mathrm{d}\mathbb{P}_0^{\otimes m}) = -h\sqrt{m}\widehat{\mathbb{P}}_m(\dot{g}/g) - \tfrac{1}{2}h^2\mathcal{I} + \mathrm{o}_{\mathbb{P}_0^{\otimes m}}(1)$$

and $\sqrt{m}\widehat{\mathbb{P}}_m(\dot{g}/g) \xrightarrow{d} \mathrm{N}_{(0,\mathcal{I})}$ under $\mathbb{P}_0^{\otimes m}$. Given a two sample location family $\mathbb{P}_{\mathbb{R}}^{m,l}$ for any $\theta \in \mathbb{R}$ the log of the likelihood-ratio satisfies $\log(\mathrm{d}\mathbb{P}_\theta^{m,l}/\mathrm{d}\mathbb{P}_0^{m,l}) = \log(\mathrm{d}\mathbb{P}_\theta^{\otimes m}/\mathrm{d}\mathbb{P}_0^{\otimes m})$. Thereby, if the location family is regular and $m/n \xrightarrow{n\to\infty} \gamma \in (0,1)$, whence $h_m := h\sqrt{m/n} \xrightarrow{n\to\infty} h\sqrt{\gamma}$, it follows

$$\begin{aligned}\ell_{h,n} := \log(\mathrm{d}\mathbb{P}_{h/\sqrt{n}}^{m,l}/\mathrm{d}\mathbb{P}_0^{m,l}) &= \log(\mathrm{d}\mathbb{P}_{h_m/\sqrt{m}}^{\otimes m}/\mathrm{d}\mathbb{P}_0^{\otimes m}) \\ &= -h\sqrt{\gamma}\sqrt{m}\,\widehat{\mathbb{P}}_m(\dot{g}/g) - \tfrac{\gamma}{2}h^2\mathcal{I} + \mathrm{o}_{\mathbb{P}_0^{\otimes n}}(1) \quad (08.01)\end{aligned}$$

We consider in the sequel a rank test $\varphi_n = \mathbb{1}_{\{U_{ml}^\star > z_{1-\alpha}\}}$ with $\mathbb{F}_{[0,1]}(-z_{1-\alpha}) = \alpha \in (0,1)$ based on the standardised test statistic $U_{ml}^\star = (U_{ml} - ml/2)/\sqrt{v_{ml}}$ and its asymptotic decomposition given in (07.01). □

§08.04 **Theorem**. *Assume a two sample regular location model $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P}_{\mathbb{R}}^{m,l})$, $n = m + l \in \mathbb{N}$. Consider a rank test $\varphi_n = \mathbb{1}_{\{U_{ml}^\star > z_{1-\alpha}\}}$ with $\mathbb{F}_{[0,1]}(-z_{1-\alpha}) = \alpha \in (0,1)$ for the testing task $H_0 : \theta = 0$ against $H_1 : \theta > 0$. If $m/n \xrightarrow{n\to\infty} \gamma \in (0,1)$, then:*

(i) *Under the null hypothesis $H_o : \theta = 0$ we have $\mathbb{P}_0^{m,l}(\varphi_n) = \mathbb{P}_0^{\otimes m+l}(U_{ml}^\star > z_{1-\alpha}) \xrightarrow{n\to\infty} \alpha$, i.e., φ_n is an asymptotic level-α test;*

(ii) *The power function $\beta_{\varphi_n}(\theta) = \mathbb{P}_\theta^{m,l}(\varphi_n)$, $\theta \in \mathbb{R}$, of the rank test $\varphi_n = \mathbb{1}_{\{U_{ml}^\star > z_{1-\alpha}\}}$ satisfies under local alternatives $\beta_{\varphi_n}(h/\sqrt{n}) = \mathbb{P}_{h/\sqrt{n}}^{m,l}(U_{ml}^\star > z_{1-\alpha}) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(-z_{1-\alpha} + h\rho)$ with $\rho = \lambda(g^2)\sqrt{12\gamma(1-\gamma)}$ for each $h \in \mathbb{R}^+$.*

§08.05 **Proof** of Theorem §08.04. is given in the lecture. □

§08.06 **Comment**. Let us briefly consider a rank test $\varphi_n = \mathbb{1}_{\{U_{ml} < ml/2 + z_\alpha\sqrt{v_{ml}}\}} = \mathbb{1}_{\{U_{ml}^\star < z_\alpha\}}$ with $\mathbb{F}_{[0,1]}(z_\alpha) = \alpha \in (0,1)$ for the testing task of the null hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : 0 > \theta$. Similar to Theorem §08.04 φ_n is an asymptotic level-α test with power for local alternatives $\beta_{\varphi_n}(-h/\sqrt{n}) = \mathbb{P}_{-h/\sqrt{n}}^{m,l}(U_{ml}^\star < z_\alpha) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(z_\alpha + h\rho)$, $h \in \mathbb{R}^+$. □

§08.07 **Two sample Gaussian location model**. For $m, l \in \mathbb{N}$, $n = m + l$ and variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$ the joint distribution of independent random variables $(X_i)_{i\in[\![n]\!]}$ with $(X_i)_{i\in[\![m]\!]} \backsim \mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}^{\otimes m}$ and $(X_{i+m})_{i\in[\![l]\!]} \sim \mathrm{N}_{(0,\sigma^2)}^{\otimes l}$ belongs to a *two sample Gaussian location model* $\mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}^{m,l} := \big(\mathrm{N}_{(\theta,\sigma^2)}^{m,l} := \mathrm{N}_{(\theta,\sigma^2)}^{\otimes m} \otimes \mathrm{N}_{(0,\sigma^2)}^{\otimes l}\big)_{\theta\in\mathbb{R}}$. □

§08.08 **Remark**. For $\sigma \in \mathbb{R}_{\setminus 0}^+$ a Gaussian location family $\mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}$ on $(\mathbb{R}, \mathscr{B})$ is regular with score function $\dot{\ell}_\theta(x) = (x-\theta)/\sigma$, $x \in \mathbb{R}$, and Fisher information $\mathcal{I}_\theta = \lambda(|\dot{g}|^2/g) = \mathrm{N}_{(\theta,\sigma^2)}(\dot{\ell}_\theta^2) = \int_{\mathbb{R}}(x-\theta)^2/\sigma^2\mathrm{N}_{(\theta,\sigma^2)}(dx) = 1$, $\theta \in \mathbb{R}$ (using the notations in Example §04.17). □

§08.09 **Example**. Assume a two sample Gaussian location model $(\mathbb{R}^n, \mathscr{B}^n, \mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}^{m,l})$, $n = m + l \in \mathbb{N}$ and $\sigma \in \mathbb{R}_{\setminus 0}^+$. Define the statistics $T_{ml} := l\sum_{i\in[\![m]\!]} X_i - m\sum_{i\in[\![l]\!]} X_{i+m} \in \mathscr{B}^n$ and $V_{ml}^2 := \frac{ml(m+l)}{(m+l)-2}\big(\sum_{i\in[\![m]\!]}(X_i - \frac{1}{m}\sum_{i\in[\![m]\!]} X_i)^2 + \sum_{i\in[\![l]\!]}(X_{i+m} - \frac{1}{l}\sum_{i\in[\![l]\!]} X_{i+m})^2\big) \in \mathscr{B}^n$, and set $V_{ml} :=$

$\sqrt{V_{ml}^2}$. Under $\mathrm{N}_{(0,\sigma^2)}^{m,l}$ the standardised (Student-) t-statistic $T_{ml}^\star := T_{ml}/V_{ml} \in \overline{\mathscr{B}}^n$ has a t_{n-2}-distribution with $n-2$ degrees of freedom, i.e. $T_{ml}^\star \sim \mathrm{t}_{n-2}$. We denote by $\mathrm{t}_{n-2,\alpha}$ a α-quantile of a t_{n-2}-distribution, i.e., $\mathbb{F}_{\mathrm{t}_{n-2}}(\mathrm{t}_{n-2,\alpha}) = \alpha \in (0,1)$. Consider for the testing task of the null hypothesis $H_o : \theta = 0$ against the alternative $H_1 : 0 < \theta$ (or $H_1 : 0 > \theta$) the t-test $\varphi_n^\star = \mathbb{1}_{\{T_{ml}^\star > \mathrm{t}_{n-2,1-\alpha}\}}$ (or $\varphi_n^\star = \mathbb{1}_{\{T_{ml}^\star < \mathrm{t}_{n-2,\alpha}\}}$), which is by construction a level-α test. Since a Gaussian location model is regular (see Remark §08.08) we can directly apply Theorem §08.04 to derive its asymptotic power function under local alternatives. However, Theorem §08.04 allows us to study a t-test in an arbitrary regular location model (Definition §08.02). More precisely, for $\theta \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\setminus 0}^+$ define $\mathbb{v}_{(\theta,\sigma)} \in \mathscr{B}$ with $x \mapsto \mathbb{v}_{(\theta,\sigma)}(x) := (x-\theta)/\sigma$. As in Remark §08.08 $\dot{\ell}_\theta = \mathbb{v}_{(\theta,\sigma)}$ and $\mathcal{I}_\theta = \mathrm{N}_{(\theta,\sigma^2)}(\mathbb{v}_{(\theta,\sigma)}^2) = 1$, $\theta \in \mathbb{R}$, is the score function and Fisher information, respectively, in a Gaussian location family $\mathrm{N}_{\mathbb{R}\times\{\sigma^2\}}$ with variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$. Considering a regular location family $\mathbb{P}_\mathbb{R}$ with $\mathbb{P}_0 \in \mathcal{W}_2(\mathscr{B})$ (see Notation §19.05) and hence $\mathbb{P}_\theta \in \mathcal{W}_2(\mathscr{B})$ for all $\theta \in \mathbb{R}$ we have $\sigma^2 := \mathbb{P}_0(\mathbb{v}_{(0,1)}^2) = \lambda(\mathbb{v}_{(0,1)}^2 g) = \lambda(\mathrm{id}_\mathbb{R}^2 g) = \int_\mathbb{R} x^2 g(x)\lambda(dx) < \infty$ and $\mathbb{P}_\theta(\mathbb{v}_{(\theta,\sigma)}^2) = 1$ for all $\theta \in \mathbb{R}$ exploiting the translation invariance of the Lebesgue measure. □

§08.10 **Regular mean location family with finite variance.** A regular location family $\mathbb{P}_\mathbb{R}$ of probability measures on $(\mathbb{R}, \mathscr{B})$ is said to have *finite variance* $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$, if $\mathbb{P}_0 \in \mathcal{W}_2(\mathscr{B})$ (and hence $\mathbb{P}_\theta \in \mathcal{W}_2(\mathscr{B})$ for all $\theta \in \mathbb{R}$), and $\sigma^2 = \mathbb{P}_0(\mathbb{v}_{(0,1)}^2)$ (and hence $\mathbb{P}_\theta(\mathbb{v}_{(\theta,\sigma)}^2) = 1$ for all $\theta \in \mathbb{R}$). We call a regular location family satisfying in addition $\mathbb{P}_0(\mathbb{v}_{(0,1)}) = 0$ (and hence $\mathbb{P}_\theta(\mathbb{v}_{(\theta,\sigma)}) = 0$ for all $\theta \in \mathbb{R}$) a *regular mean location family*. □

§08.11 **Theorem.** *Assume a two sample regular mean location model $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P}_\mathbb{R}^{m,l})$ with finite variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$. Consider a* t*-test $\varphi_n^\star = \mathbb{1}_{\{T_{m,l}^\star > \mathrm{t}_{n-2,1-\alpha}\}}$ with $1 - \mathbb{F}_{\mathrm{t}_{n-2}}(\mathrm{t}_{n-2,1-\alpha}) = \alpha \in (0,1)$ for the testing task $H_0 : \theta = 0$ against $H_1 : \theta > 0$. If $m/n \xrightarrow{n\to\infty} \gamma \in (0,1)$, then:*

(i) *Under the null hypothesis $H_0 : \theta = 0$ we have $\mathbb{P}_0^{m,l}(\varphi_n^\star) = \mathbb{P}_0^{m,l}(T_{m,l}^\star > \mathrm{t}_{n-2,1-\alpha}) \xrightarrow{n\to\infty} \alpha$, i.e., $\varphi_n^\star$ is an asymptotic level-α test;*

(ii) *The power function $\beta_{\varphi_n^\star}(\theta) = \mathbb{P}_\theta^{m,l}(\varphi_n^\star)$, $\theta \in \mathbb{R}$ of the* t*-test $\varphi_n^\star = \mathbb{1}_{\{T_{m,l}^\star > \mathrm{t}_{n-2,1-\alpha}\}}$ satisfies under local alternatives $\beta_{\varphi_n^\star}(h/\sqrt{n}) = \mathbb{P}_{h/\sqrt{n}}^{m,l}(T_{ml}^\star > \mathrm{t}_{n-2,1-\alpha}) \xrightarrow{n\to\infty} \mathbb{F}_{[0,1]}(-z_{1-\alpha} + \rho)$ with $\rho = h\sigma^{-1}\sqrt{\gamma(1-\gamma)}$.*

§08.12 **Proof** of Theorem §08.11. is given in the lecture. □

§08.13 **Remark.** Given a two sample regular mean location model $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P}_\mathbb{R}^{m,l})$, $n = m + l \in \mathbb{N}$ with density $g \in \mathscr{B}^+$ and finite variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$ let us compare the asymptotic level-α rank-test $\varphi_n = \mathbb{1}_{\{U_{ml}^\star > z_{1-\alpha}\}}$ (see Theorem §08.04) and the t-test $\varphi_n^\star = \mathbb{1}_{\{T_{ml}^\star > \mathrm{t}_{n-2,1-\alpha}\}}$ (see Theorem §08.11). Using their asymptotic power functions the *asymptotic relative efficiency* (see Definition §06.10) between both tests equals $\mathrm{are}(\varphi_n, \varphi_n^\star) = 12\sigma^2(\lambda g^2)^2$. In the particular case of a Gaussian location model, i.e., $g(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-x^2/(2\sigma^2))$ we have $\lambda g^2 = 1/(2\sigma\sqrt{\pi})$ and hence $\mathrm{are}(\varphi_n, \varphi_n^\star) = 3/\pi \approx 0.955$. On the other hand denoting by $\mathcal{D}$ the class of all Lebesgue-densities $g \in \mathscr{B}^+$ satisfying $\lambda(\mathbb{v}_{(0,\sigma)} g) = 0$ and $\lambda(\mathbb{v}_{(0,\sigma)}^2 g) = 1$ Hodges and Lehmann [1956] have shown that $\inf_{g\in\mathcal{D}} 12\sigma^2(\lambda g^2)^2 = 0.864$ and $\sup_{g\in\mathcal{D}} 12\sigma^2(\lambda g^2)^2 = \infty$. □

Chapter 3

Nonparametric estimation by projection

This chapter presents an introduction to nonparametric estimation of curves along the lines of the textbooks by **Tsybakov** *[2009] and* **Comte** *[2015] where far more details, examples and further discussions can be found.*

Overview

§09 Review

Nonparametric density estimation. Consider for a non-empty set of parameters Θ a family $\mathbb{P}_\Theta$ of probability measures on $(\mathbb{R}, \mathscr{B})$ which contains the distribution of an observable real random variable, $X \bigcirc\!\!\!\!\sim \mathbb{P}_\Theta$. The family $\mathbb{P}_\Theta$ captures the prior knowledge about the distribution of the observation. For example, a family given by a set of parameters Θ containing only one singleton, i.e., $\Theta = \{\theta_o\}$, and hence $X \sim \mathbb{P}_{\theta_o}$ for some probability measure $\mathbb{P}_{\theta_o} \in \mathcal{W}(\mathscr{B})$, means, the data generating process is known to us in advance. On the contrary, a parameter set $\Theta = \mathcal{W}(\mathscr{B})$ reflects a lack of prior knowledge. A parametric model $\mathbb{P}_\Theta$ for some parameter set $\Theta \subseteq \mathbb{R}^k$ provides in a certain sense a trade-off between both extremes. In this chapter

our aim is to avoid an assumption of a finite dimensional set of parameters. For example, consider $(X_i)_{i\in[\![n]\!]} \overset{i.i.d.}{\sim} \mathbb{P} \in \mathcal{W}(\mathscr{B})$, that is, an independent and identically distributed sample with common probability measure $\mathbb{P} \in \mathcal{W}(\mathscr{B})$. A reasonable estimator of the associated cumulative distribution function (c.d.f.) $\mathbb{F}(t) := \mathbb{P}((-\infty, t])$, $t \in \mathbb{R}$, is the empirical cumulative distribution function (e.c.d.f.) $\widehat{\mathbb{F}}_n(t) := \widehat{\mathbb{P}}_n((-\infty, t])$, $t \in \mathbb{R}$. For each $t \in \mathbb{R}$, $\widehat{\mathbb{F}}_n(t)$ is an unbiased estimator of $\mathbb{F}(t)$ with variance $\mathfrak{var}_m(\widehat{\mathbb{F}}_n(t)) = \frac{1}{n}\mathbb{F}(t)(1-\mathbb{F}(t))$. Consequently, $\widehat{\mathbb{F}}_n(t)$ converges in probability to $\mathbb{F}(t)$, and thus it is a consistent estimator. Moreover, by the law of large numbers the convergence holds almost surely in any point and also uniformly, by Glivenko-Cantelli's theorem, i.e., $\|\widehat{\mathbb{F}}_n - \mathbb{F}\|_{\mathcal{L}_\infty} = \mathrm{o}(1)$ $\mathbb{P}$-a.s.. If we assume in addition that $\mathbb{P}$ admits a Lebesgue density then $\widehat{\mathbb{F}}_n$ is a unbiased estimator with minimal variance, by Lehman-Scheffé's theorem. However, comparing different probability measures using their associated c.d.f.'s is visually difficult and as a consequence, other measures for dissimilarities are typically used. Consider, for instance, for two probability measures $\mathbb{P}$ and $\mathbb{P}_o$ on $(\mathbb{R}, \mathscr{B})$ their *total variation distance* given by $\|\mathbb{P} - \mathbb{P}_o\|_{\mathrm{TV}} := \sup\{|\mathbb{P}(B) - \mathbb{P}_o(B)|, B \in \mathscr{B}\}$. We note that for any probability measure $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ admitting a Lebesgue-density we have $\|\mathbb{P} - \widehat{\mathbb{P}}_n\|_{\mathrm{TV}} = 1$ $\mathbb{P}$-a.s. for any $n \in \mathbb{N}$. As a consequence the empirical probability measure $\widehat{\mathbb{P}}_n$ is not a consistent estimator of $\mathbb{P}$ in terms of the total variation distance. In other words, dependending on the measure of accuracy (metric, topology, etc.) a different estimator of $\mathbb{P}$ might be reasonable.

§09.01 **Lemma** *(Scheffé's theorem). Let $\mathbb{P}, \mathbb{P}_o \in \mathcal{W}(\mathscr{B})$ admit a μ-density $\mathbb{p}$ and $\mathbb{p}_o$, respectively, for some $\mu \in \mathcal{M}_\sigma(\mathscr{B})$. Then $\|\mathbb{P} - \mathbb{P}_o\|_{\mathrm{TV}} = \frac{1}{2}\mu(|\mathbb{p} - \mathbb{p}_o|) = \frac{1}{2}\|\mathbb{p} - \mathbb{p}_o\|_{\mathcal{L}_1(\mu)}$.*

§09.02 **Proof** of Lemma §09.01. is given in the lecture. □

In the sequel let $\mathcal{D}$ be the set of Lebesgue densities on $(\mathbb{R}, \mathscr{B})$, and hence $\mathcal{D} \subseteq \mathcal{L}_1 = \mathcal{L}_1(\mathscr{B}, \lambda)$. $\mathbb{P}_{\mathbb{p}} = \mathbb{p}\lambda$ and $\mathbb{E}_{\mathbb{p}}$ denote for each density $\mathbb{p} \in \mathcal{D}$ the associated probability measure and expectation, respectively. We consider the statistical product experiment $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P}_{\mathcal{D}}^{\otimes n} = (\mathbb{P}_{\mathbb{p}}^{\otimes n})_{\mathbb{p}\in\mathcal{D}})$ and $(X_i)_{i\in[\![n]\!]} \bigcirc\!\!\sim \mathbb{P}_{\mathcal{D}}^{\otimes n}$. Typically, for $s \geqslant 1$ we access the accuracy of an estimator $\widehat{\mathbb{p}}$ of $\mathbb{p}$ either by a local measure, e.g. $\mathbb{P}_{\mathbb{p}}^{\otimes n}(|\widehat{\mathbb{p}}(t) - \mathbb{p}(t)|^s)$, for $t \in \mathbb{R}$, or by a global measure, e.g. $\mathbb{P}_{\mathbb{p}}^{\otimes n}(\|\widehat{\mathbb{p}} - \mathbb{p}\|_{\mathcal{L}_s}^s) = \mathbb{P}_{\mathbb{p}}^{\otimes n}(\lambda(|\widehat{\mathbb{p}} - \mathbb{p}|^s))$, with a focus on the special cases $s = 1$ and $s = 2$. For an introduction to Kernel density estimation we refer to the lecture course Statistik 1 (§22 - §24).

Nonparametric regression. We describe the dependence of the variation of a real-valued random variable Y (response) on the variation of an explanatory random variable X by a functional relationship $\mathbb{E}(Y|X = x) = f(x)$ where f is an unknown functional parameter of interest. For a detailed discussion of the case of a deterministic explanatory variable we refer to Tsybakov [2009]. Here and subsequently, we restrict our attention to the special case of a real-valued explanatory variable X, and hence, a random vector (X, Y) taking values in $(\mathbb{R}^2, \mathscr{B}^2)$. The joint distribution of (X, Y) is uniquely determined by the functional parameter of interest f, the conditional distribution of the error $\xi := Y - f(X)$ given X and the marginal distribution of X which are generally all not known in advance. However, the joint distribution is typically parametrised by the regression function f only and we write shortly $(X, Y) \sim \mathbb{P}_f$. Thereby, the dependence on the marginal distribution $\mathbb{P}^X$ of the regressor X and the conditional distribution of the error term ξ given X is usually not made explicit. For sake of simplicity, suppose in addition that the joint distribution $\mathbb{P}_f$ of (X, Y) admits a joint Lebesgue density $\mathbb{p}$. Denoting by $\mathbb{p}^X$ the marginal density of X we use for the conditional density $\mathbb{p}^{Y|X}$ of Y given X the $\mathbb{P}_f$-a.s.

identity $\mathfrak{p}^X \mathfrak{p}^{Y|X} = \mathfrak{p}$ which allows for $\mathbb{P}_f$-a.e. $x \in \mathbb{R}$ to write

$$\mathfrak{q}(x) := f(x)\,\mathfrak{p}^X(x) = \mathbb{E}\big(Y \big| X = x\big)\,\mathfrak{p}^X(x) = \int_{\mathbb{R}} y\,\mathfrak{p}^{Y|X=x}(y)\mathrm{d}y\,\mathfrak{p}^X(x) = \int_{\mathbb{R}} y\,\mathfrak{p}(x,y)\mathrm{d}y. \quad (09.01)$$

Consequently, the function of interest is $\mathbb{P}_f$-a.s. given by $f = \mathfrak{q}/\mathfrak{p}^X$ which motivates the following estimation strategy. Given a sample of (X,Y) estimate separately $\mathfrak{q}$ and $\mathfrak{p}^X$, say by $\widehat{\mathfrak{q}}$ and $\widehat{\mathfrak{p}}^X$, and then form a estimator $\widehat{f} = \widehat{\mathfrak{q}}/\widehat{\mathfrak{p}}^X$ (possibly in addition to be regularised). There are many different approaches including local smoothing techniques, orthogonal series estimation, penalised smoothing techniques and combinations of them, to name but a few. In the sequel let $\mathcal{F}$ be a family of regression functions and for each $f \in \mathcal{F}$ denote by $\mathbb{P}_f$ and $\mathbb{E}_f$ the associated probability measure of (X,Y) and its expectation, respectively. We denote by $\mathbb{P}_{\mathcal{F}}$ the family of possible distributions of (X,Y), but keep in mind, that the distribution $\mathbb{P}_f$ of (X,Y) is generally not uniquely determined by $f \in \mathcal{F}$ only. If $\big\{(X_i,Y_i) : i \in [\![n]\!]\big\}$ form an independent and identically distributed (i.i.d.) sample of $(X,Y) \sim \mathbb{P}_f$ then $\mathbb{P}_f^{\otimes n} = \otimes_{j\in[\![n]\!]}\mathbb{P}_f$ denotes the joint product probability measure of the family $((X_i,Y_i))_{i\in[\![n]\!]}$. We write $((X_i,Y_i))_{i\in[\![n]\!]} \overset{i.i.d.}{\sim} \mathbb{P}_f$ or $((X_i,Y_i))_{i\in[\![n]\!]} \sim \mathbb{P}_f^{\otimes n}$ for short. We denote by $\mathbb{P}_{\mathcal{F}}^{\otimes n} := (\mathbb{P}_f^{\otimes n})_{f\in\mathcal{F}}$ the corresponding family of product probability measures. For $s \geqslant 1$ we access also the accuracy of an estimator $\widehat{f}$ of f either by a local measure, e.g. $\mathbb{P}_f^{\otimes n}(|\widehat{f}(t) - f(t)|^s)$, for $t \in \mathbb{R}$, or by a global measure, e.g. $\mathbb{P}_f^{\otimes n}(\|\widehat{f} - f\|_{\mathcal{L}_s}^s) = \mathbb{P}_f^{\otimes n}(\lambda(|\widehat{f} - f|^s))$ with a focus on the special cases $s = 1$ and $s = 2$. For an introduction to smoothing techniques we refer to the lecture course Statistik 1 (§22 - §24).

§10 Noisy version of the parameter

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ be a separable real Hilbert spaces. We are interested in the reconstruction of $\theta \in \mathbb{H}$ from a noisy version of it, which we formalise first in this section by introducing stochastic processes.

§10|01 Stochastic process

§10.01 **Notation**. Here and subsequently, a non-empty and generally non-finite subset $\mathcal{J}$ of $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{R}$ and a subset $\mathcal{U}$ of $\mathcal{J}$ denote an index set. We consider the product spaces $\mathbb{R}^{\mathcal{J}} = \bigtimes_{j\in\mathcal{J}}\mathbb{R}$ and $\mathbb{R}^{\mathcal{U}} = \bigtimes_{u\in\mathcal{U}}\mathbb{R}$, where we identify the family $y_\bullet = (y_j)_{j\in\mathcal{J}} \in \mathbb{R}^{\mathcal{J}}$ and the map $y_\bullet : \mathcal{J} \to \mathbb{R}$ with $j \mapsto y_j$. The map $\Pi_u : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{U}}$ given by $y_\bullet = (y_j)_{j\in\mathcal{J}} \mapsto (y_j)_{j\in\mathcal{U}} =: \Pi_u y_\bullet$ is called *canonical projection*. In particular, for each $j \in \mathcal{J}$ the *coordinate map* $\Pi_j := \Pi_{\{j\}} : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}$ is given by $y_\bullet = (y_{j'})_{j'\in\mathcal{J}} \mapsto y_j =: \Pi_j y_\bullet$. Moreover, $\mathbb{R}^{\mathcal{J}}$ is equipped with the product Borel-σ-algebra $\mathscr{B}^{\otimes\mathcal{J}} := \bigotimes_{j\in\mathcal{J}}\mathscr{B}$. Recall that $\mathscr{B}^{\otimes\mathcal{J}}$ equals the smallest σ-algebra on $\mathbb{R}^{\mathcal{J}}$ such that all coordinate maps Π_j, $j \in \mathcal{J}$ are measurable. i.e., $\mathscr{B}^{\otimes\mathcal{J}} = \sigma(\Pi_j, j \in \mathcal{J})$. Moreover, let $(\mathcal{J}, \mathscr{J}, \nu)$ be a measure space with σ-finite $\nu \in \mathcal{M}_\sigma(\mathscr{J})$ and $\mathcal{L}_2(\nu) := \mathcal{L}_2(\mathcal{J}, \mathscr{J}, \nu)$ the usual set of square integrable real-valued functions defined on $(\mathcal{J}, \mathscr{J}, \nu)$. Define the set of equivalence classes $\mathbb{J} := \mathbb{L}_2(\nu) := \mathbb{L}_2(\mathcal{J}, \mathscr{J}, \nu)$, which forms a Hilbert space endowed with usual inner product $\langle\cdot,\cdot\rangle_{\mathbb{J}} := \langle\cdot,\cdot\rangle_{\mathbb{L}_2(\nu)}$ and induced norm $\|\cdot\|_{\mathbb{J}} := \|\cdot\|_{\mathbb{L}_2(\nu)}$. Eventually, we define arithmetic operations on elements of $\mathbb{R}^{\mathcal{J}}$ coordinate-wise, for example meaning $a_\bullet b_\bullet = (a_j b_j)_{j\in\mathcal{J}}$ and $r a_\bullet = (r a_j)_{j\in\mathcal{J}}$ for $a_\bullet, b_\bullet \in \mathbb{R}^{\mathcal{J}}$ and $r \in \mathbb{R}$. Let us further introduce $0_\bullet := (0)_{j\in\mathcal{J}}$ and $1_\bullet := (1)_{j\in\mathcal{J}}$. □

§10.02 **Comment**. Given a measure space $(\Omega, \mathscr{A}, \mu)$, $s \in [1,\infty]$ and the usual space $\mathcal{L}_s(\Omega, \mathscr{A}, \mu)$ of $\mathcal{L}_s(\mu)$-integrable functions introduce for each $\mathscr{A}$-$\mathscr{B}$-measurable $h : \Omega \to \mathbb{R}$, in short $h \in \mathscr{A}$, the

μ-equivalence class $\{h\}_\mu := \{h_o \in \mathscr{A} : h = h_o\ \mu\text{-a.e.}\}$. For $s \in \overline{\mathbb{R}}^+$ define the set of equivalence classes $\mathbb{L}_s(\mu) := \mathbb{L}_s(\mathscr{A},\mu) := \mathbb{L}_s(\Omega,\mathscr{A},\mu) := \{\{h\}_\mu : h \in \mathcal{L}_s(\mathscr{A},\mu)\}$ and $\|\{h\}_\mu\|_{\mathbb{L}_s(\mu)} := \|h\|_{\mathcal{L}_s(\mu)}$ for $\{h\}_\mu \in \mathbb{L}_s(\mu)$. For $s \geqslant 1$, $(\mathbb{L}_s(\mu), \|\cdot\|_{\mathbb{L}_s(\mu)})$ is a normed vector space. Formally, we denote by $\{\bullet\}_\mu : \mathcal{L}_s(\mu) \to \mathbb{L}_s(\mu)$ the natural injection $h \mapsto \{h\}_\mu$. In case $s = 2$ the norm $\|\{h\}_\mu\|_{\mathbb{L}_2(\mu)} := \|h\|_{\mathcal{L}_2(\mu)} = (\mu(|h|^2))^{1/2}$ is induced by the inner product $(\{h\}_\mu, \{h_o\}_\mu) \mapsto \langle\{h\}_\mu, \{h_o\}_\mu\rangle_{\mathbb{L}_2(\mu)} := \mu(hh_o)$, and hence $(\mathbb{L}_2(\mu), \langle\cdot,\cdot\rangle_{\mathbb{L}_2(\mu)})$ is a Hilbert space. As usual we identify the equivalence class $\{h\}_\mu$ with its representative h, and write $h \in \mathbb{L}_2(\mu)$ for short. If $\lambda = \mu$ is the Lebesgue-measure then we write shortly $(\mathbb{L}_2, \langle\cdot,\cdot\rangle_{\mathbb{L}_2})$ and $\{\bullet\} : \mathcal{L}_2 \to \mathbb{L}_2$. □

§10.03 **Stochastic process.** Let $(Y_j)_{j\in\mathcal{J}}$ be a family of real-valued random variables on a common probability space $(\Omega,\mathscr{A},\mathbb{P})$, that is, $Y_j \in \mathscr{A}$ for each $j \in \mathcal{J}$. Consider the $\mathbb{R}^{\mathcal{J}}$-valued random variable $Y_\bullet := (Y_j)_{j\in\mathcal{J}}$ where $Y_\bullet : \Omega \to \mathbb{R}^{\mathcal{J}}$ is a $\mathscr{A}$-$\mathscr{B}^{\otimes\mathcal{J}}$-measurable map given by $\omega \mapsto (Y_j(\omega))_{j\in\mathcal{J}} =: Y_\bullet(\omega)$. $Y_\bullet$ is called a *stochastic process*. Its *distribution* $\mathbb{P}^{Y_\bullet} := \mathbb{P}\circ Y_\bullet^{-1}$ is the image probability measure of $\mathbb{P}$ under the map $Y_\bullet$, i.e. $Y_\bullet \sim \mathbb{P}^{Y_\bullet}$ for short. Further, denote by $\mathbb{P}^{Y_\mathcal{U}} = \mathbb{P}\circ Y_\mathcal{U}^{-1} = \mathbb{P}^{Y_\bullet}\circ\Pi_\mathcal{U}^{-1}$ the distribution of the stochastic process $Y_\mathcal{U} := \Pi_\mathcal{U} Y_\bullet = (Y_u)_{u\in\mathcal{U}}$ on $\mathcal{U} \subseteq \mathcal{J}$. The family $(\mathbb{P}^{Y_\mathcal{U}})_{\mathcal{U}\subseteq\mathcal{J}\text{ finite}}$ is called *family of finite-dimensional distributions* of $Y_\bullet$ or $\mathbb{P}^{Y_\bullet}$. In particular, $\mathbb{P}^{Y_j} = \mathbb{P}^{\Pi_j Y_\bullet} = \mathbb{P}^{Y_\bullet}\circ\Pi_j^{-1}$ denotes the distribution of $Y_j = \Pi_j Y_\bullet$. Furthermore, for $j, j_o \in \mathcal{J}$ we write $\mathbb{P}(Y_j) = \mathbb{P}^{Y_\bullet}(\Pi_j)$ and $\mathbb{C}\mathrm{ov}(Y_j, Y_{j_o}) := \mathbb{P}(Y_jY_{j_o}) - \mathbb{P}(Y_j)\mathbb{P}(Y_{j_o}) = \mathbb{P}^{Y_\bullet}(\Pi_j\Pi_{j_o}) - \mathbb{P}^{Y_\bullet}(\Pi_j)\mathbb{P}^{Y_\bullet}(\Pi_{j_o})$, if it exists, for the expectation of Y_j and the covariance of Y_j and Y_{j_o} with respect to $\mathbb{P}^{Y_\bullet}$. □

§10.04 **Assumption.** The stochastic process $Y_\bullet = (Y_j)_{j\in\mathcal{J}}$ on a common measurable space $(\Omega,\mathscr{A})$ as a function $\Omega\times\mathcal{J} \to \mathbb{R}$ with $(\omega, j) \mapsto Y_j(\omega)$ is $\mathscr{A}\otimes\mathscr{J}$-$\mathscr{B}$-measurable, $Y_\bullet \in \mathscr{A}\otimes\mathscr{J}$ for short. □

§10.05 **Definition.** Let $Y_\bullet = (Y_j)_{j\in\mathcal{J}} \sim \mathbb{P}^{Y_\bullet}$ be a stochastic process satisfying Assumption §10.04. If $\mathbb{P}(|Y_j|) \in \mathbb{R}^+$, i.e. $Y_j \in \mathcal{L}_1(\mathbb{P})$ or $\Pi_j \in \mathcal{L}_1(\mathbb{P}^{Y_\bullet})$ in equal, for each $j \in \mathcal{J}$, then $\mathrm{m}_\bullet := (\mathrm{m}_j := \mathbb{P}(Y_j))_{j\in\mathcal{J}} \in \mathbb{R}^{\mathcal{J}}$ is called *mean function* of $Y_\bullet$ where $\mathrm{m}_\bullet \in \mathscr{J}$ due to Assumption §10.04. If in addition $\nu(\mathrm{m}_\bullet^2) \in \mathbb{R}^+$, i.e. $\mathrm{m}_\bullet \in \mathbb{J}$ then $\mathrm{m}_\bullet$ is called *($\mathbb{J}$-)mean*. If $\mathbb{P}(|Y_j|^2) \in \mathbb{R}^+$, i.e., $Y_j \in \mathcal{L}_2(\mathbb{P})$ or $\Pi_j \in \mathcal{L}_2(\mathbb{P}^{Y_\bullet})$ in equal, for each $j \in \mathcal{J}$, then $\mathrm{cov}_{\bullet,\bullet} = (\mathrm{cov}_{j,j_o} := \mathbb{C}\mathrm{ov}(Y_j, Y_{j_o}))_{j,j_o\in\mathcal{J}} \in \mathbb{R}^{\mathcal{J}^2}$ is called *covariance function* of $Y_\bullet$, where $\mathrm{cov}_{\bullet,\bullet} \in \mathscr{J}^2$ due to Assumption §10.04. A linear and bounded (continuous) operator from $\mathbb{J}$ into itself, $\Gamma \in \mathbb{L}(\mathbb{J})$ for short, satisfying $\langle\Gamma x_\bullet, y_\bullet\rangle_\mathbb{J} = \mathbb{C}\mathrm{ov}(\nu(x_\bullet Y_\bullet), \nu(y_\bullet Y_\bullet)) = \int_\mathcal{J}\int_\mathcal{J} y_j\,\mathrm{cov}_{j,j_o}\,x_{j_o}\nu(dj)\nu(dj_o)$ for all $y_\bullet, x_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu)$ is called *covariance operator* of $Y_\bullet$ or $\mathbb{P}^{Y_\bullet}$. If $Y_\bullet$ admits a mean function $\mathrm{m}_\bullet \in \mathscr{J}$ (respectively mean $\mathrm{m}_\bullet \in \mathbb{J}$) and a covariance function $\mathrm{cov}_{\bullet,\bullet} \in \mathscr{J}^2$ (respectively covariance operator $\Gamma \in \mathbb{L}(\mathbb{J})$) then we write shortly $Y_\bullet \sim \mathrm{P}_{(\mathrm{m}_\bullet,\mathrm{cov}_{\bullet,\bullet})}$ (respectively $Y_\bullet \sim \mathrm{P}_{(\mathrm{m}_\bullet,\Gamma)}$). □

§10.06 **Remark.** A covariance operator $\Gamma \in \mathbb{L}(\mathbb{J})$ associated with a stochastic process $Y_\bullet \sim \mathbb{P}^{Y_\bullet}$ is self-adjoint and non-negative definite, $\Gamma \in \mathbb{L}^\geqslant(\mathbb{J})$ for short. If

$$\sup\left\{\mathbb{P}\left(|\nu(y_\bullet Y_\bullet)|^2\right) : y_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu), \|y_\bullet\|_\mathbb{J} \leqslant 1\right\} \in \mathbb{R}^+,$$

which holds whenever $\mathbb{P}(\|Y_\bullet\|_\mathbb{J}^2) \in \mathbb{R}^+$ or in equal $\|Y_\bullet\|_\mathbb{J} \in \mathcal{L}_2(\mathbb{P})$ (implying $Y_\bullet \in \mathbb{J}$ $\mathbb{P}$-a.s.), then there exists a covariance operator $\Gamma \in \mathbb{L}^\geqslant(\mathbb{J})$ satisfying $\langle\Gamma x_\bullet, y_\bullet\rangle_\mathbb{J} = \mathbb{C}\mathrm{ov}(\nu(x_\bullet Y_\bullet), \nu(y_\bullet Y_\bullet))$. Observe that $\|Y_\bullet\|_\mathbb{J}^2 = \sup\left\{\left|\nu(y_\bullet Y_\bullet)\right|^2 : y_\bullet \in \mathbb{J}, \|y_\bullet\|_\mathbb{J} \leqslant 1\right\}$. Note that $\|Y_\bullet\|_\mathbb{J} \in \mathcal{L}_2(\mathbb{P})$ is a sufficient condition for the existence of a covariance operator, but it is not a necessary condition. □

§10.07 **Empirical mean model.** Assume a probability space $(\mathcal{Z},\mathscr{Z},\mathbb{P})$ and a stochastic process $\psi_\bullet = (\psi_j)_{j\in\mathcal{J}} \in \mathscr{Z}\otimes\mathscr{J}$, i.e. $\mathcal{Z}\times\mathcal{J} \ni (z, j) \mapsto \psi_j(z) \in \mathbb{R}$ is $\mathscr{Z}\otimes\mathscr{J}$-$\mathscr{B}$-measurable, satisfying in addition $\psi_j \in \mathcal{L}_1(\mathbb{P}) := \mathcal{L}_1(\mathcal{Z},\mathscr{Z},\mathbb{P})$ for each $j \in \mathcal{J}$. Consider the product probability

space $(\mathcal{Z}^n, \mathscr{Z}^{\otimes n}, \mathbb{P}^{\otimes n})$ and $Y_\bullet = (Y_j)_{j\in\mathcal{J}}$ with $Y_j := \widehat{\mathbb{P}}_n(\psi_j) \in \mathscr{Z}^{\otimes n}$ where $z^n = (z_i^n)_{i\in[\![n]\!]} \mapsto Y_j(z^n) = (\widehat{\mathbb{P}}_n(\psi_j))(z^n) = n^{-1}\sum_{i\in[\![n]\!]}\psi_j(z_i^n)$ for each $j \in \mathcal{J}$ and $Y_\bullet \in \mathscr{Z}^{\otimes n} \otimes \mathscr{J}$. By construction $\mathrm{m}_\bullet = (\mathrm{m}_j = \mathbb{P}(\psi_j))_{j\in\mathcal{J}} \in \mathscr{J}$ is the mean function of $Y_\bullet$. For each $j \in \mathcal{J}$ the statistic $\varepsilon_j := n^{1/2}(\widehat{\mathbb{P}}_n(\psi_j) - \mathbb{P}(\psi_j)) \in \mathscr{Z}^{\otimes n}$ is centred, i.e. $\varepsilon_j \in \mathcal{L}_1(\mathbb{P}^{\otimes n})$ with $\mathbb{P}^{\otimes n}(\varepsilon_j) = 0$, and $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathcal{J}} \in \mathscr{Z}^{\otimes n} \otimes \mathscr{J}$. Since $Y_j = \mathrm{m}_j + n^{-1/2}\varepsilon_j$ for each $j \in \mathcal{J}$ by construction we write shortly $Y_\bullet = \mathrm{m}_\bullet + n^{-1/2}\varepsilon_\bullet$. If for each $j \in \mathcal{J}$ in addition $\psi_j \in \mathcal{L}_2(\mathbb{P}) := \mathcal{L}_2(\mathcal{Z}, \mathscr{Z}, \mathbb{P})$ then we have $Y_j = \widehat{\mathbb{P}}_n(\psi_j) \in \mathcal{L}_2(\mathbb{P}^{\otimes n})$ and, hence $\varepsilon_j \in \mathcal{L}_2(\mathbb{P}^{\otimes n})$ by construction. The *covariance function* $\mathrm{cov}_{\bullet,\bullet} \in \mathscr{J}^2$ of $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathcal{J}}$ is given for each $j, j_\circ \in \mathcal{J}$ by

$$\mathrm{cov}_{j,j_\circ} = \mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_{j_\circ}) = \mathbb{P}(\psi_j\psi_{j_\circ}) - \mathbb{P}(\psi_j)\mathbb{P}(\psi_{j_\circ}) = n\,\mathbb{C}\mathrm{ov}(Y_j, Y_{j_\circ}).$$

Consequently, we have $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}_{\bullet,\bullet})}$ and $Y_\bullet = \mathrm{m}_\bullet + n^{-1/2}\varepsilon_\bullet \sim \mathrm{P}_{(\mathrm{m}_\bullet, n^{-1}\mathrm{cov}_{\bullet,\bullet})}$. There exists a covariance operator $\Gamma \in \mathbb{L}(\mathbb{J})$, if in addition $\sup\{\mathbb{P}(|\nu(y_\bullet\psi_\bullet)|^2) : y_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu), \|y_\bullet\|_{\mathbb{J}} \leqslant 1\} \in \mathbb{R}^+$, which holds whenever $\|\psi_\bullet\|_{\mathbb{J}} \in \mathcal{L}_2(\mathbb{P})$ or in equal $\mathbb{P}(\|\psi_\bullet\|_{\mathbb{J}}^2) \in \mathbb{R}^+$. Observe that $\|\psi_\bullet\|_{\mathbb{J}}^2 = \sup\{|\nu(y_\bullet\psi_\bullet)|^2 : y_\bullet \in \mathbb{J}, \|y_\bullet\|_{\mathbb{J}} \leqslant 1\}$. Note that $\|\psi_\bullet\|_{\mathbb{J}} \in \mathcal{L}_2(\mathbb{P})$ is a sufficent condition for the existence of a covariance operator, but it is not necessary. □

§10.08 **White noise process.** A stochastic process $\dot{\mathrm{W}}_\bullet = (\dot{\mathrm{W}}_j)_{j\in\mathcal{J}}$ is called *white noise process*, if $(\dot{\mathrm{W}}_j)_{j\in\mathcal{J}}$ is a family of independent and identically distributed random variables, where each $\dot{\mathrm{W}}_j$ has zero mean and variance one, $\dot{\mathrm{W}}_j \sim \mathrm{P}_{(0,1)}$ and $\dot{\mathrm{W}}_\bullet \sim \mathrm{P}_{(0,1)}^{\otimes\mathcal{J}}$ in short. □

§10.09 **Notation.** In other words, the distribution $\mathbb{P}^{\dot{\mathrm{W}}_\bullet}$ of a white noise process $\dot{\mathrm{W}}_\bullet = (\dot{\mathrm{W}}_j)_{j\in\mathcal{J}} \sim \mathbb{P}^{\dot{\mathrm{W}}_\bullet}$ equals the product of its marginal $\mathrm{P}_{(0,1)}$-distributions, i.e. $\mathbb{P}^{\dot{\mathrm{W}}_\bullet} = \otimes_{j\in\mathcal{J}}\mathbb{P}^{\dot{\mathrm{W}}_j} = \otimes_{j\in\mathcal{J}}\mathrm{P}_{(0,1)} = \mathrm{P}_{(0,1)}^{\otimes\mathcal{J}}$. □

§10.10 **Remark.** The centred stochastic process $\varepsilon_\bullet := (\varepsilon_j)_{j\in\mathcal{J}}$ of error terms in an Empirical mean model §10.07 is in general not a white noise process. □

§10.11 **Notation.** We denote by $\ell_2 := \mathbb{L}_2(\nu_{\mathbb{N}}) = \mathbb{L}_2(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}}) = \mathbb{J}$ the space of all square-summable real-valued sequences endowed with counting measure $\nu_{\mathbb{N}} := \sum_{j\in\mathbb{N}}\delta_{\{j\}}$ over the index set $\mathbb{N}$. □

§10.12 **Property.** *Let $\dot{\mathrm{W}}_\bullet := (\dot{\mathrm{W}}_j)_{j\in\mathbb{N}} \sim \mathrm{P}_{(0,1)}^{\otimes\mathbb{N}}$ be a white noise process. By assumption $\dot{\mathrm{W}}_\bullet$ admits $0_\bullet := (0)_{j\in\mathbb{N}}$ as ℓ_2-mean and $\Gamma = \mathrm{id}_{\ell_2} \in \mathbb{L}(\ell_2)$ as covariance operator, i.e. $\dot{\mathrm{W}}_\bullet \sim \mathrm{P}_{(0,\mathrm{id}_{\ell_2})}$, since $\langle x_\bullet, y_\bullet\rangle_{\ell_2} = \sum_{j\in\mathbb{N}} y_j x_j = \sum_{j\in\mathbb{N}} y_j \sum_{j_\circ\in\mathbb{N}} \mathrm{cov}_{j,j_\circ} x_{j_\circ} = \langle \Gamma x_\bullet, y_\bullet\rangle_{\ell_2}$.* □

§10.13 **Gaussian process.** A stochastic process $Y_\bullet = (Y_j)_{j\in\mathcal{J}} \sim \mathrm{P}_{(\mathrm{m}_\bullet,\mathrm{cov}_{\bullet,\bullet})}$ satisfying Assumption §10.04 with mean function $\mathrm{m}_\bullet \in \mathscr{J}$ and covariance function $\mathrm{cov}_{\bullet,\bullet} \in \mathscr{J}^2$ is called a *Gaussian process*, if the family of finite-dimensional distributions $(\mathbb{P}^{Y_u})_{\mathcal{U}\subseteq\mathcal{J}\text{ finite}}$ consists of normal distributions, that is, $Y_u = (Y_u)_{u\in\mathcal{U}}$ is normally distributed with mean vector $(\mathrm{m}_u)_{u\in\mathcal{U}}$ and covariance matrix $(\mathrm{cov}_{u,u'})_{u,u'\in\mathcal{U}}$. We write shortly $Y_\bullet \sim \mathrm{N}_{(\mathrm{m}_\bullet,\mathrm{cov}_{\bullet,\bullet})}$ or $Y_\bullet \sim \mathrm{N}_{(\mathrm{m}_\bullet,\Gamma_{\bullet,\bullet})}$, if in addition there exist a covariance operator $\Gamma_{\bullet,\bullet} \in \mathbb{L}(\mathbb{J})$ associated with $Y_\bullet$. The Gaussian process $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,\mathrm{id}_{\mathbb{J}})}$ with $\mathbb{J}$-mean zero and covariance operator $\mathrm{id}_{\mathbb{J}}$ is called *iso-Gaussian process* or *Gaussian white noise process*, which equals $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ in the particular case $\mathbb{J} = \mathbb{L}_2(\nu_{\mathbb{N}}) = \ell_2$. □

§10.14 **Definition** *(Random function).* Let $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ be an Hilbert space equipped with its Borel-σ-algebra $\mathscr{B}_{\mathbb{H}}$, which is induced by its topology. An $\mathscr{A}$-$\mathscr{B}_{\mathbb{H}}$-measurable map $Y : (\Omega, \mathscr{A}) \to (\mathbb{H}, \mathscr{B}_{\mathbb{H}})$ is called an $\mathbb{H}$-valued random variable or a *random function* in $\mathbb{H}$. □

§10.15 **Lemma.** *Consider $(\ell_2, \langle\cdot,\cdot\rangle_{\ell_2})$. There does not exist a non-zero random function $Y_\bullet = (Y_j)_{j\in\mathbb{N}}$ in ℓ_2 which is a Gaussian white noise process.*

§10.16 **Proof** of Lemma §10.15. Exercise. □

§10|02 Noisy parameter

§10.17 **Assumption**. The Hilbert space $\mathbb{J} = \mathbb{L}_2(\mathcal{T}, \mathscr{J}, \nu)$ and the surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \mathbb{J})$, i.e. $\mathrm{U}\mathrm{U}^\star = \mathrm{id}_{\mathbb{J}}$, are fixed and presumed to be known in advance. □

§10.18 **Notation**. Here and subsequently, we write $\theta_\bullet = (\theta_j)_{j\in\mathcal{J}} := \mathrm{U}\theta \in \mathbb{J}$. Keep in mind, that we identify equivalence classes and their representatives. Our aim is the reconstruction of $\theta_\bullet$ and hence $\mathrm{U}^\star\theta_\bullet \in \mathbb{H}$ from a noisy version of $\theta_\bullet$. □

§10.19 **Noisy parameter**. Let $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathcal{J}}$ be a stochastic process satisfying Assumption §10.04 with mean zero and let $n \in \mathbb{N}$ be a sample size. The stochastic process $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\varepsilon_\bullet$ with $\mathbb{J}$-mean $\theta_\bullet$ is called a *noisy version* of the parameter $\theta_\bullet \in \mathbb{J}$, or *noisy parameter* for short. We denote by $\mathbb{P}_\theta^n$ the distribution of $\widehat{\theta}_\bullet$. If $\varepsilon_\bullet$ admits (possibly depending on $\theta_\bullet$) a covariance function, say $\mathrm{cov}_{\bullet,\bullet} \in \mathscr{J}^2$, or a covariance operator, say $\Gamma \in \mathbb{L}(\mathbb{J})$, then we eventually write $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}_{\bullet,\bullet})}$ and $\widehat{\theta}_\bullet \sim \mathrm{P}_{(\theta,n^{-1}\mathrm{cov}_{\bullet,\bullet})}$ or $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma)}$ and $\widehat{\theta}_\bullet \sim \mathrm{P}_{(\theta,n^{-1}\Gamma)}$ for short. The reconstruction of $\theta_\bullet \in \mathbb{J}$ (or in equal $\mathrm{U}^\star\theta_\bullet \in \mathbb{H}$) from a noisy version $\widehat{\theta}_\bullet \sim \mathbb{P}_\theta^n$ is called a *statistical direct problem*. □

§10.20 **Sequence space model**. Consider $\mathbb{J} = \ell_2 = \mathbb{L}_2(\nu_{\mathbb{N}})$. Let $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathbb{N}}$ be a real-valued stochastic process satisfying Assumption §10.04 with mean $0_\bullet \in \ell_2$ and let $n \in \mathbb{N}$ be a sample size. The observable noisy version $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\varepsilon_\bullet \sim \mathrm{P}_\theta^n$ with ℓ_2-mean $\theta_\bullet \in \ell_2$ as in §10.11 takes the form of a *sequence space model (SSM)*

$$\widehat{\theta}_j = \theta_j + n^{-1/2}\varepsilon_j, \quad j \in \mathbb{N}. \tag{10.01}$$

If $\varepsilon_\bullet$ admits a covariance function (possibly depending on $\theta_\bullet$), say $\mathrm{cov}_{\bullet,\bullet} \in 2^{\mathbb{N}^2}$, then we eventually write $\widehat{\theta}_\bullet \sim \mathrm{P}_{(\theta,n^{-1}\mathrm{cov}_{\bullet,\bullet})}$ for short. If in addition $\varepsilon_\bullet$ admits a covariance operator $\Gamma \in \mathbb{L}(\ell_2)$ (an infinite matrix) then we write $\widehat{\theta}_\bullet \sim \mathrm{P}_{(\theta,n^{-1}\Gamma)}$. □

§10.21 **Gaussian sequence space model**. Let $\dot{\mathrm{B}}_\bullet := (\dot{\mathrm{B}}_j)_{j\in\mathbb{N}} \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ be a Gaussian white noise process. The observable noisy version $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet$ with ℓ_2-mean $\theta_\bullet \in \ell_2$ takes the form of a *Gaussian sequence space model (GSSM)*

$$\widehat{\theta}_j = \theta_j + n^{-1/2}\dot{\mathrm{B}}_j,\ j \in \mathbb{N} \quad \text{with} \quad (\dot{\mathrm{B}}_j)_{j\in\mathbb{N}} \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}} \tag{10.02}$$

and we denote by N_θ^n the distribution of the stochastic process $\widehat{\theta}_\bullet$. □

§10.22 **Notation**. Consider the measure space $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$ where $\lambda_{[0,1]}$ denotes the restriction of the Lebesgue measure to the Borel-σ-algebra $\mathscr{B}_{[0,1]}$ over $[0,1]$, and the Hilbert space $\mathbb{L}_2(\lambda_{[0,1]}) := \mathbb{L}_2([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. Assume that $\theta \in \mathbb{L}_2(\lambda_{[0,1]}) =: \mathbb{H}$. Consider an *orthonormal system* $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$. Then $\mathrm{U} : \mathbb{L}_2(\lambda_{[0,1]}) \to \ell_2$ with $h \mapsto \mathrm{U}h := h_\bullet = (h_j := \langle h, \mathrm{u}_j\rangle_{\mathbb{H}})_{j\in\mathbb{N}}$ is a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$. Its adjoint operator $\mathrm{U}^\star \in \mathbb{L}(\ell_2, \mathbb{L}_2(\lambda_{[0,1]}))$ satisfies $\mathrm{U}^\star a_\bullet = \sum_{j\in\mathbb{N}} a_j \mathrm{u}_j$ for all $a_\bullet \in \ell_2$. We call $h_\bullet = (h_j)_{j\in\mathbb{N}}$ *(generalised) Fourier coefficients* and U *(generalised) Fourier series transform*. □

§10.23 **Nonparametric density estimation on $[0,1]$**. Let $\mathbb{D}_2$ be a set of square-integrable Lebesgue densities on $([0,1], \mathscr{B}_{[0,1]})$, and hence $\mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]}) =: \mathbb{H}$. We denote for each density $\mathfrak{p} \in \mathbb{D}_2$ by $\mathbb{P}_\mathfrak{p} := \mathfrak{p}\lambda_{[0,1]}$ and $\mathbb{E}_\mathfrak{p}$ the associated probability measure and expectation, respectively. Assuming an iid. sample $(X_i)_{i\in[\![n]\!]}$ of size $n \in \mathbb{N}$ we consider the statistical product experiment $\big([0,1]^n, \mathscr{B}_{[0,1]}^{\otimes n}, \mathbb{P}_{\mathbb{D}_2}^{\otimes n} := (\mathbb{P}_\mathfrak{p}^{\otimes n})_{\mathfrak{p}\in\mathbb{D}_2}\big)$. Let $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$ be a generalised Fourier series transform (see Notation §10.22) which is fixed and known in advanced. Evidently, for each $\mathfrak{p} \in \mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ the generalised Fourier coefficients $\mathfrak{p}_\bullet = (\mathfrak{p}_j)_{j\in\mathbb{N}} = \mathrm{U}\mathfrak{p}$ satisfy

$\mathbb{p}_j = \langle \mathbb{p}, \mathrm{u}_j \rangle_{\mathbb{H}} = \lambda_{[0,1]}(\mathbb{p}\,\mathrm{u}_j) = \mathbb{p}\,\lambda_{[0,1]}(\mathrm{u}_j) = \mathbb{P}_{\mathbb{p}}(\mathrm{u}_j)$, i.e. $\mathrm{u}_j \in \mathbb{L}_1([0,1], \mathscr{B}_{[0,1]}, \mathbb{P}_{\mathbb{p}}) =: \mathbb{L}_1(\mathbb{P}_{\mathbb{p}})$, for each $j \in \mathbb{N}$. Moreover, the stochastic process $(\mathrm{u}_j)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]}, \mathbb{P}_{\mathbb{p}})$ is $\mathscr{B}_{[0,1]} \otimes 2^{\mathbb{N}}$-$\mathscr{B}$-measurable. Similar to an Empirical mean model §10.07 we define $\widehat{\mathbb{p}}_\bullet = (\widehat{\mathbb{p}}_j := \widehat{\mathbb{P}}_n(\mathrm{u}_j))_{j\in\mathbb{N}} \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ where $x^n = (x_i^n)_{i\in[\![n]\!]} \mapsto \widehat{\mathbb{p}}_j(x^n) = (\widehat{\mathbb{P}}_n(\mathrm{u}_j))(x^n) = n^{-1}\sum_{i\in[\![n]\!]} \mathrm{u}_j(x_i^n)$ for each $j \in \mathbb{N}$. By construction $\mathbb{p}_\bullet = (\mathbb{p}_j = \mathbb{P}_{\mathbb{p}}(\mathrm{u}_j))_{j\in\mathbb{N}} \in 2^{\mathbb{N}}$ is the mean function of $\widehat{\mathbb{p}}_\bullet$. For each $j \in \mathbb{N}$ the statistic $\varepsilon_j := n^{1/2}(\widehat{\mathbb{P}}_n(\mathrm{u}_j) - \mathbb{P}_{\mathbb{p}}(\mathrm{u}_j)) \in \mathscr{B}_{[0,1]}^{\otimes n}$ is centred, i.e. $\varepsilon_j \in \mathbb{L}_1([0,1]^n, \mathscr{B}_{[0,1]}^{\otimes n}, \mathbb{P}_{\mathbb{p}}^{\otimes n}) =: \mathbb{L}_1(\mathbb{P}_{\mathbb{p}}^{\otimes n})$ with $\mathbb{P}_{\mathbb{p}}^{\otimes n}(\varepsilon_j) = 0$, and $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathbb{N}} \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$. Since $\widehat{\mathbb{p}}_j = \mathbb{p}_j + n^{-1/2}\varepsilon_j$ for each $j \in \mathbb{N}$ by construction $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ is a noisy version of $\mathbb{p}_\bullet$. □

§11 Orthogonal projection

§11.01 **Notation** *(Reminder).* Consider a measure space $(\mathcal{J}, \mathscr{J}, \nu)$ as in Notation §10.01. For $w_\bullet \in \mathbb{R}^{\mathcal{J}}$ define the multiplication map $\mathrm{M}_{w_\bullet} : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{J}}$ with $a_\bullet \mapsto \mathrm{M}_{w_\bullet} a_\bullet := w_\bullet a_\bullet := (w_j a_j)_{j\in\mathcal{J}}$. If $w_\bullet \in \mathscr{J}$, i.e. $w_\bullet$ is $\mathscr{J}$-$\mathscr{B}$-measurable, then we have $\mathrm{M}_{w_\bullet} : \mathscr{J} \to \mathscr{J}$ too. We denote by $\mathbb{M}_{\mathscr{J}}$ the set of all multiplication maps defined on $\mathscr{J}$. If in addition $w_\bullet \in \mathcal{L}_\infty(\nu)$ then we have also $\mathrm{M}_{w_\bullet} : \mathbb{L}_2(\nu) = \mathbb{J} \to \mathbb{J}$ identifying eventually equivalence classes and representatives. We set $\mathbb{L}^{\mathrm{M}}(\mathbb{J}) := \{\mathrm{M}_{w_\bullet} \in \mathbb{M}_{\mathscr{J}} : w_\bullet \in \mathbb{L}_\infty(\nu)\} \subseteq \mathbb{L}(\mathbb{J})$ noting that $\|\mathrm{M}_{w_\bullet}\|_{\mathbb{L}(\mathbb{J})} = \sup\{\|w_\bullet a_\bullet\|_{\mathbb{J}} : \|a_\bullet\|_{\mathbb{J}} \leqslant 1\} \leqslant \|w_\bullet\|_{\mathbb{L}_\infty(\nu)}$ for each $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$. □

§11.02 **Notation**. For $A \in \mathscr{J}$ we denote by $\mathbb{1}_\bullet^A = (\mathbb{1}_j^A)_{j\in\mathcal{J}}$ the indicator function where for each $j \in \mathcal{J}$, $\mathbb{1}_j^A = 1$ if $j \in A$ and $\mathbb{1}_j^A = 0$ otherwise. Obviously, $\mathbb{1}_\bullet^A$ is $\mathscr{J}$-$\mathscr{B}$-measurable, i.e. $\mathbb{1}_\bullet^A \in \mathscr{J}$, and it belongs to $\mathbb{L}_\infty(\nu)$, and to $\mathbb{L}_2(\nu)$ whenever $\nu(A) \in \mathbb{R}^+$. Since $\{j\} \in \mathscr{J}$ we have $\mathbb{1}_\bullet^{\{j\}} \in \mathscr{J}$ and $\mathbb{1}_\bullet^{\{j\}} \in \mathbb{L}_\infty(\nu)$. Obviously, we have $\mathbb{1}_\bullet = \mathbb{1}_\bullet^{\mathcal{J}} \in \mathbb{L}_\infty(\nu)$ and $\mathrm{M}_{\mathbb{1}_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$. For each $w_\bullet \in \mathbb{L}_\infty(\nu)$ set $\mathbb{J}w_\bullet := \{\{a_\bullet w_\bullet\}_\nu : a_\bullet \in \mathcal{L}_2(\nu)\} = \{a_\bullet w_\bullet : a_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu)\}$ and hence in particular $\mathbb{J}\mathbb{1}_\bullet^A = \{a_\bullet \mathbb{1}_\bullet^A : a_\bullet \in \mathbb{J}\}$. Given $0_\bullet = (0)_{j\in\mathcal{J}}$ for $w_\bullet \in \mathscr{J}$ we write further $\mathcal{N}_w := \{w_\bullet = 0_\bullet\} := \{j \in \mathcal{J} : w_j = 0\} \in \mathscr{J}$, and denote by $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \{a_\bullet \in \mathbb{J} : a_\bullet w_\bullet \in \mathbb{J}\}$, $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \{a_\bullet w_\bullet : a_\bullet \in \mathrm{dom}(\mathrm{M}_{w_\bullet}) \subseteq \mathbb{J}\}$ and $\ker(\mathrm{M}_{w_\bullet}) = \{a_\bullet \in \mathbb{J} : \{a_\bullet w_\bullet\}_\nu = 0_\bullet\}$, respectively, the domain, range and nullspace of $\mathrm{M}_{w_\bullet} : \mathbb{J} \supseteq \mathrm{dom}(\mathrm{M}_{w_\bullet}) \to \mathbb{J}$. We write $w_\bullet \in \mathscr{J}_{\setminus 0}$, if $w_\bullet \in \mathscr{J}$ and $\nu(\mathcal{N}_w) = 0$. Similarly, if $w_\bullet \in (\mathbb{R}^+)^{\mathcal{J}}$ is $\mathscr{J}$-$\mathscr{B}^+$-measurable, then we write $w_\bullet \in \mathscr{J}^+$, and $w_\bullet \in \mathscr{J}^+_{\setminus 0}$ assuming additionally $\nu(\mathcal{N}_w) = 0$. □

§11.03 **Property**. *For each $w_\bullet \in \mathscr{J}^+ \cap \mathcal{L}_\infty(\nu)$ the multiplication $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\geqslant}(\mathbb{J}) \subseteq \mathbb{L}(\mathbb{J})$ is a positive semi-definite operator. Keeping $\mathcal{N}_w = \{w_\bullet = 0\} \in \mathscr{J}$ in mind its range and null space is given by $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \mathbb{J}w_\bullet$ and $\ker(\mathrm{M}_{w_\bullet}) = \mathbb{J}\mathbb{1}_\bullet^{\mathcal{N}_w} = \mathrm{ran}(\mathrm{M}_{\mathbb{1}_\bullet^{\mathcal{N}}})$, respectively. $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$ is consequently injective if and only if $w_\bullet \in \mathscr{J}_{\setminus 0}$, i.e. $w_\bullet \in \mathscr{J}$ and $\nu(\mathcal{N}_w) = 0$. For each $A \in \mathscr{J}$ setting $A^c := \mathcal{J}\setminus A \in \mathscr{J}$ the range and null space of $\mathrm{M}_{\mathbb{1}_\bullet^A} \in \mathbb{L}^{\geqslant}(\mathbb{J}) \subseteq \mathbb{L}(\mathbb{J})$ is given by $\mathrm{ran}(\mathrm{M}_{\mathbb{1}_\bullet^A}) = \mathbb{J}\mathbb{1}_\bullet^A$ and $\ker(\mathrm{M}_{\mathbb{1}_\bullet^A}) = \mathbb{J}\mathbb{1}_\bullet^{A^c}$, respectively. Obviously, we have $\mathrm{M}_{\mathbb{1}_\bullet^A}^2 = \mathrm{M}_{\mathbb{1}_\bullet^A}$ and hence $\mathrm{M}_{\mathbb{1}_\bullet^A}$ is an orthogonal projection and $\mathbb{J} = \mathbb{J}\mathbb{1}_\bullet^A \oplus \mathbb{J}\mathbb{1}_\bullet^{A^c}$. Moreover, the map $\mathrm{M}_{\mathbb{1}_\bullet} = \mathrm{id}_{\mathbb{J}}$ equals the identity on $\mathbb{J}$.* □

§11|01 Weigthed norms and inner products

§11.04 **Notation**. Extending the real line by the points $-\infty$ and $+\infty$ we define $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. We denote by $\overline{\mathscr{B}}$ the Borel-σ-field over $\overline{\mathbb{R}}$ and note that the trace of $\overline{\mathscr{B}} \cap \mathbb{R}$ over $\mathbb{R}$ equals $\mathscr{B}$. Thereby, each $a_\bullet \in \mathscr{J}^+$ is in a canonical way also $\mathscr{J}$-$\overline{\mathscr{B}}^+$ measurable, $a_\bullet \in \overline{\mathscr{J}}^+$ for short. For $w_\bullet \in \overline{\mathscr{J}}$ and hence $w_\bullet^2 \in \overline{\mathscr{J}}^+$, consider the measure $w_\bullet^2\nu$ on $(\mathcal{J}, \mathscr{J})$, i.e., $w_\bullet^2 = \mathrm{d}w_\bullet^2\nu/\mathrm{d}\nu$ is the Radon-Nikodym density of $w_\bullet^2\nu$ with respect to ν. We write shortly $\langle\cdot,\cdot\rangle_w := \langle\cdot,\cdot\rangle_{\mathbb{L}_2(w_\bullet^2\nu)}$ and $\|\cdot\|_w := \|\cdot\|_{\mathbb{L}_2(w_\bullet^2\nu)}$. For $w_\bullet \in \mathscr{J}$ we denote its Moore-Penrose inverse by $w_\bullet^\dagger := w_\bullet^{-1}\mathbb{1}_\bullet^{\mathcal{N}_w^c} \in \mathscr{J}$ meaning $w_j^\dagger := w_j^{-1}$ if $j \in \mathcal{N}_w^c$ and $w_j^\dagger := 0$ if $j \in \mathcal{N}_w$. Obviously, we have $w_\bullet^\dagger w_\bullet w_\bullet^\dagger = w_\bullet^\dagger$,

$w_\bullet w_\bullet^\dagger w_\bullet = w_\bullet$ and $w_\bullet w_\bullet^\dagger = w_\bullet^\dagger w_\bullet = \mathbb{1}_\bullet^{\mathcal{N}_w^c}$. We set $\mathbb{J}^w := \mathbb{L}_2^w(\nu) := \mathrm{dom}(\mathrm{M}_{w^\dagger})$ and write $w_\bullet^{2|\dagger} := (w_\bullet^\dagger)^2 = (w_\bullet^2)^\dagger$ for short. □

§11.05 **Property.** *Let $w_\bullet \in \overline{\mathscr{J}}$. Then for each $a_\bullet \in \mathcal{L}_2(w_\bullet^2\nu)$ we have $w_\bullet^2\nu(|a_\bullet|^2) = \nu(|w_\bullet a_\bullet|^2)$. If $w_\bullet \in \mathbb{R}$ ν-a.e., then $w_\bullet^2\nu \in \mathcal{M}_\sigma(\mathscr{J})$ is a σ-finite measure and $\mathbb{L}_2(w_\bullet^2\nu)$ endowed with inner product $\langle\cdot,\cdot\rangle_w = \langle\cdot,\cdot\rangle_{\mathbb{L}_2(w_\bullet^2\nu)} = \langle \mathrm{M}_{w_\bullet}\cdot, \mathrm{M}_{w_\bullet}\cdot\rangle_{\mathbb{L}_2(\nu)}$ is a separable Hilbert space. If in addition $w_\bullet \in \mathcal{L}_\infty(\nu)$, then*

$$\mathcal{L}_2(w_\bullet^{2|\dagger}\nu) = \mathcal{L}_{2(\nu)}w_\bullet + \mathscr{J}\mathbb{1}_\bullet^{\mathcal{N}_w} = \{w_\bullet h_\bullet : h_\bullet \in \mathcal{L}_2(\nu)\} + \{h_\bullet \mathbb{1}_\bullet^{\mathcal{N}_w} : h_\bullet \in \mathscr{J}\}. \tag{11.01}$$

Indeed, for each $h_\bullet \in \mathscr{J}$ consider the decomposition $h_\bullet = w_\bullet w_\bullet^\dagger h_\bullet + h_\bullet \mathbb{1}_\bullet^{\mathcal{N}_w}$. The claim follows immediately from the equivalence of $h_\bullet \in \mathcal{L}_2(w_\bullet^{2|\dagger}\nu)$ and $w_\bullet^\dagger h_\bullet \in \mathcal{L}_2(\nu)$. Under $w_\bullet \in \mathcal{L}_\infty(\nu)$ the map $\mathrm{M}_{w_\bullet} : \mathcal{L}_2(\nu) \to \mathcal{L}_2(\nu)$ is well-defined, and setting $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \{h_\bullet \in \mathcal{L}_2(\nu) : w_\bullet^\dagger h_\bullet \in \mathcal{L}_2(\nu)\} = \mathcal{L}_{2(\nu)}w_\bullet + \mathcal{L}_{2(\nu)}\mathbb{1}_\bullet^{\mathcal{N}_w} \subseteq \mathcal{L}_2(w_\bullet^{2|\dagger}\nu)$ (similar to (11.01)). Consequently, if in addition $\nu(\mathcal{N}_w) = 0$, then $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \mathcal{L}_2(w_\bullet^{2|\dagger}\nu)$. If $w_\bullet \in \mathbb{L}_\infty(\nu)$ then $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$, and $\mathrm{M}_{w_\bullet^\dagger} : \mathbb{J} \supseteq \mathrm{dom}(\mathrm{M}_{w_\bullet^\dagger}) \to \mathbb{J}$. Moreover, we have $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \mathbb{J}$, $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \mathbb{J}w_\bullet$ and $\ker(\mathrm{M}_{w_\bullet}) = \mathbb{J}\mathbb{1}_\bullet^{\mathcal{N}_w}$ (see Property §11.03). Therewith, it follows $\mathrm{dom}(\mathrm{M}_{w_\bullet^\dagger}) = \mathbb{J}w_\bullet \oplus \mathbb{J}\mathbb{1}_\bullet^{\mathcal{N}_w}$. Consequently, if in addition $\nu(\mathcal{N}_w) = 0$, then $\mathbb{J}^w = \mathbb{L}_2^w(\nu) = \mathrm{dom}(\mathrm{M}_{w_\bullet^\dagger}) = \mathbb{J}w_\bullet = \mathbb{L}_2(w_\bullet^{2|\dagger}\nu)$. The last equality follows from (11.01) since both measures $w_\bullet^{2|\dagger}\nu$ and ν share the same null sets (i.e. they mutually dominate each other). □

§11|02 Orthogonal projection

§11.06 **Notation.** For a non-empty and generally non-finite subset $\mathcal{J}$ of $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{R}$ and $m \in \mathbb{N}$ we set $[\![m]\!] := [-m, m] \cap \mathcal{J}$ and we write shortly $\mathbb{1}_\bullet^m = (\mathbb{1}_j^m)_{j\in\mathcal{J}} := \mathbb{1}_\bullet^{[\![m]\!]}$. Furthermore, we define $\mathbb{1}_\bullet^{m|\perp} := \mathbb{1}_\bullet - \mathbb{1}_\bullet^m$. □

§11.07 **Property.** *For each $m \in \mathbb{N}$, $\mathrm{M}_{\mathbb{1}_\bullet^m} \in \mathbb{L}(\mathbb{J})$ and $\mathrm{M}_{\mathbb{1}_\bullet^{m|\perp}} \in \mathbb{L}(\mathbb{J})$ is the orthogonal projection onto the linear subspace $\mathbb{J}\mathbb{1}_\bullet^m \subseteq \mathbb{J}$ and its orthogonal complement $\mathbb{J}\mathbb{1}_\bullet^{m|\perp} = (\mathbb{J}\mathbb{1}_\bullet^m)^\perp \subseteq \mathbb{J}$, respectively, that is $\mathbb{J} = \mathbb{J}\mathbb{1}_\bullet^m \oplus \mathbb{J}\mathbb{1}_\bullet^{m|\perp}$. We have point-wise $\mathbb{1}_\bullet^m - \mathbb{1}_\bullet = \mathrm{o}(1)$ as $m \to \infty$ meaning that for each $j \in \mathcal{J}$ holds $\mathbb{1}_j^m - \mathbb{1}_j = \mathrm{o}(1)$ as $m \to \infty$. Considering the orthogonal projection $\mathrm{M}_{\mathbb{1}_\bullet^m} \in \mathbb{L}(\mathbb{J})$ and the identity $\mathrm{M}_{\mathbb{1}_\bullet} = \mathrm{id}_\mathbb{J} \in \mathbb{L}(\mathbb{J})$ point-wise convergence $\mathrm{M}_{\mathbb{1}_\bullet^m} - \mathrm{id}_\mathbb{J} = \mathrm{o}(1)$ as $m \to \infty$ holds too, that is, $\|(\mathrm{M}_{\mathbb{1}_\bullet^m} - \mathrm{id}_\bullet)a_\bullet\|_\mathbb{J} = \|(\mathbb{1}_\bullet^m - \mathbb{1}_\bullet)a_\bullet\|_\mathbb{J} = \|\mathbb{1}_\bullet^{m|\perp}a_\bullet\|_\mathbb{J} = \mathrm{o}(1)$ as $m \to \infty$ for all $a_\bullet \in \mathbb{J}$.* □

§11.08 **Orthogonal projection.** Given $m \in \mathbb{N}$ we define for each $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ its orthogonal projection $\theta_\bullet^m := \theta_\bullet \mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m$ (and $\theta^m := \mathrm{U}^\star\theta_\bullet^m \in \mathbb{H}$). □

§11|03 Global and maximal global $\mathfrak{v}$-error

We shall measure first globally the accuracy of the orthogonal projection $\theta_\bullet^m := \theta_\bullet \mathbb{1}_\bullet^m$ of $\theta_\bullet \in \mathbb{J}$.

§11.09 **Property.** *If $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$ (i.e. $\nu(\mathcal{N}_\mathfrak{v}) = 0$) and $\theta_\bullet \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ (i.e. $\|\theta_\bullet\|_\mathfrak{v}^2 = \mathfrak{v}_\bullet^2\nu(\theta_\bullet^2) \in \mathbb{R}^+$), then for each $m \in \mathbb{N}$ we have $\theta_\bullet^m \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ too, since $\|\theta_\bullet^m\|_\mathfrak{v}^2 = \mathfrak{v}_\bullet^2\nu(\theta_\bullet^2\mathbb{1}_\bullet^m) \leqslant \mathfrak{v}_\bullet^2\nu(\theta_\bullet^2)$. Moreover, it holds $\|\theta_\bullet^m - \theta_\bullet\|_\mathfrak{v}^2 = \|\theta_\bullet\mathbb{1}_\bullet^{m|\perp}\|_\mathfrak{v}^2 = \mathfrak{v}_\bullet^2\nu(\theta_\bullet^2\mathbb{1}_\bullet^{m|\perp}) \leqslant \mathfrak{v}_\bullet^2\nu(\theta_\bullet^2) \in \mathbb{R}^+$ and $\|\theta_\bullet^m - \theta_\bullet\|_\mathfrak{v}^2 = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence.* □

§11.10 **Comment.** We assume throughout this chapter that the Hilbert space $\mathbb{J} = \mathbb{L}_2(\mathcal{J}, \mathscr{J}, \nu)$ and the *surjective partial isometry* $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \mathbb{J})$ is fixed and known in advance. Considering a $\mathfrak{v}$-error means the weight sequences $\mathfrak{v}_\bullet \in \mathscr{J}$ is also fixed and known in advance. Consequently, the condition $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$ does not impose an additional restriction. □

§11.11 **Global $\mathfrak{v}$-error.** Given $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$, $m \in \mathbb{N}$, a parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ and its orthogonal projection $\theta_\bullet^m = \theta_\bullet \mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m$ we call $\|\theta_\bullet^m - \theta_\bullet\|_\mathfrak{v} = \|\theta_\bullet\mathbb{1}_\bullet^{m|\perp}\|_\mathfrak{v} \in \mathbb{R}^+$ *global $\mathfrak{v}$-error*. □

§11.12 **Assumption**. Consider weights $\mathfrak{a}_\bullet, \mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$, i.e. $\nu(\mathcal{N}_\mathfrak{a}) = 0 = \nu(\mathcal{N}_\mathfrak{v})$, such that $\mathfrak{a}_\bullet \in \mathbb{L}_\infty(\nu)$ and $(\mathfrak{a}\mathfrak{v})_\bullet := (\mathfrak{a}_j\mathfrak{v}_j)_{j\in\mathcal{J}} = \mathfrak{a}_\bullet\mathfrak{v}_\bullet \in \mathbb{L}_\infty(\nu)$. We write $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathbb{L}_\infty(\nu)} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. □

§11.13 **Reminder**. Under Assumption §11.12 we have $\mathbb{J}^\mathfrak{a} = \mathbb{L}^\mathfrak{a}_2(\nu) = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}'}) = \mathbb{J}\mathfrak{a}_\bullet = \mathbb{L}_2(\mathfrak{a}_\bullet^{2|\dagger}\nu)$ and the three measures ν, $\mathfrak{a}_\bullet^{2|\dagger}\nu$ and $\mathfrak{v}_\bullet^2\nu$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). Consequently, $\mathbb{J}^\mathfrak{a} \subseteq \mathbb{J} = \mathbb{L}_2(\nu)$ and if $h_\bullet \in \mathbb{L}_2(\mathfrak{a}_\bullet^{2|\dagger}\nu)$ satisfies $\mathfrak{v}_\bullet^2\nu(h_\bullet^2) \in \mathbb{R}^+$, for example, then $h_\bullet \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ too. □

§11.14 **Notation**. Under Assumption §11.12 and given a constant $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ we consider $\mathbb{J}^\mathfrak{a} = \mathbb{L}^\mathfrak{a}_2(\nu) = \mathbb{L}_2(\mathfrak{a}_\bullet^{2|\dagger}\nu)$ endowed with $\|\cdot\|_{\mathfrak{a}^\dagger} := \|\cdot\|_{\mathbb{J}^\mathfrak{a}} := \|\cdot\|_{\mathbb{L}_2(\mathfrak{a}_\bullet^{2|\dagger}\nu)}$ and the ellipsoid

$$\mathbb{J}^{\mathfrak{a},\mathrm{r}} := \{h_\bullet \in \mathbb{J}^\mathfrak{a} : \|h_\bullet\|^2_{\mathfrak{a}^\dagger} = \mathfrak{a}_\bullet^{2|\dagger}\nu(h_\bullet^2) = \nu(\mathfrak{a}_\bullet^{2|\dagger}h_\bullet^2) \leqslant \mathrm{r}^2\} \subseteq \mathbb{J}^\mathfrak{a}.$$

Keep in mind that $(\mathfrak{a}\mathfrak{v})_\bullet \in \mathbb{L}_\infty(\nu)$ implies $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathbb{L}_\infty(\nu)} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. □

§11.15 **Property**. *Under Assumption §11.12 we have* $\mathbb{J}^{\mathfrak{a}\cdot} \subseteq \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$. *Indeed, for each* $h_\bullet \in \mathbb{J}^\mathfrak{a}$, *i.e.,* $\|h_\bullet\|_{\mathfrak{a}^\dagger_\bullet} \in \mathbb{R}^+$, *follows* $\|h_\bullet\|^2_\mathfrak{v} = \nu(h_\bullet^2\mathfrak{a}_\bullet^{2|\dagger}(\mathfrak{a}\mathfrak{v})_\bullet^2) \leqslant \|h_\bullet\|^2_{\mathfrak{a}^\dagger}\|(\mathfrak{a}\mathfrak{v})_\bullet\|^2_{\mathbb{L}_\infty(\nu)} \in \mathbb{R}^+$. □

§11.16 **Abstract smoothness condition**. Under Assumption §11.12 the parameter $\theta_\bullet \in \mathbb{J}$ satisfies an *abstract smoothness condition* if there is $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ such that $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}} \subseteq \mathbb{J}^\mathfrak{a}$. □

§11.17 **Lemma**. *Under Assumption §11.12 for each* $m \in \mathbb{N}$ *the orthogonal projection* $\theta_\bullet^m := \theta_\bullet\mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m$ *of* $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}} \subseteq \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ *satisfies* $\|\theta_\bullet^m - \theta_\bullet\|_\mathfrak{v} = \|\theta_\bullet\mathbb{1}_\bullet^{m|\perp}\|_\mathfrak{v} \leqslant \mathrm{r}\,(\mathfrak{a}\mathfrak{v})_{(m)}$.

§11.18 **Proof** of Lemma §11.17. is given in the lecture. □

§11.19 **Maximal global $\mathfrak{v}$-error**. Under Assumption §11.12 for $m \in \mathbb{N}$, a parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ and its orthogonal projection $\theta_\bullet^m = \theta_\bullet\mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m$ we call $\sup\{\|\theta_\bullet^m - \theta_\bullet\|_\mathfrak{v} : \theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\}$ *maximal global $\mathfrak{v}$-error* over the class of parameters $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$. □

§11|04 Local and maximal local ϕ-error

Secondly, we measure locally the accuracy of the orthogonal projection $\theta_\bullet^m := \theta_\bullet\mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m$ of $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$.

§11.20 **Notation**. For $\phi_\bullet \in \mathscr{J}$ and $\mathrm{dom}(\phi\nu) := \{h_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu) : \phi_\bullet h_\bullet \in \mathbb{L}_1(\nu)\}$ we consider the linear functional $\phi\nu : \mathbb{J} \supseteq \mathrm{dom}(\phi\nu) \to \mathbb{R}$ given by $h_\bullet \mapsto \phi\nu(h_\bullet) := \nu(\phi_\bullet h_\bullet)$ with a slight abuse of notations. □

§11.21 **Comment**. If $\phi_\bullet \in \mathbb{J} = \mathbb{L}_2(\nu)$, then it follows $\mathrm{dom}(\phi\nu) = \mathbb{J}$ and $\|\phi\nu\|_{\mathbb{L}(\mathbb{J},\mathbb{R})} = \|\phi_\bullet\|_\mathbb{J} \in \mathbb{R}^+$. Consequently, we have $\phi\nu \in \mathbb{L}(\mathbb{J},\mathbb{R})$ and $\phi\nu(h_\bullet) = \langle h_\bullet, \phi_\bullet\rangle_\mathbb{J}$, in other words $\phi_\bullet$ is a Fréchet-Riesz representative of the continuous linear functional $\phi\nu$. □

§11.22 **Property**. *If* $\phi_\bullet \in \mathscr{J}_{\setminus 0}$ *(i.e.* $\nu(\mathcal{N}_\phi) = 0$*) and* $\theta_\bullet \in \mathrm{dom}(\phi\nu)$ *(i.e.* $\theta_\bullet\phi_\bullet \in \mathbb{L}_1(\nu)$*), then for each* $m \in \mathbb{N}$ *we have* $\theta_\bullet^m \in \mathrm{dom}(\phi\nu)$ *too, since* $\|\phi_\bullet\theta_\bullet^m\|_{\mathbb{L}_1(\nu)} = \nu(|\phi_\bullet\theta_\bullet|\mathbb{1}_\bullet^m) \leqslant \nu(|\phi_\bullet\theta_\bullet|)$. *Moreover, it holds* $|\phi\nu(\theta_\bullet) - \phi\nu(\theta_\bullet^m)| \leqslant |\phi_\bullet|\nu(|\theta_\bullet^m - \theta_\bullet|) = |\phi_\bullet|\nu(|\theta_\bullet|\mathbb{1}_\bullet^{m|\perp}) \leqslant \nu(|\phi_\bullet\theta_\bullet|) \in \mathbb{R}^+$ *and* $|\phi\nu(\theta_\bullet) - \phi\nu(\theta_\bullet^m)| = \mathrm{o}(1)$ *as* $m \to \infty$ *by dominated convergence.* □

§11.23 **Comment**. We assume throughout this chapter that the Hilbert space $\mathbb{J} = \mathbb{L}_2(\mathcal{J}, \mathscr{J}, \nu)$ and the *surjective partial isometry* $\mathrm{U} \in \mathbb{L}(\mathbb{H},\mathbb{J})$ is fixed and known in advance. Considering a ϕ-error means the linear function $\phi\nu$ and hence in equal $\phi_\bullet \in \mathscr{J}$ is also fixed and known in advance. Consequently, the condition $\phi_\bullet \in \mathscr{J}_{\setminus 0}$ does not impose an additional restriction. □

§11.24 **Local ϕ-error.** Given $\phi_\bullet \in \mathscr{J}_{\setminus 0}$, $m \in \mathbb{N}$, a parameter $\theta_\bullet = \mathrm{U}\theta \in \mathrm{dom}(\phi\nu)$ and its orthogonal projection $\theta^m_\bullet = \theta_\bullet \mathbb{1}^m_\bullet \in \mathbb{J}\mathbb{1}^m_\bullet$ we call $|\phi\nu(\theta_\bullet) - \phi\nu(\theta^m_\bullet)| = |\phi\nu(\theta_\bullet \mathbb{1}^{m|\perp}_\bullet)| \in \mathbb{R}^+$ *local ϕ-error*. □

§11.25 **Assumption.** Consider $\phi_\bullet, \mathfrak{a}_\bullet \in \mathscr{J}_{\setminus 0}$, i.e. $\nu(\mathcal{N}_\phi) = 0 = \nu(\mathcal{N}_\mathfrak{a})$, such that $\mathfrak{a}_\bullet \in \mathbb{L}_\infty(\nu)$ and $(\mathfrak{a}\phi)_\bullet := (\mathfrak{a}_j \phi_j)_{j\in\mathcal{J}} = \mathfrak{a}_\bullet \phi_\bullet \in \mathbb{L}_2(\nu)$ and hence $\|\mathfrak{a}_\bullet \mathbb{1}^{m|\perp}_\bullet\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\mathbb{L}_2(\nu)} = \mathrm{o}(1)$ as $m \to \infty$. □

§11.26 **Reminder.** Under Assumption §11.25 we have $\mathbb{J}^\mathfrak{a} = \mathbb{L}^\mathfrak{a}_2(\nu) = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}^\dagger}) = \mathbb{J}\mathfrak{a}_\bullet = \mathbb{L}_2(\mathfrak{a}^{2|\dagger}_\bullet \nu)$ and the three measures ν, $|\phi_\bullet|\nu$ and $\mathfrak{a}^{2|\dagger}_\bullet \nu$ dominate mutually each other (see Property §11.05). Consequently, $\mathbb{J}^\mathfrak{a} \subseteq \mathbb{J} = \mathbb{L}_2(\nu)$ and if $h_\bullet \in \mathbb{L}_2(\mathfrak{a}^{2|\dagger}_\bullet \nu)$ satisfies $\nu(|\phi_\bullet h_\bullet|) \in \mathbb{R}^+$, for example, then $h_\bullet \in \mathbb{L}_1(|\phi_\bullet|\nu)$ too. □

§11.27 **Property.** *Under Assumption §11.25 we have $\mathbb{J}^\mathfrak{a} \subseteq \mathrm{dom}(\phi\nu)$. Indeed, for each $h_\bullet \in \mathbb{J}^\mathfrak{a}$, i.e. $\|h_\bullet\|_{\mathfrak{a}^\dagger} \in \mathbb{R}^+$, we have $\|\phi_\bullet h_\bullet\|_{\mathbb{L}_1(\nu)} = \nu(|h_\bullet \mathfrak{a}^\dagger_\bullet(\mathfrak{a}\phi)_\bullet|) \leqslant \|h_\bullet\|_{\mathfrak{a}^\dagger} \|(\mathfrak{a}\phi)_\bullet\|_{\mathbb{L}_2(\nu)} \in \mathbb{R}^+$.* □

§11.28 **Notation** *(Reminder).* Under Assumption §11.25 the parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ satisfies an abstract smoothness condition if there is $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ such that $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}} = \{h_\bullet \in \mathbb{J}^\mathfrak{a} : \|h_\bullet\|^2_{\mathfrak{a}^\dagger} \leqslant \mathrm{r}^2\} \subseteq \mathbb{J}^\mathfrak{a}$ where $\|\cdot\|_{\mathfrak{a}^\dagger} = \|\cdot\|_{\mathbb{J}^\mathfrak{a}} := \|\cdot\|_{\mathbb{L}_2(\mathfrak{a}^{2|\dagger}_\bullet \nu)}$ (see Definition §11.16). Since $(\mathfrak{a}\phi)_\bullet \in \mathbb{L}_2(\nu)$ we have $\|\mathfrak{a}_\bullet \mathbb{1}^{m|\perp}_\bullet\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\mathbb{L}_2(\nu)} = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence. □

§11.29 **Lemma.** *Under Assumption §11.25 for each $m \in \mathbb{N}$ the orthogonal projection $\theta^m_\bullet := \theta_\bullet \mathbb{1}^m_\bullet \in \mathbb{J}\mathbb{1}^m_\bullet$ of $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}} \subseteq \mathrm{dom}(\phi\nu)$ satisfies $|\phi\nu(\theta_\bullet - \theta^m_\bullet)| = |\phi\nu(\theta_\bullet \mathbb{1}^{m|\perp}_\bullet)| \leqslant \nu(|\phi_\bullet \theta_\bullet| \mathbb{1}^{m|\perp}_\bullet) \leqslant \mathrm{r}\, \|\mathfrak{a}_\bullet \mathbb{1}^{m|\perp}_\bullet\|_\phi$.*

§11.30 **Proof** of Lemma §11.29. is given in the lecture. □

§11.31 **Maximal local ϕ-error.** Under Assumption §11.25 for $m \in \mathbb{N}$, a parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ and its orthogonal projection $\theta^m_\bullet = \theta_\bullet \mathbb{1}^m_\bullet \in \mathbb{J}\mathbb{1}^m_\bullet$ we call $\sup\{|\phi\nu(\theta_\bullet) - \phi\nu(\theta^m_\bullet)| : \theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\}$ *maximal local ϕ-error* over the class of parameters $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$. □

§12 Orthogonal projection estimator

§12.01 **Notation** *(Reminder).* Consider a measure space $(\mathcal{J}, \mathscr{J}, \nu)$ as in Notation §10.01. For $w_\bullet \in \mathbb{R}^\mathcal{J}$ define the multiplication map $\mathrm{M}_{w_\bullet} : \mathbb{R}^\mathcal{J} \to \mathbb{R}^\mathcal{J}$ with $a_\bullet \mapsto \mathrm{M}_{w_\bullet} a_\bullet := w_\bullet a_\bullet$. For $w_\bullet \in \mathscr{J}$ we have $\mathrm{M}_{w_\bullet} : \mathscr{J} \to \mathscr{J}$ too. We denote by $\mathbb{M}_\mathscr{J}$ the set of all multiplication maps defined on $\mathscr{J}$. If in addition $w_\bullet \in \mathcal{L}_\infty(\nu)$ then we have also $\mathrm{M}_{w_\bullet} : \mathbb{L}_2(\nu) = \mathbb{J} \to \mathbb{J}$ identifying eventually equivalence classes and representatives. We set $\mathbb{L}^{\mathrm{M}}(\mathbb{J}) := \mathbb{M}_\mathbb{J} := \{\mathrm{M}_{w_\bullet} \in \mathbb{M}_\mathscr{J} : w_\bullet \in \mathbb{L}_\infty(\nu)\} \subseteq \mathbb{L}(\mathbb{J})$ noting that $\|\mathrm{M}_{w_\bullet}\|_{\mathbb{L}(\mathbb{J})} = \sup\{\|w_\bullet a_\bullet\|_\mathbb{J} : \|a_\bullet\|_\mathbb{J} \leqslant 1\} \leqslant \|w_\bullet\|_{\mathbb{L}_\infty(\nu)}$ for each $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$. □

§12.02 **Reminder.** If $w_\bullet \in \mathbb{L}_\infty(\nu)$ then $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\mathbb{J})$, and $\mathrm{M}_{w^\dagger_\bullet} : \mathbb{J} \supseteq \mathrm{dom}(\mathrm{M}_{w^\dagger_\bullet}) \to \mathbb{J}$. Moreover, we have $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \mathbb{J}$, $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \mathbb{J}w_\bullet$ and $\ker(\mathrm{M}_{w_\bullet}) = \mathbb{J}\mathbb{1}^{\mathcal{N}_w}_\bullet$ (see Property §11.03), and $\mathrm{dom}(\mathrm{M}_{w^\dagger_\bullet}) = \mathbb{J}w_\bullet \oplus \mathbb{J}\mathbb{1}^{\mathcal{N}_w}_\bullet$ (see Property §11.05). Consequently, if in addition $\nu(\mathcal{N}_w) = 0$, then $\mathbb{J}^w = \mathbb{L}^w_2(\nu) = \mathrm{dom}(\mathrm{M}_{w^\dagger_\bullet}) = \mathbb{J}w_\bullet = \mathbb{L}_2(w^{2|\dagger}_\bullet \nu)$. For each $m \in \mathbb{N}$ we write $\mathbb{1}^m_\bullet = (\mathbb{1}^m_j)_{j\in\mathcal{J}} := \mathbb{1}^{[\![m]\!]}_\bullet$ and $\mathbb{1}^{m|\perp}_\bullet := \mathbb{1}_\bullet - \mathbb{1}^m_\bullet$ with $[\![m]\!] := [-m, m] \cap \mathcal{J}$. Consequently, $\mathrm{M}_{\mathbb{1}^m_\bullet} \in \mathbb{L}^{\geqslant}(\mathbb{J})$ and $\mathrm{M}_{\mathbb{1}^{m|\perp}_\bullet} \in \mathbb{L}^{\geqslant}(\mathbb{J})$ is the *orthogonal projection* onto the linear subspace $\mathbb{J}\mathbb{1}^m_\bullet \subseteq \mathbb{J}$ and its orthogonal complement $\mathbb{J}\mathbb{1}^{m|\perp}_\bullet = (\mathbb{J}\mathbb{1}^m_\bullet)^\perp \subseteq \mathbb{J}$, respectively, that is $\mathbb{J} = \mathbb{J}\mathbb{1}^m_\bullet \oplus \mathbb{J}\mathbb{1}^{m|\perp}_\bullet$ (see Property §11.07). Finally, given $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ we consider the orthogonal projections $\theta^m_\bullet = \theta_\bullet \mathbb{1}^m_\bullet \in \mathbb{J}\mathbb{1}^m_\bullet$ (and $\theta^m := \mathrm{U}^\star \theta^m_\bullet \in \mathbb{H}$) (Definition §11.08). □

§12.03 **Notation** *(Reminder).* Consider a centred stochastic processes $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathcal{J}}$ satisfying Assumption §10.04 and let $n \in \mathbb{N}$ be a sample size. The observable noisy version $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\varepsilon_\bullet$ of the parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ takes the form of a *statistical direct problem* (see Definition §10.19). We

denote by $\mathbb{P}_{\theta}^{n}$ the distribution of $\widehat{\theta}_{\bullet}$. We write $\varepsilon_{\bullet} \sim \mathrm{P}_{(0,\Gamma)}$ if $\varepsilon_{\bullet}$ admits a covariance operator $\Gamma \in \mathbb{L}(\mathbb{J})$ possibly depending on $\theta_{\bullet}$. □

§12.04 **Definition**. Given a noisy version $\widehat{\theta}_{\bullet} \sim \mathbb{P}_{\theta}^{n}$ of the parameter $\theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}$ for each $m \in \mathbb{N}$ we call $\widehat{\theta}_{\bullet}^{m} := \widehat{\theta}_{\bullet}\mathbb{1}_{\bullet}^{m}$ *orthogonal projection estimator (OPE)* of $\theta_{\bullet}$. □

§12.05 **GSSM (§10.21 continued)**. Considering $\ell_2 = \mathbb{L}_2(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ we illustrate the OPE in a Gaussian sequence space model §10.21. Here the observable stochastic process $\widehat{\theta}_{\bullet} = \theta_{\bullet} + n^{-1/2}\dot{\mathrm{B}}_{\bullet}$ is a noisy version of $\theta_{\bullet} = \mathrm{U}\theta \in \ell_2$ and $\dot{\mathrm{B}}_{\bullet} \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Consequently, $\widehat{\theta}_{\bullet}$ admits a N_{θ}^{n}-distribution belonging to the family $\mathrm{N}_{\Theta}^{n} := (\mathrm{N}_{\theta}^{n})_{\theta\in\Theta}$. Summarising the observations satisfy a statistical product experiment $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\otimes\mathbb{N}}, \mathrm{N}_{\Theta}^{n})$ where $\Theta \subseteq \ell_2$. □

§12|01 Global and maximal global $\mathfrak{v}$-risk

We measure first the accuracy of the OPE $\widehat{\theta}_{\bullet}^{m} = \widehat{\theta}_{\bullet}\mathbb{1}_{\bullet}^{m}$ of $\theta_{\bullet}^{m} = \theta_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{J}\mathbb{1}_{\bullet}^{m}$ with $\theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}$ by a global mean-$\mathfrak{v}$-error, i.e. $\mathfrak{v}$-risk.

§12.06 **Reminder**. If $\mathfrak{v}_{\bullet} \in \mathscr{J}_{\setminus 0}$ and $\theta_{\bullet} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ then we have $\theta_{\bullet}^{m} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ too and $\|\theta_{\bullet}^{m} - \theta_{\bullet}\|_{\mathfrak{v}}^2 = \mathrm{o}(1)$ as $m \to \infty$ (Property §11.09). □

§12.07 **Assumption**. Consider a noisy version $\widehat{\theta}_{\bullet} = \theta_{\bullet} + n^{-1/2}\varepsilon_{\bullet} \sim \mathbb{P}_{\theta}^{n}$ of $\theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}$ satisfying Assumption §10.04, $\mathbb{v}_{\bullet}^{\theta} := \mathbb{P}_{\theta}^{n}(\varepsilon_{\bullet}^2) := (\mathbb{P}_{\theta}^{n}(\varepsilon_{j}^2))_{j\in\mathcal{J}} \in \mathbb{L}_{\infty}(\nu)$ and $\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_{\infty}(\nu)$ $\mathbb{P}_{\theta}^{n}$-a.s., for each $m \in \mathbb{N}$. □

§12.08 **Comment**. Under Assumption §12.07 if $\mathfrak{v}_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_2(\nu)$ then we have $\mathfrak{v}_{\bullet}\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_2(\nu)$ $\mathbb{P}_{\theta}^{n}$-a.s.. If in addition $\theta_{\bullet} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$, and hence $\theta_{\bullet}^{m} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ (Property §11.09), then it follows

$$\mathfrak{v}_{\bullet}\widehat{\theta}_{\bullet}^{m} = n^{-1/2}\mathfrak{v}_{\bullet}\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m} + \mathfrak{v}_{\bullet}\theta_{\bullet}^{m} \in \mathbb{L}_2(\nu) \quad \mathbb{P}_{\theta}^{n}\text{-a.s.}. \tag{12.01}$$

If $\mathcal{J} \subseteq \mathbb{Z}$ (at most countable) then Assumption §10.04 and $\mathbb{v}_{\bullet}^{\theta} = \mathbb{P}_{\theta}^{n}(\varepsilon_{\bullet}^2) \in \mathbb{L}_{\infty}(\nu)$ implies the additional assumption $\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_{\infty}(\nu)$ $\mathbb{P}_{\theta}^{n}$-a.s.. However, the last implication does generally not hold, if $\mathcal{J} \in \{\mathbb{R}, \mathbb{R}^{+}\}$ for example. □

§12|01|01 Global $\mathfrak{v}$-risk

§12.09 **Definition**. Under Assumption §12.07, $\mathfrak{v}_{\bullet} \in \mathscr{J}_{\setminus 0}$, $\theta_{\bullet} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ and $\mathfrak{v}_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{J}$ for $m \in \mathbb{N}$ the *global $\mathfrak{v}$-risk* of an OPE $\widehat{\theta}_{\bullet}^{m} = \widehat{\theta}_{\bullet}\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ $\mathbb{P}_{\theta}^{n}$-a.s. satisfies

$$\mathbb{P}_{\theta}^{n}(\|\widehat{\theta}_{\bullet}^{m} - \theta_{\bullet}\|_{\mathfrak{v}}^2) = \mathbb{P}_{\theta}^{n}(\|(\widehat{\theta}_{\bullet} - \theta_{\bullet})\mathbb{1}_{\bullet}^{m}\|_{\mathfrak{v}}^2) + \|\theta_{\bullet}\mathbb{1}_{\bullet}^{m|\perp}\|_{\mathfrak{v}}^2 \tag{12.02}$$

with *variance* term $\mathbb{P}_{\theta}^{n}(\|(\widehat{\theta}_{\bullet} - \theta_{\bullet})\mathbb{1}_{\bullet}^{m}\|_{\mathfrak{v}}^2) = n^{-1}\mathbb{P}_{\theta}^{n}(\|\mathfrak{v}_{\bullet}\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m}\|_{\mathbb{J}}^2)$ and *bias* term $\|\theta_{\bullet}\mathbb{1}_{\bullet}^{m|\perp}\|_{\mathfrak{v}_{\bullet}}$. □

§12.10 **Property**. *Under Assumption §12.07, $\mathfrak{v}_{\bullet} \in \mathscr{J}_{\setminus 0}$ and $\mathbb{1}_{\bullet}^{m} \in \mathbb{L}_2(\mathfrak{v}_{\bullet}^2\nu)$ for $m \in \mathbb{N}$ we have*

$$\mathbb{P}_{\theta}^{n}\big(\|\mathfrak{v}_{\bullet}\varepsilon_{\bullet}\mathbb{1}_{\bullet}^{m}\|_{\mathbb{J}}^2\big) = \int_{\mathcal{J}} \mathbb{P}_{\theta}^{n}(\varepsilon_{j}^2)\mathfrak{v}_{j}^2\mathbb{1}_{j}^{m}\nu(dj) = \nu(\mathbb{v}_{\bullet}^{\theta}\mathfrak{v}_{\bullet}^2\mathbb{1}_{\bullet}^{m}) \tag{12.03}$$

and consequently $\mathbb{P}_{\theta}^{n}(\|(\widehat{\theta}_{\bullet} - \theta_{\bullet})\mathbb{1}_{\bullet}^{m}\|_{\mathfrak{v}_{\bullet}}^2) \leqslant n^{-1}\|\mathbb{v}_{\bullet}^{\theta}\|_{\mathbb{L}_{\infty}(\nu)}\|\mathbb{1}_{\bullet}^{m}\|_{\mathfrak{v}_{\bullet}}^2 \in \mathbb{R}^{+}$. □

§12.11 **Notation**. For $a_{\bullet} \in (\mathbb{R})^{\mathbb{N}}$ with minimal value in $B \subseteq \mathbb{N}$ we define

$$\arg\min\big\{a_m : m \in B\big\} := \min\big\{m \in B : a_m \leqslant a_j,\ \forall j \in B\big\}. \qquad \square$$

§12.12 **Proposition** *(Upper bound).* *Let Assumption §12.07,* $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ *and* $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ *for all* $m \in \mathbb{N}$ *be satisfied. For all* $n, m \in \mathbb{N}$ *setting*

$$\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) := \|\theta_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\circ := \arg\min\left\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\right\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\circ}(\theta_\bullet, \mathfrak{v}_\bullet) = \min\left\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\right\} \quad (12.04)$$

we have $\mathbb{P}_\theta^n(\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2) \leqslant (1 \vee \|\mathbb{v}_\bullet^\theta\|_{\mathbb{L}_\infty(\nu)})\, \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet)$.

§12.13 **Proof** of Proposition §12.12. is given in the lecture. □

§12.14 **Definition**. Let $\theta_\bullet \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ and $\widehat{\theta}_\bullet^m \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ $\mathbb{P}_\theta^n$-a.s. for all $m \in \mathbb{N}$. If there exist $\mathrm{C} \in \mathbb{R}_{\setminus 0}^+$ and for each $n \in \mathbb{N}$, $\mathrm{R}_n^\circ \in \mathbb{R}_{\setminus 0}^+$ and $m_n^\circ \in \mathbb{N}$ satisfying

$$\mathrm{C}^{-1}\,\mathrm{R}_n^\circ \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2 \leqslant \mathbb{P}_\theta^n\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2 \leqslant \mathrm{C}\,\mathrm{R}_n^\circ,$$

then we call R_n° *oracle bound*, m_n° *oracle dimension* and $\widehat{\theta}_\bullet^{m_n^\circ}$ *oracle optimal* (up to the constant C). As a consequence, up to the constant C^2 the statistic $\widehat{\theta}_\bullet^{m_n^\circ}$ attains the lower global $\mathfrak{v}$-risk bound within the family of OPE's, that is, $\mathbb{P}_\theta^n\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2 \leqslant \mathrm{C}^2 \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2$. □

§12.15 **Oracle inequality**. *Under Assumption §12.07 let* $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ *and* $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\mathfrak{v}_\bullet^2\nu)$ *for all* $m \in \mathbb{N}$. *If in addition* $1 \leqslant \max(\|\mathbb{v}_\bullet^\theta\|_{\mathbb{L}_\infty(\nu)}, \|(\mathbb{v}_\bullet^\theta)^{-1}\|_{\mathbb{L}_\infty(\nu)}) \leqslant \mathbb{v}_\theta \in \mathbb{R}_{\setminus 0}^+$ *then* (12.04) *implies*

$$\mathbb{v}_\theta^{-1}\,\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{P}_\theta^n(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2) = n^{-1}\nu(\mathbb{v}_\bullet^\theta \mathfrak{v}_\bullet^2 \mathbb{1}_\bullet^m) + \|\theta_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2$$
$$\leqslant \mathbb{v}_\theta \mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) \quad \textit{for all } m, n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality

$$\mathbb{v}_\theta^{-1}\,\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2) \leqslant \mathbb{P}_\theta^n(\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2)$$
$$\leqslant \mathbb{v}_\theta\, \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{v}_\theta^2 \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2), \quad (12.05)$$

and, hence $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet)$, m_n° *and the statistic* $\widehat{\theta}_\bullet^{m_n^\circ}$, *respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant* $\mathbb{v}_\theta^2$*).* □

§12.16 **Remark**. We shall emphasise that for each fixed $m \in \mathbb{N}$ with $\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}} \in \mathbb{R}^+$ we have $n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}} = \mathrm{o}(1)$ as $n \to \infty$. As a consequence, if $\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}} \in \mathbb{R}^+$ for all $m \in \mathbb{N}$ and $\|\theta_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}} = \mathrm{o}(1)$ as $m \to \infty$ then we obtain $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) = \mathrm{o}(1)$ as $n \to \infty$, and thus, $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet)$ is also called an *oracle rate*. Indeed, for all $\delta \in \mathbb{R}_{\setminus 0}^+$ there exists $m_\delta \in \mathbb{N}$ and $n_\delta \in \mathbb{N}$ such that we have both $\|\theta_\bullet \mathbb{1}_\bullet^{m_\delta|\perp}\|_{\mathfrak{v}}^2 \leqslant \delta/2$ and $n^{-1}\|\mathbb{1}_\bullet^{m_\delta}\|_{\mathfrak{v}}^2 \leqslant \delta/2$ for all $n \geqslant n_\delta$, and whence $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) \leqslant \mathrm{R}_n^{m_\delta}(\theta_\bullet, \mathfrak{v}_\bullet) \leqslant \delta$. However, note that the oracle dimension $m_n^\circ = m_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet)$ as defined in Proposition §12.12 depends on the unknown parameter of interest $\theta_\bullet$, and thus also the oracle optimal statistic $\widehat{\theta}_\bullet^{m_n^\circ}$. In other words $\widehat{\theta}_\bullet^{m_n^\circ}$ is not a feasible estimator. □

§12.17 **Corollary** *(GSSM §12.05 continued).* *Let* $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\mathfrak{q}}^n$ *as in Model §12.05, where* $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ *and* $\theta_\bullet = \mathrm{U}\theta \in \ell_2$. *For* $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *and* $\theta_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ *the (infeasible) OPE* $\widehat{\theta}_\bullet^{m_n^\circ} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{m_n^\circ} \in \ell_2\mathbb{1}_\bullet^{m_n^\circ} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ *with oracle dimension* m_n° *as in* (12.04) *satisfies*

$$\mathrm{N}_{\mathfrak{q}}^n\left(\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2\right) = \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) = \inf_{m\in\mathbb{N}} \mathrm{N}_{\mathfrak{q}}^n\left(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2\right),$$

and hence it is oracle optimal (with constant 1*).*

§12.18 **Proof** of Corollary §12.17. is given in the lecture. □

§12.19 **Illustration**. Here and subsequently, we use for two sequences $a_\bullet, b_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ the notation $a_n \simeq b_n$ if the sequence $a_\bullet/b_\bullet$ is bounded away both from zero and infinity. We illustrate the last results considering usual behaviour for the bias and variance term. We distinguish the following two cases

(p) $\mathfrak{v}_\bullet \in \mathbb{J}$ or there is $m \in \mathbb{N}$ with $\|\theta^m_\bullet - \theta_\bullet\|^2_{\mathfrak{v}} = 0$,

(np) $\mathfrak{v}_\bullet \notin \mathbb{J}$ and for all $m \in \mathbb{N}$ holds $\|\theta^m_\bullet - \theta_\bullet\|^2_{\mathfrak{v}} \in \mathbb{R}^+_{\setminus 0}$.

Interestingly, in case **(p)** the oracle bound is parametric, that is, $n\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet) = \mathrm{O}(1)$, in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet) = \infty$. In case **(np)** consider the following two specifications:

Table 01 [§12]

Order of the oracle rate $\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$ $\mathfrak{v}_j^2 = j^{2v}$	$(a \in \mathbb{R}^+_{\setminus 0})$ θ_j^2	(squarred bias) $\|\theta_\bullet \mathbb{1}^{m\vert\perp}_\bullet\|^2_{\mathfrak{v}}$	(variance) $\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}$	m°_n	$\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet)$
(o) $v \in (-1/2, a)$	j^{-2a-1}	$m^{-2(a-v)}$	m^{2v+1}	$n^{\frac{1}{2a+1}}$	$n^{-\frac{2(a-v)}{2a+1}}$
$v = -1/2$	j^{-2a-1}	m^{-2a-1}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a+1}}$	$\frac{\log n}{n}$
(s) $v + 1/2 \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+} e^{-m^{2a}}$	m^{2v+1}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{2v+1}{2a}}}{n}$
$v = -1/2$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 01 [§12] the order of the oracle rate $\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet)$ is depict for $v \geqslant -1/2$ only. In case $v < -1/2$ the oracle rate $\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet)$ is parametric. □

§12|01|02 Maximal global $\mathfrak{v}$-risk

§12.20 **Reminder**. Under Assumption §11.12 we have $\mathbb{J}^{\mathfrak{a}} = \mathbb{L}_2(\nu) = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}^\dagger}) = \mathbb{J}_{\mathfrak{a}_\bullet} \subseteq \mathbb{J}$ and the three measures ν, $\mathfrak{a}^{2|\dagger}_\bullet \nu$ and $\mathfrak{v}^2_\bullet \nu$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\mathbb{J}^{\mathfrak{a}}$ endowed with $\|\cdot\|_{\mathfrak{a}^\dagger} = \|\mathrm{M}_{\mathfrak{a}^\dagger}\cdot\|_{\mathbb{J}}$ and given a constant $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ the ellipsoid $\mathbb{J}^{\mathfrak{a},\mathrm{r}} := \{h_\bullet \in \mathbb{J}^{\mathfrak{a}} : \|h_\bullet\|_{\mathfrak{a}^\dagger} \leqslant \mathrm{r}\} \subseteq \mathbb{J}^{\mathfrak{a}}$. Since $(\mathfrak{a}\mathfrak{v})_\bullet \in \mathbb{L}_\infty(\nu)$, and hence $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\mathbb{L}_\infty(\nu)} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ we have $\mathbb{J}^{\mathfrak{a}} \subseteq \mathbb{L}_2(\mathfrak{v}^2_\bullet \nu)$ (Property §11.15), and $\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\mathfrak{v}} \leqslant \mathrm{r}\,(\mathfrak{a}\mathfrak{v})_{(m)}$ for all $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.17). □

§12.21 **Proposition**. *Let the Assumptions §12.07 and §11.12 and $\mathbb{1}^m_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet \nu)$ for all $m \in \mathbb{N}$ be satisfied. For all $n, m \in \mathbb{N}$ setting*

$$\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := [(\mathfrak{a}\mathfrak{v})^2_{(m)} \vee n^{-1}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}], \quad m^\star_n := \arg\min\{\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\}$$
$$\text{and} \quad \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}^{m^\star_n}_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\{\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\} \quad (12.06)$$

we have $\mathbb{P}^n_\theta(\|\widehat{\theta}^{m^\star_n}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}) \leqslant (\|\mathbb{v}^\theta_\bullet\|_{\mathbb{L}_\infty(\nu)} + \mathrm{r}^2)\,\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ for all $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ and $n \in \mathbb{N}$.

§12.22 **Proof** of Proposition §12.21. is given in the lecture. □

§12.23 **Remark**. Under the assumptions of Proposition §12.21 if there exists in addition $\mathbb{v} \in \mathbb{R}^+$ satisfying $\|\mathbb{v}^\theta_\bullet\|_{\mathbb{L}_\infty(\nu)} \leqslant \mathbb{v}$ for all $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ then

$$\sup\{\mathbb{P}^n_\theta(\|\widehat{\theta}^{m^\star_n}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}) : \theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\} \leqslant (\mathbb{v} + \mathrm{r}^2)\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \quad \text{for all } n \in \mathbb{N}.$$

Arguing similarly as in Remark §12.16 we note that $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)=\mathrm{o}(1)$ as $n\to\infty$, whenever $\|\mathbb{1}^m_\bullet\|_{\mathfrak{v}_\bullet}\in\mathbb{R}^+$ for all $m\in\mathbb{N}$ and $(\mathfrak{a}\mathfrak{v})_{(m)}=\mathrm{o}(1)$ as $m\to\infty$. The latter is satisfied, for example, if $(\mathfrak{a}\mathfrak{v})_\bullet=\mathfrak{a}_\bullet\mathfrak{v}_\bullet\in\mathbb{J}$ (in equal $\mathfrak{a}_\bullet\in\mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$). Note that the dimension $m^\star_n:=m^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ as defined in (12.06) does not depend on the unknown parameter of interest $\theta_\bullet$ but on the class $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$ only, and thus also the statistic $\widehat{\theta}^{m^\star_n}_\bullet$. In other words, if the regularity of $\theta_\bullet$ is known in advance, then the OPE $\widehat{\theta}^{m^\star_n}_\bullet$ is a feasible estimator. □

§12.24 **Corollary** *(GSSM §12.05 continued). Let* $\widehat{\theta}_\bullet=\theta_\bullet+n^{-1/2}\dot{\mathrm{B}}_\bullet\sim\mathrm{N}^n_{\mathfrak{a}}$ *as in Model §12.05, where* $\dot{\mathrm{B}}_\bullet\sim\mathrm{N}^{\otimes\mathbb{N}}_{(0,1)}$ *and* $\theta_\bullet=\mathrm{U}\theta\in\ell_2$. *Under Assumption §11.12 the OPE* $\widehat{\theta}^{m^\star_n}_\bullet=\widehat{\theta}_\bullet\mathbb{1}^{m^\star_n}_\bullet\in\ell_2\mathbb{1}^{m^\star_n}_\bullet\subseteq\ell_2(\mathfrak{v}^2_\bullet)$ *with dimension* $m^\star_n$ *as in* (12.06) *satisfies*

$$\sup\big\{\mathrm{N}^n_{\mathfrak{a}}\big(\|\widehat{\theta}^{m^\star_n}_\bullet-\theta_\bullet\|^2_{\mathfrak{v}}\big):\theta_\bullet\in\ell^{\mathfrak{a},\mathrm{r}}_2\big\}\leqslant\mathrm{C}\,\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)\quad\textit{for all } n\in\mathbb{N}\tag{12.07}$$

with constant $\mathrm{C}=1+\mathrm{r}^2$.

§12.25 **Proof** of Corollary §12.24. is given in the lecture. □

§12.26 **Illustration**. We illustrate the last results considering usual behaviour for $(\mathfrak{a}\mathfrak{v})_\bullet,\mathfrak{v}_\bullet\in\mathscr{J}$. We distinguish the following two cases **(p)** $\mathfrak{v}_\bullet\in\mathbb{J}$, and **(np)** $\mathfrak{v}_\bullet\notin\mathbb{J}$. Interestingly, in case **(p)** the bound in Proposition §12.21 is parametric, that is, $n\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)=\mathrm{O}(1)$, in case **(np)** the bound is nonparametric, i.e. $\lim_{n\to\infty}n\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)=\infty$. In case **(np)** consider the following two specifications:

Table 02 [§12]

Order of the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ as $n\to\infty$

$(j\in\mathbb{N})$ $\mathfrak{v}_j=j^v$	$(a\in\mathbb{R}^+_{\setminus 0})$ $\mathfrak{a}^2_j$	(squarred bias) $(\mathfrak{a}\mathfrak{v})^2_{(m)}$	(variance) $\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}$	$m^\star_n$	$\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$
(o) $v\in(-1/2,a)$	j^{-2a}	$m^{-2(a-v)}$	m^{2v+1}	$n^{\frac{1}{2a+1}}$	$n^{-\frac{2(a-v)}{2a+1}}$
$v=-1/2$	j^{-2a}	m^{-2a-1}	$\log m$	$\big(\frac{n}{\log n}\big)^{\frac{1}{2a+1}}$	$\frac{\log n}{n}$
(s) $v+1/2\in\mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{2v}e^{-m^{2a}}$	m^{2v+1}	$(\log n)^{\frac{1}{2a}}$	$n^{-1}(\log n)^{\frac{2v+1}{2a}}$
$v=-1/2$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 02 [§12] the order of the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ is depict for $\geqslant-1/2$ only. In case $v<-1/2$ the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ is parametric. □

§12|02 Local and maximal local ϕ-risk

We measure secondly the accuracy of the OPE $\widehat{\theta}^m_\bullet=\widehat{\theta}_\bullet\mathbb{1}^m_\bullet$ of $\theta^m_\bullet=\theta_\bullet\mathbb{1}^m_\bullet\in\mathbb{J}\mathbb{1}^m_\bullet$ with $\theta_\bullet=\mathrm{U}\theta\in\mathbb{J}$ by a local mean-ϕ-error, i.e. ϕ-risk.

§12.27 **Reminder**. If $\phi_\bullet\in\mathscr{J}_{\setminus 0}$ and $\theta_\bullet\in\mathrm{dom}(\phi\nu):=\{h_\bullet\in\mathbb{J}=\mathbb{L}_2(\nu):\phi_\bullet h_\bullet\in\mathbb{L}_1(\nu)\}$ then we have $|\phi\nu(\theta_\bullet)-\phi\nu(\theta^m_\bullet)|=\mathrm{o}(1)$ as $m\to\infty$ (Property §11.22). □

§12.28 **Assumption**. Consider a noisy version $\widehat{\theta}_\bullet=\theta_\bullet+n^{-1/2}\varepsilon_\bullet\sim\mathbb{P}^n_\theta$ of $\theta_\bullet=\mathrm{U}\theta\in\mathbb{J}$ satisfying Assumption §10.04, $\varepsilon_\bullet\sim\mathrm{P}_{(0,\Gamma_\theta)}$ with $\Gamma_\theta\in\mathbb{L}^{\geqslant}(\mathbb{J})$ and $\varepsilon_\bullet\mathbb{1}^m_\bullet\in\mathbb{L}_2(\nu)$ $\mathbb{P}^n_\theta$-a.s. for each $m\in\mathbb{N}$. □

§12.29 **Comment**. Under Assumption §12.28 if $\mathbb{1}^m_\bullet\in\mathbb{L}_2(\phi^2_\bullet\nu)$ then $\mathbb{P}^n_\theta$-a.s. we have $|\nu(|\phi_\bullet\varepsilon_\bullet\mathbb{1}^m_\bullet|)|^2\leqslant\nu(\phi^2_\bullet\mathbb{1}^m_\bullet)\nu(\varepsilon^2_\bullet\mathbb{1}^m_\bullet)\in\mathbb{R}^+$ and hence $\varepsilon_\bullet\mathbb{1}^m_\bullet\in\mathrm{dom}(\phi\nu)$. If in addition $\theta_\bullet\in\mathrm{dom}(\phi\nu)$, and hence $\theta^m_\bullet\in\mathrm{dom}(\phi\nu)$ (Property §11.22), then it follows

$$n^{-1/2}\varepsilon_\bullet\mathbb{1}^m_\bullet+\theta^m_\bullet=\widehat{\theta}^m_\bullet\in\mathrm{dom}(\phi\nu)\quad\mathbb{P}^n_\theta\text{-a.s.}.\tag{12.08}$$

If $\mathcal{J} \subseteq \mathbb{Z}$ (at most countable) then Assumption §10.04 and $\Gamma_\vartheta \in \mathbb{L}(\mathbb{J})$ implies $\mathbb{v}_\bullet^\vartheta = \mathbb{P}_\vartheta^n(\varepsilon_\bullet^2) \in \mathbb{L}_\infty(\nu)$ and hence the additional assumption $\varepsilon_\bullet \mathbb{1}_\bullet^m \in \mathbb{L}_2(\nu)$ $\mathbb{P}_\vartheta^n$-a.s.. However, the last implication does generally not hold, if $\mathcal{J} \in \{\mathbb{R}, \mathbb{R}^+\}$ for example. □

§12|02|01 Local ϕ-risk

§12.30 **Definition**. Under Assumption §12.28, $\phi_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathrm{dom}(\phi\nu)$ and $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\phi_\bullet^2\nu)$ for $m \in \mathbb{N}$ the *local ϕ-risk* of an OPE $\widehat{\theta}_\bullet^m = \widehat{\theta}_\bullet \mathbb{1}_\bullet^m \in \mathrm{dom}(\phi\nu)$ $\mathbb{P}_\vartheta^n$-a.s. satisfies

$$\mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2) = \mathbb{P}_\vartheta^n(|\phi_\bullet\nu((\widehat{\theta}_\bullet - \theta_\bullet)\mathbb{1}_\bullet^m)|^2) + |\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|^2. \tag{12.09}$$

with *variance* $\mathbb{P}_\vartheta^n(|\phi\nu((\widehat{\theta}_\bullet - \theta_\bullet)\mathbb{1}_\bullet^m)|^2) = n^{-1}\mathrm{P}_{(\vartheta,\Gamma_\vartheta)}(|\phi\nu(\varepsilon_\bullet \mathbb{1}_\bullet^m)|^2)$ and *bias* $|\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|$. □

§12.31 **Property**. *Under Assumption §12.28, $\phi_\bullet \in \mathscr{J}_{\setminus 0}$ and $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\phi_\bullet^2\nu)$ for $m \in \mathbb{N}$ we have*

$$\mathrm{P}_{(\vartheta,\Gamma_\vartheta)}(|\phi\nu(\varepsilon_\bullet \mathbb{1}_\bullet^m)|^2) = \langle \Gamma_\vartheta(\phi_\bullet \mathbb{1}_\bullet^m), \phi_\bullet \mathbb{1}_\bullet^m\rangle_{\mathbb{J}} =: \|\phi_\bullet \mathbb{1}_\bullet^m\|_{\Gamma_\vartheta}^2 \leqslant \|\Gamma_\vartheta\|_{\mathbb{L}(\mathbb{J})}\|\mathbb{1}_\bullet^m\|_\phi^2 \tag{12.10}$$

and consequently $\mathbb{P}_\vartheta^n(|\nu(\phi_\bullet(\widehat{\theta}_\bullet - \theta_\bullet)\mathbb{1}_\bullet^m)|^2) \leqslant n^{-1}\|\Gamma_\vartheta\|_{\mathbb{L}(\mathbb{J})}\|\mathbb{1}_\bullet^m\|_\phi^2 \in \mathbb{R}^+$. □

§12.32 **Proposition** *(Upper bound). Let Assumption §12.28, $\phi_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathrm{dom}(\phi\nu)$ and $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\phi_\bullet^2\nu)$ for all $m \in \mathbb{N}$ be satisfied. For all $m, n \in \mathbb{N}$ setting*

$$\mathrm{R}_n^m(\theta_\bullet, \phi_\bullet) := |\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\circ := \arg\min\left\{\mathrm{R}_n^m(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\right\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(\theta_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\circ}(\theta_\bullet, \phi_\bullet) := \min\left\{\mathrm{R}_n^m(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\right\} \tag{12.11}$$

we have $\mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet)|^2) \leqslant (1 \vee \|\Gamma_\vartheta\|_{\mathbb{L}(\mathbb{J})})\mathrm{R}_n^\circ(\theta_\bullet, \phi_\bullet)$.

§12.33 **Proof** of Proposition §12.32. is given in the lecture. □

§12.34 **Definition**. Let $\theta_\bullet \in \mathrm{dom}(\phi_\bullet\nu)$ and $\widehat{\theta}_\bullet^m \in \mathrm{dom}(\phi\nu)$ $\mathbb{P}_\vartheta^n$-a.s. for all $m \in \mathbb{N}$. If there exist $\mathrm{C} \in \mathbb{R}_{\setminus 0}^+$ and for each $n \in \mathbb{N}$, $\mathrm{R}_n^\circ \in \mathbb{R}_{\setminus 0}^+$ and $m_n^\circ \in \mathbb{N}$ satisfying

$$\mathrm{C}^{-1}\,\mathrm{R}_n^\circ \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2) \leqslant \mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet)|^2) \leqslant \mathrm{C}\,\mathrm{R}_n^\circ,$$

then we call R_n° *oracle bound*, m_n° *oracle dimension* and $\widehat{\theta}_\bullet^{m_n^\circ}$ *oracle optimal* (up to the constant C). As a consequence, up to the constant C^2 the statistik $\widehat{\theta}_\bullet^{m_n^\circ}$ attains the lower local ϕ-risk bound within the family of OPE's, that is, $\mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet)|^2) \leqslant \mathrm{C}^2 \inf_{m\in\mathbb{N}} \mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2)$. □

§12.35 **Comment**. If $\Gamma_\vartheta \in \mathbb{L}(\mathbb{J})$ is invertible with inverse $\Gamma_\vartheta^{-1} \in \mathbb{L}(\mathbb{J})$, i.e. $\Gamma_\vartheta\Gamma_\vartheta^{-1} = \mathrm{id}_{\mathbb{J}} = \Gamma_\vartheta^{-1}\Gamma_\vartheta$, then we write shortly $\mathbb{v}_\vartheta := \max(\|\Gamma_\vartheta\|_{\mathbb{L}(\mathbb{J})}, \|\Gamma_\vartheta^{-1}\|_{\mathbb{L}(\mathbb{J})}) \in \mathbb{R}_{\setminus 0}^+$. In this situation for all $a_\bullet \in \mathbb{J}$ we have $\mathbb{v}_\vartheta^{-1}\|a_\bullet\|_{\mathbb{J}}^2 \leqslant \|a_\bullet\|_{\Gamma_\vartheta}^2 = \langle \Gamma_\vartheta a_\bullet, a_\bullet\rangle_{\mathbb{J}} \leqslant \mathbb{v}_\vartheta\|a_\bullet\|_{\mathbb{J}}^2$. □

§12.36 **Oracle inequality**. *Under Assumption §12.28 let $\phi_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathrm{dom}(\phi\nu)$ and $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\phi_\bullet^2\nu)$ for all $m \in \mathbb{N}$. If in addition $1 \leqslant \max(\|\Gamma_\vartheta\|_{\mathbb{L}(\mathbb{J})}, \|\Gamma_\vartheta^{-1}\|_{\mathbb{L}(\mathbb{J})}) \leqslant \mathbb{v}_\vartheta \in \mathbb{R}_{\setminus 0}^+$. then (12.11) and Comment §12.35 imply*

$$\mathbb{v}_\vartheta^{-1}\mathrm{R}_n^m(\theta_\bullet, \phi_\bullet) \leqslant \mathbb{P}_\vartheta^n(|\phi\nu(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2) = n^{-1}\|\phi_\bullet \mathbb{1}_\bullet^m\|_{\Gamma_\vartheta}^2 + |\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|^2$$
$$\leqslant \mathbb{v}_\vartheta \mathrm{R}_n^m(\theta_\bullet, \phi_\bullet) \quad \textit{for all } m, n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality

$$\mathbb{v}_\theta^{-1}\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n(|\phi\nu(\widehat{\theta}_\bullet^m-\theta_\bullet)|^2) \leqslant \mathbb{P}_\theta^n(|\phi\nu(\widehat{\theta}_\bullet^{m_n^\circ}-\theta_\bullet)|^2)$$
$$\leqslant \mathbb{v}_\theta \mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) \leqslant \mathbb{v}_\theta^2 \inf_{m\in\mathbb{N}} \mathbb{P}_\theta^n(|\phi\nu(\widehat{\theta}_\bullet^m-\theta_\bullet)|^2), \quad (12.12)$$

and hence, $\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet)$, m_n° and the statistic $\widehat{\theta}_\bullet^{m_n^\circ}$, respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant $\mathbb{v}_\theta^2$). □

§12.37 **Remark**. Arguing similarly as in Remark §12.16 we note that $\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) = \mathrm{o}(1)$ as $n\to\infty$, whenever $\|\mathbb{1}_\bullet^m\|_\phi^2 \in \mathbb{R}^+$ for all $m\in\mathbb{N}$ and $|\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})| = \mathrm{o}(1)$ as $m\to\infty$. The latter is satisfied, for example, if $\theta_\bullet \in \mathrm{dom}(\phi\nu)$. The oracle dimension $m_n^\circ = m_n^\circ(\theta_\bullet,\phi_\bullet)$ as defined in (12.11) depends again on the unknown parameter of interest $\theta_\bullet$, and thus also the oracle optimal statistic $\widehat{\theta}_\bullet^{m_n^\circ}$. In other words $\widehat{\theta}_\bullet^{m_n^\circ}$ is not a feasible estimator. □

§12.38 **Corollary** *(GSSM §12.05 continued). Let $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\theta_\bullet}^n$ as in Model §12.05, where $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ and $\theta_\bullet = \mathrm{U}\theta \in \ell_2$. For $\theta_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ the (infeasible) OPE $\widehat{\theta}_\bullet^{m_n^\circ} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{m_n^\circ} \in \ell_2 \mathbb{1}_\bullet^{m_n^\circ} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with oracle dimension m_n° as in (12.11) satisfies*

$$\mathrm{N}_{\theta_\bullet}^n(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^{m_n^\circ}-\theta_\bullet)|^2) = \mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) = \inf_{m\in\mathbb{N}} \mathrm{N}_{\theta_\bullet}^n(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^m-\theta_\bullet)|^2),$$

and hence it is oracle optimal (with constant 1).

§12.39 **Proof** of Corollary §12.38. is given in the lecture. □

§12.40 **Illustration**. We illustrate the last results considering usual behaviour for both the variance and the bias term. Similar to the two cases **(p)** and **(np)** in Illustration §12.19 we distinguish here the following two cases

(p) $\phi_\bullet \in \mathbb{J}$ or there is $K\in\mathbb{N}$ with $\sup\{|\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m\in\mathbb{N}\cap[K,\infty)\} = 0$,

(np) $\phi_\bullet \notin \mathbb{J}$ and for all $m\in\mathbb{N}$ holds $\sup\{|\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m\in\mathbb{N}\cap[K,\infty)\} \in \mathbb{R}_{\setminus 0}^+$.

In case **(p)** the oracle bound is again parametric, i.e. $n\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) = \mathrm{O}(1)$, while in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet) = \infty$. In case **(np)** consider the following two specifications

Table 03 [§12]

Order of the oracle rate $\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet)$ as $n\to\infty$

$(j\in\mathbb{N})$ $\phi_\bullet = j^{v-1/2}$	$(a\in\mathbb{R}_{\setminus 0}^+)$ θ_j	(squarred bias) $\lvert\phi\nu(\theta_\bullet\mathbb{1}_\bullet^{m\lvert\perp})\rvert^2$	(variance) $\lVert\mathbb{1}_\bullet^m\rVert_\phi^2$	m_n°	$\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet)$
(o) $v\in(0,a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$
$v=0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	$\left(\frac{n}{\log n}\right)^{\frac{1}{2a}}$	$\frac{\log n}{n}$
(s) $v\in\mathbb{R}_{\setminus 0}^+$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+}e^{-2m^{2a}}$	m^{2v}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$
$v=0$	$e^{-j^{2a}}$	$m^{(1-2a)_+}e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 03 [§12] the order of the oracle rate $\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet)$ is depict for $v\geqslant 0$ only. For $v<0$ the oracle rate $\mathrm{R}_n^\circ(\theta_\bullet,\phi_\bullet)$ is parametric. □

§12|02|02 Maximal local ϕ-risk

§12.41 **Reminder**. Under Assumption §11.25 we have $\mathbb{J}^{\mathfrak{a}} = \mathbb{L}_2^{\mathfrak{a}}(\nu) = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}^\dagger}) = \mathbb{J}_{\mathfrak{a}_\bullet} \subseteq \mathbb{J}$ and the three measures ν, $\mathfrak{a}_\bullet^{2|\dagger}\nu$ and $|\phi_\bullet|\nu$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\mathbb{J}^{\mathfrak{a}}$ endowed with $\|\cdot\|_{\mathfrak{a}^\dagger} = \|\mathrm{M}_{\mathfrak{a}^\dagger}\cdot\|_{\mathbb{J}}$ and given a constant $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ the ellipsoid $\mathbb{J}^{\mathfrak{a},\mathrm{r}} := \{h_\bullet \in \mathbb{J}^{\mathfrak{a}} : \|h_\bullet\|_{\mathfrak{a}^\dagger} \leqslant \mathrm{r}\} \subseteq \mathbb{J}^{\mathfrak{a}}$. Since $(\mathfrak{a}\phi)_\bullet \in \mathbb{J}$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathbb{J}} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ ($\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence) we have $\mathbb{J}^{\mathfrak{a}} \subseteq \mathrm{dom}(\phi\nu)$ (Property §11.27), and $|\phi\nu(\theta_\bullet \mathbb{1}_\bullet^{m|\perp})| \leqslant \mathrm{r}\,\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi$ for all $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.29). □

§12.42 **Proposition**. *Let the Assumptions §11.25 and §12.28, and $\mathbb{1}_\bullet^m \in \mathbb{L}_2(\phi_\bullet^2\nu)$ for all $m \in \mathbb{N}$ be satisfied. For all $n, m \in \mathbb{N}$ setting*

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\star := \arg\min\left\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\right\}$$
$$\textit{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\left\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\right\} \quad (12.13)$$

we have $\mathbb{P}_\theta^n(|\phi\nu(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2) \leqslant (\|\Gamma_\theta\|_{\mathbb{L}(\mathbb{J})} \vee \mathrm{r}^2)\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ for all $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ and $n \in \mathbb{N}$.

§12.43 **Proof** of Proposition §12.42. is given in the lecture. □

§12.44 **Remark**. Under the assumptions of Proposition §12.42 if there exists in addition $\mathbb{v} \in \mathbb{R}^+$ satisfying $\|\Gamma_\theta\|_{\mathbb{L}(\mathbb{J})} \leqslant \mathbb{v}$ for all $\theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ then

$$\sup\left\{\mathbb{P}_\theta^n(|\phi\nu(\widehat{\theta}_\bullet^{m_n^\star} - \theta_\bullet)|^2) : \theta_\bullet \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\right\} \leqslant (\mathbb{v} \vee \mathrm{r}^2)\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \quad \text{for all } n \in \mathbb{N}.$$

Arguing similarly as in Remark §12.16 we note that $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) = \mathrm{o}(1)$ as $n \to \infty$, whenever $\|\mathbb{1}_\bullet^m\|_\phi^2 \in \mathbb{R}^+$ for all $m \in \mathbb{N}$ and $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \mathrm{o}(1)$ as $m \to \infty$. The latter is satisfied, for example, if $(\mathfrak{a}\phi)_\bullet \in \mathbb{J}$ (in equal $\mathfrak{a}_\bullet \in \mathbb{L}_2(\phi_\bullet^2\nu)$). Note that the dimension $m_n^\star := m_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ as defined in (12.13) does not depend on the unknown parameter of interest $\theta_\bullet$ but on the class $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$ only, and thus also the statistic $\widehat{\theta}_\bullet^{m_n^\star}$. In other words, if the regularity of $\theta_\bullet$ is known in advance, then the OPE $\widehat{\theta}_\bullet^{m_n^\star}$ is a feasible estimator. □

§12.45 **Corollary** *(GSSM §12.05 continued)*. *Let $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\mathfrak{a}}^n$ as in Model §12.05, where $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ and $\theta_\bullet = \mathrm{U}\theta \in \ell_2$. Under Assumption §11.25 the OPE $\widehat{\theta}_\bullet^{m_n^\star} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{m_n^\star} \in \ell_2 \mathbb{1}_\bullet^{m_n^\star} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m_n^\star$ as in (12.13) satisfies*

$$\sup\left\{\mathrm{N}_{\mathfrak{a}}^n(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^{m_n^\star} - \theta_\bullet)|^2) : \theta_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}\right\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \quad \textit{for all } n \in \mathbb{N} \quad (12.14)$$

with constant $\mathrm{C} = 1 \vee \mathrm{r}^2$.

§12.46 **Proof** of Corollary §12.45. is given in the lecture. □

§12.47 **Illustration**. We illustrate the last results considering usual behaviour for $\mathfrak{a}_\bullet, \phi_\bullet \in \mathscr{J}$. We distinguish the following two cases **(p)** $\phi_\bullet \in \mathbb{J}$, and **(np)** $\phi_\bullet \notin \mathbb{J}$. Interestingly, in case **(p)** the bound in Proposition §12.42 is parametric, that is, $n\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) = \mathrm{O}(1)$, in case **(np)** the bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) = \infty$. In case **(np)** consider the following two specifications:

Table 04 [§12]

Order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$ $\phi_j = j^{v-1/2}$	$(a \in \mathbb{R}_{\setminus 0}^+)$ $\mathfrak{a}_j^2$	(squarred bias) $\lVert \mathfrak{a}_\bullet \mathbb{1}_\bullet^{m\vert\perp} \rVert_\phi^2$	(variance) $\lVert \mathbb{1}_\bullet^m \rVert_\phi^2$	$m_n^\star$	$\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$
(o) $v \in (0, a)$	j^{-2a}	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$
$v = 0$	j^{-2a}	m^{-2a}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a}}$	$\frac{\log n}{n}$
(s) $v \in \mathbb{R}_{\setminus 0}^+$	$e^{-j^{2a}}$	$m^{2(v-a)_+} e^{-m^{2a}}$	m^{2v}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$
$v = 0$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 04 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for $v \geqslant 0$ only. For $v < 0$ the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is parametric. □

§13 Minimax optimal estimation

§13|01 Minimax theory: a general approach

Suppose that the function of interest θ belongs to a class $\Theta \subseteq \mathbb{H}$. For each noise level $n \in \mathbb{N}$ let $\mathbb{P}_\Theta^n := (\mathbb{P}_\theta^n)_{\theta\in\Theta}$ denote a family of probability measures and let $\mathbb{E}_\theta^n$ be the expectation with respect to the measure $\mathbb{P}_\theta^n$ in $\mathbb{P}_\Theta^n$. Moreover, we assume that the probability measure associated with an observable quantity belongs to $\mathbb{P}_\Theta^n$.

§13.01 **GSSM (§10.21 continued).** Considering $\ell_2 = \mathbb{L}_2(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ and a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \ell_2)$, which is fixed and presumed to be known in advance, we illustrate the minimax approach in a Gaussian sequence space model §10.21. Here the observable stochastic process $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet$ is a noisy version of $\theta_\bullet = \mathrm{U}\theta \in \ell_2$ and $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Consequently, $\widehat{\theta}_\bullet$ admits a $\mathrm{N}_{\theta_\bullet}^n$-distribution belonging to the family $\mathrm{N}_\Theta^n := (\mathrm{N}_{\theta_\bullet}^n)_{\theta_\bullet\in\Theta}$. Summarising the observations satisfy a statistical product experiment $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\otimes\mathbb{N}}, \mathrm{N}_\Theta^n)$ where $\Theta \subseteq \ell_2$. □

Assume furthermore, that an estimator $\widetilde{\theta}$ of θ based on the observable quantities is available which takes its values in $\mathbb{H}$ but does not necessarily belong to Θ. We shall measure the accuracy of any estimator $\widetilde{\theta}$ of θ by its distance $\mathfrak{d}_{\text{ist}}(\widetilde{\theta}, \theta)$ where $\mathfrak{d}_{\text{ist}}(\cdot,\cdot)$ is a certain semi metric to be specified below. Moreover, we call the quantity $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widetilde{\theta}, \theta)\big)$ risk of the estimator $\widetilde{\theta}$ of θ.

§13.02 **Definition.** Given an estimator $\widetilde{\theta}$ of a function of interest θ belonging to a class of solutions Θ based on observable quantities with probability measure $\mathbb{P}_\theta^n \in \mathbb{P}_\Theta^n$ we call

$$\mathcal{R}_n[\,\widetilde{\theta}\,|\,\Theta\,] := \sup\big\{\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widetilde{\theta}, \theta)\big) \colon \theta \in \Theta\big\}$$

its *maximal risk* over Θ. □

§13.03 **Remark.** An advantage of taking a maximal risk instead of a risk is that the former does not depend on the unknown function θ. Imagine we would have taken a constant estimator, say $\widetilde{\theta} = h$, of θ. This would be the perfect estimator if by chance $\theta = h$, but in all other cases this estimator is likely to perform poorly. Therefore it is reasonable to consider the supremum over the whole class of possible functions in order to get consolidated findings. However, considering the maximal risk may be a very pessimistic point of view. □

§13.04 **Definition.** Consider a maximal risk $\mathcal{R}_n[\,\bullet\,|\,\Theta\,]$ over a family $\mathbb{P}_\Theta^n$ of probability measures. Let $\widehat{\theta}$ be an estimator of $\theta \in \Theta$, $\mathrm{C} \in \mathbb{R}_{\setminus 0}^+$ and for each $n \in \mathbb{N}$ let $\mathrm{R}_n^\star \in \mathbb{R}^+$ satisfy

(lower) $\mathrm{R}_n^\star$ is a *lower bound* up to the constant C^{-1} of the maximal risk over Θ, that is

$$\inf_{\widetilde{\theta}} \mathcal{R}_n[\widetilde{\theta}\,|\,\Theta] \geqslant \mathrm{C}^{-1}\,\mathrm{R}_n^\star$$

where the infimum is taken over all possible estimators of θ;

(upper) $\mathrm{R}_n^\star$ is an *upper bound* up to the constant C of the maximal risk over Θ, that is

$$\mathcal{R}_n[\widehat{\theta}\,|\,\Theta] \leqslant \mathrm{C}\,\mathrm{R}_n^\star$$

Then we call $\mathrm{R}_n^\star$ *minimax-bound* and the estimator $\widehat{\theta}$ *minimax-optimal* (up to the constant C). As a consequence, up to the constant C^2 the estimator $\widehat{\theta}$ attains the lower maximal risk bound that is, $\mathcal{R}_n[\widehat{\theta}\,|\,\Theta] \leqslant \mathrm{C}^2 \inf_{\widetilde{\theta}} \mathcal{R}_n[\widetilde{\theta}\,|\,\Theta]$. □

§13.05 **Remark**. We call a minimax-bound $(\mathrm{R}_n^\star)_{n\in\mathbb{N}}$ a *minimax-optimal rate* (of convergence) if in addition $\mathrm{R}_n^\star = \mathrm{o}(1)$ as $n \to \infty$. It is worth noting that a minimax-optimal rate is not unique since every other rate that is equivalent of order is also minimax-optimal. □

§13.06 **Nonparametric regression with uniform design (nRu)**. Let the $[0,1]\times\mathbb{R}$-valued random vector (X,Y) obeys $\mathbb{P}^X$-a.e. a nonparametric regression model $\mathbb{P}_f\big(Y\big|X\big) = f$ (see section §09). For convenience, in addition the regressor X is supposed to be uniformly distributed on the interval $[0,1]$, i.e. $X \sim \mathcal{U}_{[0,1]}$. As a consequence, we have $\mathbb{p}^X = \mathbb{1}_{[0,1]}$ and $\mathbb{L}_2([0,1],\mathscr{B}_{[0,1]},\mathbb{P}^X) = \mathbb{L}_2([0,1],\mathscr{B}_{[0,1]},\lambda_{[0,1]}) = \mathbb{L}_2(\lambda_{[0,1]})$. Here and subsequently we assume that the conditional distribution $\mathbb{P}_f^{Y|X}$ of Y given X is regular, and thus $\mathbb{P}_f^{Y|X}(\mathrm{id}_\mathbb{R}) = \mathbb{P}_f\big(Y\big|X\big) = f$ $\mathbb{P}^X$-a.s.. Let us denote in this situation by $\mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}_f^{Y|X}$ the joint distribution of (X,Y) defined on $([0,1]\times\mathbb{R}, \mathscr{B}_{[0,1]}\otimes\mathscr{B})$, but keep in mind, that the conditional distribution $\mathbb{P}_f^{Y|X}$ of Y given X is still not specified. We assume that $f \in \mathbb{F} \subseteq \mathbb{L}_2(\lambda_{[0,1]})$. Summarising the observations satisfy a statistical product experiment $(([0,1]\times\mathbb{R})^n, \mathscr{B}_{[0,1]\times\mathbb{R}}^n, \mathcal{U}_\mathbb{F}^{\otimes n} = (\mathcal{U}_f^{\otimes n})_{f\in\mathbb{F}})$ where $\mathbb{F} \subseteq \mathbb{L}_2(\lambda_{[0,1]})$. Let us assume in addition that Y given X is normally distributed with conditional mean $f(X)$ and conditional variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$, that is $\mathbb{P}_f^{Y|X} = \mathrm{N}_{(f(X),\sigma^2)}$. In this situation we denote by $\mathcal{U}_{f,\sigma} := \mathcal{U}_{[0,1]} \odot \mathrm{N}_{(f(X),\sigma^2)}$ the joint distribution of (X,Y). We first consider the case that the variance σ^2 is known *a priori* (i.e. $\sigma^2 = 1$), and in a second step we dismiss this information. Obviously, the distribution $\mathcal{U}_{f,\sigma}^{\otimes n}$ depends not only on the parameter of interest $f \in \mathbb{F}$ and the noise level $n \in \mathbb{N}$, but also on the variance $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$ which plays the role of a nuisance parameter. Consequently, let $\mathcal{U}_{\mathbb{F}\times\mathbb{R}_{\setminus 0}^+} := (\mathcal{U}_{f,\sigma})_{f\in\mathbb{F},\sigma\in\mathbb{R}_{\setminus 0}^+}$ denote the family of possible distributions of (X,Y). Summarising, if the variance is unknown then the observations satisfy a statistical product experiment $(([0,1]\times\mathbb{R})^n, \mathscr{B}_{[0,1]\times\mathbb{R}}^n, \mathcal{U}_{\mathbb{F}\times\mathbb{R}_{\setminus 0}^+}^{\otimes n})$ where $\mathbb{F} \subseteq \mathbb{L}_2(\lambda_{[0,1]})$. □

More generally, given a class of solutions Θ, a class of nuisance parameters Ξ and a noise level $n \in \mathbb{N}$ let $\mathbb{P}_{\Theta\times\Xi}^n := (\mathbb{P}_{\theta,\xi}^n)_{\theta\in\Theta,\xi\in\Xi}$ denote a family of probability measures. Moreover, we assume that the probability measure associated with an observable quantity belongs to $\mathbb{P}_{\Theta\times\Xi}^n$. Note that dismissing in Model §13.06 the assumption of a normally distributed error the class of nuisances parameters Ξ equals the family of possible conditional distributions of the error terms.

§13.07 **Definition**. Given an estimator $\widetilde{\theta}$ of a function of interest θ belonging to a class of solutions Θ based on observable quantities with probability measure $\mathbb{P}_{\theta,\xi}^n \in \mathbb{P}_{\Theta\times\Xi}^n$ we call

$$\mathcal{R}_n[\widetilde{\theta}\,|\,\Theta,\Xi] := \sup\big\{\mathbb{P}_{\theta,\xi}^n\big(\mathfrak{d}_{\mathrm{ist}}^2(\widetilde{\theta},\theta)\big)\colon \theta\in\Theta, \xi\in\Xi\big\}$$

its *maximal risk* over $\Theta\times\Xi$. □

§13.08 **Remark**. Taking the supremum over the class of nuisances parameters allows us to quantify the additional complexity due to the presence of the nuisance parameter. Moreover, if there exist an estimator $\widehat{\theta}$, a constant $\mathrm{C} \in \mathbb{R}^{+}_{\setminus 0}$ and for each $n \in \mathbb{N}$ there is $\mathrm{R}^{\star}_{n} \in \mathbb{R}^{+}$ such that

(lower) $\mathrm{R}^{\star}_{n}$ is a *lower bound* up to the constant C of the maximal risk over $\Theta \times \Xi$, that is

$$\inf_{\widetilde{\theta}} \mathcal{R}_{n}[\widetilde{\theta} \,|\, \Theta, \Xi] \geqslant \mathrm{C}^{-1}\, \mathrm{R}^{\star}_{n}$$

where the infimum is taken over all possible estimators of θ;

(upper) $\mathrm{R}^{\star}_{n}$ is an *upper bound* up to the constant C of the maximal risk over $\Theta \times \Xi$, that is

$$\mathcal{R}_{n}[\widehat{\theta} \,|\, \Theta, \Xi] \leqslant \mathrm{C}\, \mathrm{R}^{\star}_{n},$$

then we call $\mathrm{R}^{\star}_{n}$ *minimax-bound* and the estimator $\widehat{\theta}$ *minimax-optimal* (up to the constant C). As a consequence, up to the constant C^{2} the estimator $\widehat{\theta}$ attains the lower maximal risk bound that is, $\mathcal{R}_{n}[\widehat{\theta} \,|\, \Theta, \Xi] \leqslant \mathrm{C}^{2} \inf_{\widetilde{\theta}} \mathcal{R}_{n}[\widetilde{\theta} \,|\, \Theta, \Xi]$. Typically, we assume first that the nuisance parameter ξ is known *a priori*, and hence $\mathbb{P}^{n}_{\Theta \times \{\xi\}}$ is a family of probability measures associated with the observable quantities. In this situation, we consider the maximal risk $\big\{\mathbb{P}^{n}_{\theta,\xi}\big(\mathfrak{d}^{2}_{\mathrm{ist}}(\widetilde{\theta}, \theta)\big) \colon \theta \in \Theta\big\}$ and we seek a bound $\mathrm{R}^{\star}_{n}$ up to a constant which depends possibly on the nuisance parameter ξ. However, if the bound $\mathrm{R}^{\star}_{n}$ is a valid lower and upper bound up to a constant uniformly for all nuisance parameters $\xi \in \Xi$, then it is, obviously, also a bound of the maximal risk $\mathcal{R}_{n}[\widehat{\theta} \,|\, \Theta, \Xi]$. □

§13.09 **Reminder**. Considering a Hilbert space $\mathbb{J} = \mathbb{L}_{2}(\mathcal{T}, \mathscr{I}, \nu)$ and a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \mathbb{J})$, which are fixed and presumed to be known in advance, we study *statistical direct problems* as in Definition §10.19. Given weights $\mathfrak{a}_{\bullet} \in \mathscr{I}_{\setminus 0}$ we introduce $\mathbb{J}^{\mathfrak{a}} = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}}) = \mathbb{J}\mathfrak{a}_{\bullet} = \mathbb{L}_{2}(\mathfrak{a}^{2|\dagger}_{\bullet}\nu)$ endowed with $\|\cdot\|_{\mathfrak{a}^{\dagger}} := \|\cdot\|_{\mathbb{L}_{2}(\mathfrak{a}^{2|\dagger}_{\bullet}\nu)}$ and the ellipsod $\mathbb{J}^{\mathfrak{a},\mathrm{r}} := \big\{h_{\bullet} \in \mathbb{J}^{\mathfrak{a}} \colon \|h_{\bullet}\|^{2}_{\mathfrak{a}^{\dagger}} \leqslant \mathrm{r}^{2}\big\} \subseteq \mathbb{J}^{\mathfrak{a}}$, where the measures ν and $\mathfrak{a}^{2|\dagger}_{\bullet}\nu$ dominate mutually each other. We consider the following global and local measures of accuracy (compare Subsections §12|01 and §12|02).

(global) Given weights $\mathfrak{v}_{\bullet} \in \mathscr{I}_{\setminus 0}$ satisfying Assumption §11.12 introduce $\mathbb{L}_{2}(\mathfrak{v}^{2}_{\bullet}\nu) = \mathrm{dom}(\mathrm{M}_{\mathfrak{v}}) = \mathbb{J}\mathfrak{v}^{\dagger}_{\bullet} \subseteq \mathbb{J}$ and $\|\cdot\|_{\mathfrak{v}} = \|\mathrm{M}_{\mathfrak{v}_{\bullet}}\cdot\|_{\mathbb{J}}$, where $\mathbb{J}^{\mathfrak{a},\mathrm{r}} \subseteq \mathbb{L}_{2}(\mathfrak{v}^{2}_{\bullet}\nu)$ (Property §11.15). For $\theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ we call $\mathfrak{d}_{\mathrm{ist}}(\widetilde{\theta}_{\bullet}, \theta_{\bullet}) = \|\widetilde{\theta}_{\bullet} - \theta_{\bullet}\|_{\mathfrak{v}}$ *global $\mathfrak{v}$-error*, $\mathbb{P}^{n}_{\theta}\big(\|\widetilde{\theta}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}}\big)$ *global $\mathfrak{v}$-risk* and

$$\mathcal{R}^{\mathfrak{v}}_{n}[\widetilde{\theta}_{\bullet} \,|\, \mathbb{J}^{\mathfrak{a},\mathrm{r}}] := \sup\big\{\mathbb{P}^{n}_{\theta}\big(\|\widetilde{\theta}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}}\big) \colon \theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\big\}$$

maximal $\mathfrak{v}$-risk over $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$.

(local) Given $\phi_{\bullet} \in \mathscr{I}_{\setminus 0}$ satisfying Assumption §11.25 introduce $\mathrm{dom}(\phi\nu) := \big\{h_{\bullet} \in \mathbb{J} \colon \phi_{\bullet} h_{\bullet} \in \mathbb{L}_{1}(\nu)\big\}$ and the linear functional $\phi\nu : \mathbb{J} \supseteq \mathrm{dom}(\phi\nu) \to \mathbb{R}$ with $h_{\bullet} \mapsto \phi\nu(h_{\bullet}) := \nu(\phi_{\bullet} h_{\bullet})$ where $\mathbb{J}^{\mathfrak{a},\mathrm{r}} \subseteq \mathrm{dom}(\phi\nu)$ (Property §11.27). For $\theta_{\bullet} \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}$ we call $\mathfrak{d}_{\mathrm{ist}}(\widetilde{\theta}_{\bullet}, \theta_{\bullet}) = |\phi\nu(\widetilde{\theta}_{\bullet} - \theta_{\bullet})|$ *local ϕ-error*, $\mathbb{P}^{n}_{\theta}\big(|\phi\nu(\widetilde{\theta}_{\bullet} - \theta_{\bullet})|^{2}\big)$ *local ϕ-risk* and

$$\mathcal{R}^{\phi}_{n}[\widetilde{\theta}_{\bullet} \,|\, \mathbb{J}^{\mathfrak{a},\mathrm{r}}] := \sup\big\{\mathbb{P}^{n}_{\theta}\big(|\phi\nu(\widetilde{\theta}_{\bullet} - \theta_{\bullet})|^{2}\big) \colon \theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\big\}$$

maximal ϕ-risk over $\mathbb{J}^{\mathfrak{a},\mathrm{r}}$.

We formulate the results in terms of $\theta_{\bullet} = \mathrm{U}\theta \in \mathbb{J}$ rather than directly for $\theta \in \mathbb{H}$. Since U is known, considering the class $\mathbb{H}^{\mathfrak{a},\mathrm{r}} := \mathrm{U}^{\star}\mathbb{J}^{\mathfrak{a},\mathrm{r}} := \big\{\mathrm{U}^{\star}\theta_{\bullet} \colon \theta_{\bullet} \in \mathbb{J}^{\mathfrak{a},\mathrm{r}}\big\}$ we obtain immediately also bounds over $\mathbb{H}^{\mathfrak{a},\mathrm{r}}$ for the maximal global risk

$$\mathcal{R}^{\mathfrak{v}}_{n}[\widetilde{\theta} \,|\, \mathrm{U}^{\star}\mathbb{J}^{\mathfrak{a},\mathrm{r}}] := \sup\big\{\mathbb{P}^{n}_{\theta}\big(\|\mathrm{U}(\widetilde{\theta} - \theta)\|^{2}_{\mathfrak{v}}\big) \colon \theta \in \mathbb{H}^{\mathfrak{a},\mathrm{r}}\big\}$$

and maximal local risk

$$\mathcal{R}^{\phi}_{n}[\widetilde{\theta} \,|\, \mathrm{U}^{\star}\mathbb{J}^{\mathfrak{a},\mathrm{r}}] := \sup\big\{\mathbb{P}^{n}_{\theta}\big(|\phi\nu(\mathrm{U}(\widetilde{\theta} - \theta))|^{2}\big) \colon \theta \in \mathbb{H}^{\mathfrak{a},\mathrm{r}}\big\}$$

which we do not explicitly state in the sequel. □

§13|02 Deriving a lower bound: a general reduction scheme

For a detailed discussion of several other strategies to derive lower bounds we refer the reader, for example, to the text book by Tsybakov [2009].

§13.10 **Definition**. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be two probability measures on a measurable space $(\mathcal{X}, \mathscr{X})$.

(a) The function

$$\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) = \begin{cases} \mathbb{P}_0\big(\log \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1}\big) = \int \log\big(\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1}\big) d\mathbb{P}_0, & \text{if } \mathbb{P}_0 \ll \mathbb{P}_1, \\ +\infty, & \text{otherwise} \end{cases}$$

is called *Kullback-Leibler-divergence* of $\mathbb{P}_0$ with respect to $\mathbb{P}_1$.

Let $\mu \in \mathcal{M}_\sigma(\mathscr{X})$ be a $\mathbb{P}_0$ and $\mathbb{P}_1$ dominating σ-finite measure (e.g. $\mathbb{P}_0, \mathbb{P}_1 \ll \mu = \mathbb{P}_0 + \mathbb{P}_1$). We write $\mathrm{d}\mathbb{P}_0 := \mathrm{d}\mathbb{P}_0/\mathrm{d}\mu$ and $\mathrm{d}\mathbb{P}_1 := \mathrm{d}\mathbb{P}_1/\mathrm{d}\mu$ for short.

(b) The *Hellinger distance* between $\mathbb{P}_0$ and $\mathbb{P}_1$ is defined by

$$\mathrm{H}(\mathbb{P}_0, \mathbb{P}_1) := \Big(\int |\sqrt{\mathrm{d}\mathbb{P}_0} - \sqrt{\mathrm{d}\mathbb{P}_1}|^2 \Big)^{1/2} := \|\sqrt{\mathrm{d}\mathbb{P}_0} - \sqrt{\mathrm{d}\mathbb{P}_1}\|_{\mathbb{L}_2(\mu)}$$

(c) and the *Hellinger affinity* is given by

$$\rho(\mathbb{P}_0, \mathbb{P}_1) := \int \sqrt{\mathrm{d}\mathbb{P}_0}\sqrt{\mathrm{d}\mathbb{P}_1} := \langle \sqrt{\mathrm{d}\mathbb{P}_0}, \sqrt{\mathrm{d}\mathbb{P}_1} \rangle_{\mathbb{L}_2(\mu)},$$

where both do not depend on the choice of the dominating measure μ. □

§13.11 **Remark**. The Kullback-Leibler-divergence satisfies $\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) \geqslant 0$ as well as $\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) = 0$ if and only if $\mathbb{P}_0 = \mathbb{P}_1$, but $\mathrm{KL}(\cdot|\cdot)$ is not symmetric. Moreover, for product measures holds $\mathrm{KL}(\mathbb{P}_{0,1} \otimes \mathbb{P}_{0,2}|\mathbb{P}_{1,1} \otimes \mathbb{P}_{1,2}) = \mathrm{KL}(\mathbb{P}_{0,1}|\mathbb{P}_{1,1}) + \mathrm{KL}(\mathbb{P}_{0,2}|\mathbb{P}_{1,2})$. □

§13.12 **Lemma**. *(i)* $0 \leqslant \mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leqslant 2$*; (ii)* $\rho(\mathbb{P}_0, \mathbb{P}_1) = 1 - \frac{1}{2}\mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1)$*; and (iii)* $\mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leqslant \mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1)$.

§13.13 **Proof** of Lemma §13.12. Exercise. □

§13.14 **Lemma**. *For* $a_\bullet, b_\bullet \in \ell_2$ *and* $n \in \mathbb{N}$ *we have* $\mathrm{KL}(\mathrm{N}_{a_\bullet}^n|\mathrm{N}_{b_\bullet}^n) = \frac{n}{2}\|a_\bullet - b_\bullet\|_{\ell_2}^2$.

§13.15 **Proof** of Lemma §13.14. Exercise. □

§13.16 **Notation**. Recall that the semi metric $\mathfrak{d}_{\text{ist}}(\cdot,\cdot)$ is symmetric and satisfies the triangular inequality. Morever, here and subsequently we suppose that for an estimator $\widetilde{\theta}$ and parameter θ^0 and θ^1 such that $\mathfrak{d}_{\text{ist}}(\theta^0, \theta^1) \in \mathbb{R}^+_{\setminus 0}$ the quantities $\mathfrak{d}_{\text{ist}}(\widetilde{\theta}, \theta^0)$ and $\mathfrak{d}_{\text{ist}}(\widetilde{\theta}, \theta^1)$ are measurable. □

§13.17 **Lemma**. *Let* $\mathbb{P}_0$ *and* $\mathbb{P}_1$ *be two probability measures on a measurable space* $(\mathcal{X}, \mathscr{X})$*. Suppose that for an estimator* $\widetilde{\theta}$ *and parameter* θ^0 *and* θ^1 *with* $\mathfrak{d}_{\text{ist}}(\theta^0, \theta^1) \in \mathbb{R}^+_{\setminus 0}$ *the quantities* $\mathfrak{d}_{\text{ist}}(\widetilde{\theta}, \theta^0)$ *and* $\mathfrak{d}_{\text{ist}}(\widetilde{\theta}, \theta^1)$ *are measurable. Then, we have*

$$\mathbb{P}_0\big(\mathfrak{d}_{\text{ist}}^2(\widetilde{\theta}, \theta^0)\big) + \mathbb{P}_1\big(\mathfrak{d}_{\text{ist}}^2(\widetilde{\theta}, \theta^1)\big) \geqslant \frac{1}{2}\,\mathfrak{d}_{\text{ist}}^2(\theta^0, \theta^1)\,\rho^2(\mathbb{P}_0, \mathbb{P}_1). \tag{13.01}$$

§13.18 **Proof** of Lemma §13.17. is given in the lecture. □

§13|03 Lower bound based on two hypotheses

§13.19 **Lemma** *(Lower bound based on two hypotheses).* *Given a noise level $n \in \mathbb{N}$ let $\mathbb{P}_\Theta^n := (\mathbb{P}_\theta^n)_{\theta\in\Theta}$ be a family of probability measures. If there are $\theta^0, \theta^1 \in \Theta$ with associated probability measures $\mathbb{P}_0 := \mathbb{P}_{\theta^0}^n$ and $\mathbb{P}_1 := \mathbb{P}_{\theta^1}^n$ such that $\mathrm{H}(\mathbb{P}_0, \mathbb{P}_1) \leqslant 1$ then we have*

$$\inf_{\widetilde{\theta}} \mathcal{R}_n[\widetilde{\theta} \,|\, \Theta] \geqslant \frac{1}{16}\mathfrak{d}_{\mathrm{ist}}^2(\theta^0, \theta^1).$$

where the infimum is taken over all possible estimators.

§13.20 **Proof** of Lemma §13.19. is given in the lecture. □

§13.21 **Remark** *(Lower bound for a local ϕ-risk).* Due to the bounded Hellinger distance in Lemma §13.19, Le Cam's general method (see Le Cam [1973]) and Pinsker's inequality allow to derive a lower bound for a local ϕ-risk as in Reminder §13.09. However, in this special setting a lower bound can be obtained elementarily from Lemma §13.19, which in this situation states

$$\inf_{\widetilde{\theta}} \mathcal{R}_n^\phi[\widetilde{\theta}_\bullet \,|\, \Theta] \geqslant \frac{1}{16}|\phi\nu(\theta_\bullet^0 - \theta_\bullet^1)|^2.$$

If we consider furthermore candidates $\theta_\bullet^0 := \theta_\bullet^*$ and $\theta_\bullet^1 = -\theta_\bullet^*$ for some $\theta_\bullet^* \in \Theta$ such that $-\theta_\bullet^* \in \Theta$, then trivially $|\phi\nu(\theta_\bullet^0 - \theta_\bullet^1)|^2 = 4|\phi\nu(\theta_\bullet^*)|^2$ which in turn implies due to the last assertion

$$\inf_{\widetilde{\theta}} \mathcal{R}_n^\phi[\widetilde{\theta}_\bullet \,|\, \Theta] \geqslant \frac{1}{4}|\phi\nu(\theta_\bullet^*)|^2. \tag{13.02}$$

Often a minimax-optimal lower bound can be found by constructing a candidate $\theta_\bullet^* = \mathrm{U}\theta^* \in \Theta$ that has the largest possible $|\phi\nu(\theta_\bullet^*)|^2$-value but $\mathbb{P}_{\theta^*}^n$ and $\mathbb{P}_{-\theta^*}^n$ are still statistically indistinguishable in the sense that $\mathrm{H}(\mathbb{P}_{\theta^*}^n, \mathbb{P}_{-\theta^*}^n) \leqslant 1$. □

§13.22 **Reminder** *(Maximal local ϕ-risk in GSSM §13.01).* Given Model §13.01 we consider an OPE as in Section §12. Here the observable stochastic process $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\mathfrak{a}}^n$ is a noisy version of $\theta_\bullet = \mathrm{U}\theta \in \Theta \subseteq \ell_2$ and $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Consequently, $\widehat{\theta}_\bullet$ admits a $\mathrm{N}_{\mathfrak{a}}^n$-distribution belonging to the family $\mathrm{N}_\Theta^n := (\mathrm{N}_{\mathfrak{a}}^n)_{\mathfrak{a}\in\Theta}$. Summarising the observations satisfy a statistical product experiment $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}}, \mathrm{N}_\Theta^n)$ where $\Theta \subseteq \ell_2$. Under Assumption §11.25 in Corollary §12.45 an upper bound for a maximal local ϕ-risk of an OPE is shown. More precisely, the performance of the OPE $\widehat{\theta}_\bullet^m = \widehat{\theta}_\bullet \mathbb{1}_\bullet^m \in \ell_2\mathbb{1}_\bullet^m \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m \in \mathbb{N}$ is measured by its maximal local ϕ-risk, that is

$$\mathcal{R}_n^\phi[\widehat{\theta}_\bullet^m \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}] := \sup\big\{\mathrm{N}_{\mathfrak{a}}^n\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^m - \theta_\bullet)|^2\big) : \theta_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.13) where for $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \tag{13.03}$$

By Corollary §12.45 under Assumption §11.25 the maximal local ϕ-risk of an OPE $\widehat{\theta}_\bullet^{m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (13.03) satisfies

$$\mathcal{R}_n^\phi[\widehat{\theta}_\bullet^{m_n^\star} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}] \leqslant \mathrm{C}\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$$

with $\mathrm{C} = 1 \vee \mathrm{r}^2$. □

§13.23 **Notation**. For sequences $a_\bullet, b_\bullet \in (\mathbb{K})^{\mathbb{N}}$ taking its values in $\mathbb{K} \in \{\mathbb{R}, \mathbb{R}^+, \mathbb{R}^+_{\setminus 0}, \mathbb{Q}, \mathbb{Z}, \dots\}$ we write $a_\bullet \in (\mathbb{K})^{\mathbb{N}}_{\nearrow}$ and $b_\bullet \in (\mathbb{K})^{\mathbb{N}}_{\searrow}$ if $a_\bullet$ and $b_\bullet$, respectively, is monotonically *non-decreasing* and *non-increasing*. If in addition $a_n \to \infty$ and $b_n \to 0$ as $n \to \infty$, then we write $a_\bullet \in (\mathbb{K})^{\mathbb{N}}_{\uparrow\infty}$ and $b_\bullet \in (\mathbb{K})^{\mathbb{N}}_{\downarrow 0}$ for short. □

§13.24 **Assumption**. Consider $\phi_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ such that $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\phi)_\bullet \in \ell_2$ (i.e. Assumption §11.25 is satisfied and $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$), and in addition $\mathfrak{a}^2_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$. □

§13.25 **Comment**. Assuming $\mathfrak{a}^2_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\searrow}$ is rather weak. If we suppose in addition $\liminf_{j\to\infty} \mathfrak{a}^2_j \geqslant c > 0$, and hence $\mathfrak{a}^2_\bullet \notin (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$, then the assumption $(\mathfrak{a}\phi)_\bullet \in \ell_2$ implies $\phi_\bullet \in \ell_2$ and hence the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \phi_\bullet)$ is parametric (Illustration §12.47). Since we are interested in the case of a non-parametric, the additional assumption $\mathfrak{a}^2_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$ imposes a rather weak condition satisfied also in Illustration §12.47.
If $\mathfrak{a}_2^2 > n^{-1}$ then exploiting the definition (13.03) and $\phi_2 \in \mathbb{R}_{\setminus 0}$ we have

$$\mathrm{R}^1_n(\mathfrak{a}_\bullet, \phi_\bullet) = n^{-1}\phi_1^2 + (\mathfrak{a}\phi)_2^2 + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{2|\perp}\|^2_\phi > n^{-1}\phi_1^2 + n^{-1}\phi_2^2 + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{2|\perp}\|^2_\phi = \mathrm{R}^2_n(\mathfrak{a}_\bullet, \phi_\bullet),$$

and consequently $m^\star_n - 1 \in \mathbb{N}$. In this situation, from (definition of the arg min)

$$\begin{aligned} n^{-1}\|\mathbb{1}_\bullet^{m^\star_n - 1}\|^2_\phi + (\mathfrak{a}\phi)^2_{m^\star_n} + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m^\star_n|\perp}\|^2_\phi \\ = \mathrm{R}^{m^\star_n - 1}_n(\mathfrak{a}_\bullet, \phi_\bullet) > \mathrm{R}^{m^\star_n}_n(\mathfrak{a}_\bullet, \phi_\bullet) = n^{-1}\|\mathbb{1}_\bullet^{m^\star_n - 1}\|^2_\phi + n^{-1}\phi^2_{m^\star_n} + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m^\star_n|\perp}\|^2_\phi \end{aligned}$$

follows $(\mathfrak{a}\phi)^2_{m^\star_n} > n^{-1}\phi^2_{m^\star_n}$, and hence $\mathfrak{a}^2_{m^\star_n} > n^{-1}$ (since $\phi_{m^\star_n} \in \mathbb{R}_{\setminus 0}$). On the other hand from

$$\begin{aligned} n^{-1}\|\mathbb{1}_\bullet^{m^\star_n}\|^2_\phi + (\mathfrak{a}\phi)^2_{m^\star_n+1} + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m^\star_n+1|\perp}\|^2_\phi \\ = \mathrm{R}^{m^\star_n}_n(\mathfrak{a}_\bullet, \phi_\bullet) \leqslant \mathrm{R}^{m^\star_n+1}_n(\mathfrak{a}_\bullet, \phi_\bullet) = n^{-1}\|\mathbb{1}_\bullet^{m^\star_n}\|^2_\phi + n^{-1}\phi^2_{m^\star_n} + \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m^\star_n+1|\perp}\|^2_\phi \end{aligned}$$

follows $(\mathfrak{a}\phi)^2_{m^\star_n+1} \leqslant n^{-1}\phi^2_{m^\star_n+1}$, and hence $\mathfrak{a}^2_{m^\star_n+1} \leqslant n^{-1}$ (since $\phi_{m^\star_n+1} \in \mathbb{R}_{\setminus 0}$). Assuming $\mathfrak{a}_2^2 > n^{-1}$ we use the property $\mathfrak{a}^2_{m^\star_n} > n^{-1} \geqslant \mathfrak{a}^2_{m^\star_n+1}$ in the next proof. □

§13.26 **Proposition** *(GSSM §13.01 continued). Let* $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}^n_\theta$ *as in Model §13.01 where* $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}^{\otimes\mathbb{N}}_{(0,1)}$ *and* $\theta_\bullet = \mathrm{U}\theta \in \ell_2$. *Given Assumption §13.24 and the notations in* (13.03) *for all* $n \in \mathbb{N} \cap (\mathfrak{a}_2^{-2}, \infty)$ *we have*

$$\inf_{\widetilde{\theta}} \mathcal{R}^\phi_n[\widetilde{\theta}_\bullet \,|\, \ell^{\mathfrak{a},\mathrm{r}}_2] \geqslant 8^{-1}(1 \wedge 2\mathrm{r}^2)\ \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \phi_\bullet) \tag{13.04}$$

where the infimum is taken over all estimators $\widetilde{\theta}_\bullet$.

§13.27 **Proof** of Proposition §13.26. is given in the lecture. □

§13.28 **Illustration**. Consider the two specifications **(o)** and **(s)** depict in Table 04 [§12] of the Illustration §12.47. In both cases Assumption §13.24 is satisfied. Consequently, due to Proposition §13.26 the Table 04 [§12] presents the order of the *minimax rate* $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \phi_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{\theta}^{m^\star_n}_\bullet = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{m^\star_n} \in \ell_2 \mathbb{1}_\bullet^{m^\star_n} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with optimally selected dimension $m^\star_n$ (Corollary §12.45). We shall stress, that the order of $m^\star_n$ given in the Table 04 [§12] depends on the parameter $a \in \mathbb{R}^+_{\setminus 0}$ characterising the (abstract) smoothness of the solution which is generally not known in advance. □

§13|04 Lower bound based on m hypotheses

§13.29 **Notation.** For $m \in \mathbb{N}$ set $\mathcal{T}_m := \{-1,1\}^m$ and for each $\tau := (\tau_j)_{j\in[\![m]\!]} \in \mathcal{T}_m$ and $j \in [\![m]\!]$ introduce $\tau^{(j)} \in \mathcal{T}_m$ given by $\tau_j^{(j)} := -\tau_j$ and $\tau_l^{(j)} := \tau_l$ for $l \in [\![m]\!]\backslash\{j\}$. □

§13.30 **Lemma** *(Assouad's cube technique).* *Given a noise level $n \in \mathbb{N}$ let $\mathbb{P}_\Theta^n := (\mathbb{P}_\theta^n)_{\theta\in\Theta}$ be a family of probability measures. Suppose there exist $m \in \mathbb{N}$ and distances $\mathfrak{d}_{\text{ist}}^{(j)}(\cdot,\cdot)$, $j \in [\![m]\!]$ such that $\mathfrak{d}_{\text{ist}}^2(\cdot,\cdot) \geqslant \sum_{j\in[\![m]\!]} |\mathfrak{d}_{\text{ist}}^{(j)}(\cdot,\cdot)|^2$. If for each $\tau \in \mathcal{T}_m$ there is $\theta^\tau \in \Theta$ with associated probability measure $\mathbb{P}_\tau := \mathbb{P}_{\theta^\tau}^n$ such that for all $\tau \in \mathcal{T}_m$ and $j \in [\![m]\!]$ we have $\mathrm{H}(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 1$ then we obtain*

$$\inf_{\widetilde{\theta}} \mathcal{R}_n[\,\widetilde{\theta}\,|\,\Theta\,] \geqslant 2^{-m} \sum_{\tau\in\mathcal{T}_m} \frac{1}{16} \sum_{j\in[\![m]\!]} |\mathfrak{d}_{\text{ist}}^{(j)}(\theta^\tau, \theta^{\tau^{(j)}})|^2$$

where the infimum is taken over all possible estimators.

§13.31 **Proof** of Lemma §13.30. is given in the lecture. □

§13.32 **Remark** *(Lower bound of a global $\mathfrak{v}$-risk).* The last result allows to derive a lower bound for a global $\mathfrak{v}$-risk as in Reminder §13.09 which in case $\mathbb{J} = \ell_2 = \mathbb{L}_2(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ states

$$\inf_{\widetilde{\theta}} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{\theta}_\bullet\,|\,\Theta\,] \geqslant 2^{-m} \sum_{\tau\in\mathcal{T}_m} \frac{1}{16} \sum_{j\in[\![m]\!]} \mathfrak{v}_j^2 |\theta_j^\tau - \theta_j^{\tau^{(j)}}|^2.$$

If we assume furthermore candidates $\theta_\bullet^\tau := (\tau_j \theta_j^\star \mathbb{1}_j^m)_{j\in\mathbb{N}} \in \Theta$, $\tau \in \mathcal{T}_m$, for some $\theta_\bullet^\star = \mathrm{U}\theta^\star \in \Theta$, then it is easily seen that $\sum_{j\in[\![m]\!]} \mathfrak{v}_j^2 |\theta_j^\tau - \theta_j^{\tau^{(j)}}|^2 = 4\sum_{j\in[\![m]\!]} \mathfrak{v}_j^2 |\theta_j^\star|^2 = 4\|\theta_\bullet^\star \mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$ which in turn implies

$$\inf_{\widetilde{\theta}} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{\theta}_\bullet\,|\,\Theta\,] \geqslant 2^{-m} \sum_{\tau\in\mathcal{T}_m} \frac{1}{4}\|\theta_\bullet^\star \mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 = \frac{1}{4}\|\theta_\bullet^\star \mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2. \tag{13.05}$$

Often a minimax-optimal lower bound can be found by choosing the parameter m and the function θ^* that have the largest possible $\|\theta_\bullet^\star \mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$-value although that the associated $\mathbb{P}_\tau$, $\tau \in \mathcal{T}_m$ are still statistically indistinguishable in the sense that $\mathrm{H}(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 1$ for all $j \in [\![m]\!]$ and $\tau \in \mathcal{T}_m$. □

§13.33 **Reminder** *(Maximal global $\mathfrak{v}$-risk in GSSM §13.01).* Given Model §13.01 we consider an OPE as in Section §12. Here the observable stochastic processi $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_\theta^n$ is a noisy version of $\theta_\bullet = \mathrm{U}\theta \in \Theta \subseteq \ell_2$ and $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Consequently, $\widehat{\theta}_\bullet$ admits a N_θ^n-distribution belonging to the family $\mathrm{N}_\Theta^n := (\mathrm{N}_\theta^n)_{\theta\in\Theta}$. Summarising the observations satisfy a statistical product experiment $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}}, \mathrm{N}_\Theta^n)$ where $\Theta \subseteq \ell_2$. Under Assumption §11.12 in Corollary §12.24 an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE is shown. More precisely, the performance of the OPE $\widehat{\theta}_\bullet^m = \widehat{\theta}_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}$-risk over the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\theta}_\bullet^m\,|\,\ell_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathrm{N}_\theta^n\big(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2\big) : \theta_\bullet \in \ell_2^{\mathfrak{a}_\bullet,\mathrm{r}}\big\}.$$

Let us recall (12.06) where for $n, m \in \mathbb{N}$ we have defined $(\mathfrak{a}\mathfrak{v})_{(m)}^2 := \|(\mathfrak{a}\mathfrak{v})_\bullet^2 \mathbb{1}_\bullet^{m|\perp}\|_{\ell_\infty}$ and

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := (\mathfrak{a}\mathfrak{v})_{(m)}^2 \vee n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \tag{13.06}$$

By Corollary §12.24 under Assumption §11.12 the maximal global $\mathfrak{v}$-risk of an OPE $\widehat{\theta}_\bullet^{m_n^\star}$ with optimally chosen dimension $m_n^\star$ as in (13.06) satisfies

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\theta}_\bullet^{m_n^\star}\,|\,\ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$$

with $\mathrm{C} = 1 + \mathrm{r}^2$. □

§13.34 **Notation.** For $w_\bullet \in \ell_\infty \cap (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ we set $w_{(0)}^2 := \|w_\bullet^2\|_{\ell_\infty}$ and $w_{(\bullet)}^2 = (w_{(j)}^2 := \|w_\bullet^2 \mathbb{1}_\bullet^{j\downarrow\perp}\|_{\ell_\infty})_{j\in\mathbb{N}}$, where by construction $w_{(j)}^2 = \sup\{w_i^2 : i \in \mathbb{N} \cap [j+1, \infty)\}$, $j \in \mathbb{N}_0$ and $w_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_\searrow^{\mathbb{N}}$. □

§13.35 **Assumption.** Consider $\mathfrak{v}_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ such that $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\mathfrak{v})_\bullet \in \ell_\infty$ (i.e. Assumption §11.12 is satisfied), and in addition $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ and there exists $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} \in (0, 1]$ such that for all $m \in \mathbb{N}$

$$(\mathfrak{a}\mathfrak{v})_{(m-1)}^2 \geqslant \min\{(\mathfrak{a}\mathfrak{v})_j^2 : j \in [\![m]\!]\} \geqslant \mathrm{C}_{(\mathfrak{a}\mathfrak{v})}(\mathfrak{a}\mathfrak{v})_{(m-1)}^2$$

or in equal $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\|(\mathfrak{a}\mathfrak{v})_\bullet^{-2}\mathbb{1}_\bullet^m\|_{\ell_\infty} \leqslant (\mathfrak{a}\mathfrak{v})_{(m-1)}^{-2}$. □

§13.36 **Comment.** Note that $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_\searrow^{\mathbb{N}}$ by definition, hence $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ is satisfied if and only if $(\mathfrak{a}\mathfrak{v})_{(m)}^2 = \mathrm{o}(1)$ as $m \to \infty$ (i.e. the maximal global approximation is consistent). Moreover if $(\mathfrak{a}\mathfrak{v})_\bullet^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ then we have trivially $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ and $\|(\mathfrak{a}\mathfrak{v})_\bullet^{-2}\mathbb{1}_\bullet^m\|_{\ell_\infty} = (\mathfrak{a}\mathfrak{v})_m^{-2} = (\mathfrak{a}\mathfrak{v})_{(m-1)}^{-2}$ for all $m \in \mathbb{N}$, i.e. Assumption §13.35 is satisfied with $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} = 1$.
For $m_n^\star$ and $\mathrm{R}_n^\star := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ as in (13.06) we distinguish case i) : $\mathrm{R}_n^\star = n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2 > (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2$ and case ii) : $\mathrm{R}_n^\star = (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 \geqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$. Consider case i) first. If $(\mathfrak{a}\mathfrak{v})_{(1)}^2 > n^{-1}\mathfrak{v}_1^2$ then $\mathrm{R}_n^1(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = n^{-1}\mathfrak{v}_1^2 \vee (\mathfrak{a}\mathfrak{v})_{(1)}^2 = (\mathfrak{a}\mathfrak{v})_{(1)}^2$ and hence $m_n^\star - 1 \in \mathbb{N}$. In this situation, from $n^{-1}\|\mathbb{1}_\bullet^{m_n^\star - 1}\|_{\mathfrak{v}}^2 < n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$ (since $\mathfrak{v}_{m_n^\star}^2 \in \mathbb{R}_{\setminus 0}^+$), the definition (13.06) and

$$n^{-1}\|\mathbb{1}_\bullet^{m_n^\star - 1}\|_{\mathfrak{v}}^2 \vee (\mathfrak{a}\mathfrak{v})_{(m_n^\star - 1)}^2 = \mathrm{R}_n^{m_n^\star - 1}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) > \mathrm{R}_n^\star = n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$$

it follows $\mathrm{R}_n^{m_n^\star - 1}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = (\mathfrak{a}\mathfrak{v})_{(m_n^\star - 1)}^2$ and hence $(\mathfrak{a}\mathfrak{v})_{(m_n^\star - 1)}^2 > n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$. Consider case ii). We set

$$m_n^\diamond := \min\{m \in \mathbb{N} \cap [m_n^\star + 1, \infty) : n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \geqslant (\mathfrak{a}\mathfrak{v})_{(m)}^2\} \tag{13.07}$$

where the defining set is not empty since $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$. We note that $(\mathfrak{a}\mathfrak{v})_{(m_n^\diamond - 1)}^2 = (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2$. Indeed, in the non trivial case $m_n^\diamond - 1 > m_n^\star$ for each $m \in [\![m_n^\star + 1, m_n^\diamond - 1]\!]$ we have $\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = (\mathfrak{a}\mathfrak{v})_{(m)}^2 > n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$, which together with $(\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 = \mathrm{R}_n^\star \leqslant \mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = (\mathfrak{a}\mathfrak{v})_{(m)}^2 \leqslant (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2$ (since $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$) implies the equality $(\mathfrak{a}\mathfrak{v})_{(m)}^2 = (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2$ for all $m \in [\![m_n^\star + 1, m_n^\diamond - 1]\!]$. Moreover, from $(\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 = \mathrm{R}_n^\star \leqslant \mathrm{R}_n^{m_n^\diamond}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = n^{-1}\|\mathbb{1}_\bullet^{m_n^\diamond}\|_{\mathfrak{v}}^2$ it follows $(\mathfrak{a}\mathfrak{v})_{(m_n^\diamond - 1)}^2 = (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 \leqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\diamond}\|_{\mathfrak{v}}^2$. To summarise, assuming $(\mathfrak{a}\mathfrak{v})_{(1)}^2 > n^{-1}\mathfrak{v}_1^2$ we use in the next proof the properties case i) $(\mathfrak{a}\mathfrak{v})_{(m_n^\star - 1)}^2 > n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$ and case ii) $(\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 = (\mathfrak{a}\mathfrak{v})_{(m_n^\diamond - 1)}^2 \leqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\diamond}\|_{\mathfrak{v}}^2$. □

§13.37 **Proposition** *(GSSM §13.01 continued). Let $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_\theta^n$ as in Model §13.01 where $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$ and $\theta_\bullet = \mathrm{U}\theta \in \ell_2$. Given Assumption §13.35 and the notations in (13.06) for all $n \in \mathbb{N} \cap (\mathfrak{v}_1^2(\mathfrak{a}\mathfrak{v})_{(1)}^{-2}, \infty)$ we have*

$$\inf_{\widetilde{\theta}} \mathcal{R}_n^{\mathfrak{v}}[\widetilde{\theta}_\bullet \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}] \geqslant 8^{-1}(1 \wedge 2\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\mathrm{r}^2)\ \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \tag{13.08}$$

where the infimum is taken over all estimators $\widetilde{\theta}_\bullet$.

§13.38 **Proof** of Proposition §13.37. is given in the lecture. □

§13.39 **Illustration.** Consider the two specifications **(o)** and **(s)** depict in Table 02 [§12] of the Illustration §12.26. In both cases Assumption §13.35 is satisfied. Consequently, due to Proposition §13.37 the Table 02 [§12] presents the order of the *minimax rate* $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{\theta}_\bullet^{m_n^\star} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{m_n^\star} \in \ell_2 \mathbb{1}_\bullet^{m_n^\star} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with optimally selected dimension $m_n^\star$ (Corollary §12.24). We shall stress, that the order of $m_n^\star$ given in the Table 02 [§12] depends on the parameter $a \in \mathbb{R}_{\setminus 0}^+$ characterising the (abstract) smoothness of the solution which is generally not known in advance. □

§14 Data-driven estimation

§14|01 Data-driven estimation procedures

Considering a Hilbert space $\mathbb{J} = \mathbb{L}_2(\mathcal{J}, \mathscr{J}, \nu)$ and a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \mathbb{J})$, which are fixed and presumed to be known in advance, we study data-driven estimation procedures in *statistical direct problems* as in Definition §10.19. Precisely, we consider the observable noisy version $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\varepsilon_\bullet$ of the parameter $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ where the centred stochastic processes $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathcal{J}}$ satisfies Assumption §10.04 and $n \in \mathbb{N}$ is a sample size. We denote by $\mathbb{P}_\theta^n$ the distribution of $\widehat{\theta}_\bullet$. Based on the noisy parameter $\widehat{\theta}_\bullet$ we consider the family $(\widehat{\theta}_\bullet^m = \widehat{\theta}_\bullet \mathbb{1}_\bullet^m)_{m\in\mathbb{N}}$ of orthogonal projections estimators (OPE's) of $\theta_\bullet$ defined in Definition §12.04. For each $m \in \mathbb{N}$ we shall measure the accuracy of the OPE $\widehat{\theta}_\bullet^m$ by its risk $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^m, \theta)\big)$ where $\mathfrak{d}_{\text{ist}}(\cdot,\cdot)$ is a certain semi metric such as a global $\mathfrak{v}$-error (Definition §12.09) or a local ϕ-error (Definition §12.30). Moreover, given $\theta_\bullet = \mathrm{U}\theta \in \mathbb{J}$ we consider the family of orthogonal projections (OP's) $(\theta_\bullet^m = \theta_\bullet \mathbb{1}_\bullet^m \in \mathbb{J}\mathbb{1}_\bullet^m)_{m\in\mathbb{N}}$ (Definition §11.08) where we tactically set $\theta_\bullet^\infty := \theta_\bullet$. Let us here assume that there exist $\mathrm{C} \in \mathbb{R}_{\setminus 0}^+$, and for each $n, m \in \mathbb{N}$, $\mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \in \mathbb{R}^+$ and $\mathrm{R}_n^m(\theta_\bullet, \mathfrak{d}_{\text{ist}}) = \mathfrak{d}_{\text{ist}}^2(\theta_\bullet^m, \theta_\bullet) + \mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\text{ist}})$ such that the risk of the estimator $\widehat{\theta}^m$ satisfy

$$\mathrm{C}^{-1}\mathrm{R}_n^m(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \leqslant \mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^m, \theta_\bullet)\big) \leqslant \mathrm{C}\{\mathfrak{d}_{\text{ist}}^2(\theta_\bullet^m, \theta_\bullet) + \mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\text{ist}})\} = \mathrm{C}\mathrm{R}_n^m(\theta_\bullet, \mathfrak{d}_{\text{ist}}). \tag{14.01}$$

Minimising the right hand side in the last display as a function of $m \in \mathbb{N}$ leads to an optimal dimension (if it exists) and upper bound

$$\begin{aligned} m_n^\circ := m_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}}) := \arg\min\big\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{d}_{\text{ist}}) = \mathfrak{d}_{\text{ist}}^2(\theta_\bullet^m, \theta_\bullet) + \mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \colon m \in \mathbb{N}\big\} \quad \text{and} \\ \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}}) := \mathrm{R}_n^{m_n^\circ}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) = \min\big\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \colon m \in \mathbb{N}\big\}. \end{aligned} \tag{14.02}$$

Combining (14.01) and (14.02) (up to the constant C) we have that m_n° is an *oracle dimension*, $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}})$ an *oracle bound* and the OPE $\widehat{\theta}^{m_n^\circ}$ with oracle dimension m_n° is *oracle optimal*. However, the oracle dimension $m_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}})$ given in (14.02) depends on the unknown parameter of interest $\theta_\bullet$, and thus also the oracle optimal statistic $\widehat{\theta}_\bullet^{m_n^\circ}$. In other words $\widehat{\theta}_\bullet^{m_n^\circ}$ is not a feasible estimator. We present in what follows two data-driven procedures to select a dimension $\widehat{m}$ within an admissible subset $[\![\mathrm{M}]\!] \subseteq \mathbb{N}$ of dimension parameters given by an integer $\mathrm{M} \in \mathbb{N}$, which eventually leads to a feasible data-driven estimator $\widehat{\theta}_\bullet^{\widehat{m}}$ depending on the observable quantities only. We call any data-driven estimator $\widehat{\theta}_\bullet$ *adaptive* for a class Θ of solutions if for all $\theta_\bullet \in \Theta$ there is a constant $\mathrm{K}_\theta \in \mathbb{R}_{\setminus 0}^+$ possibly depending on $\theta_\bullet$ such that $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet, \theta_\bullet)\big) \leqslant \mathrm{K}_\theta \mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}})$ for all $n \in \mathbb{N}$. Each of those two different data-driven strategies involves in addition a sequence $\mathfrak{p}\mathrm{en}_\bullet = (\mathfrak{p}\mathrm{en}_m)_{m\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ of penalties. Both, the upper bound M and the sequence of penalties, depend on the noise level n and possibly on the class Θ of solutions. However, for ease of presentation we omit the additional subscripts. We eventually show that there are constants $\mathrm{C}_1, \mathrm{C}_2 \in \mathbb{R}_{\setminus 0}^+$ possibly depending on the solution $\theta_\bullet \in \Theta$ and $(\mathfrak{b}\mathrm{ias}_m(\theta_\bullet, \mathfrak{d}_{\text{ist}}))_{m\in\mathbb{N}}, (\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}))_{n\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ such that for all $n \in \mathbb{N}$ the risk of the data-driven estimator $\widehat{\theta}_\bullet^{\widehat{m}}$ satisfies

$$\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) \leqslant \mathrm{C}_1 \min\big\{\mathfrak{b}\mathrm{ias}_m^2(\theta_\bullet, \mathfrak{d}_{\text{ist}}) + \mathfrak{p}\mathrm{en}_m \colon m \in [\![\mathrm{M}]\!]\big\} + \mathrm{C}_2\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}). \tag{14.03}$$

If in addition for all $\theta_\bullet \in \Theta$ there is a constant $\mathrm{C}_3 \in \mathbb{R}_{\setminus 0}^+$ such that for all $n \in \mathbb{N}$ we have also

$$m_n^\circ \in [\![\mathrm{M}]\!], \quad \mathfrak{b}\mathrm{ias}_{m_n^\circ}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \leqslant \mathrm{C}_3\mathfrak{d}_{\text{ist}}(\theta_\bullet^{m_n^\circ}, \theta_\bullet) \quad \text{and} \quad \mathfrak{p}\mathrm{en}_{m_n^\circ} \leqslant \mathrm{C}_3\mathfrak{v}\mathrm{ar}_{n,m_n^\circ}(\theta_\bullet, \mathfrak{d}_{\text{ist}}). \tag{14.04}$$

Then due to (14.03) the data-driven estimator $\widehat{\theta}_\bullet^{\widehat{m}}$ satisfies

$$\begin{aligned} \mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) &\leqslant \mathrm{C}_1\mathrm{C}_3 \min\big\{\mathfrak{d}_{\text{ist}}^2(\theta_\bullet^m, \theta_\bullet) + \mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \colon m \in [\![\mathrm{M}]\!]\big\} + \mathrm{C}_2\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}) \\ &= \mathrm{C}_1\mathrm{C}_3\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{d}_{\text{ist}}) + \mathrm{C}_2\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}). \end{aligned}$$

and hence, if in addition $\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) \leqslant \mathrm{C}_4 \mathrm{R}^{\circ}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}})$ for a constant $\mathrm{C}_4 \in \mathbb{R}^+_{\setminus 0}$, then $\widehat{\theta}^{\widehat{m}}_\bullet$ is *adaptive*. Indeed, we have $\mathbb{P}^n_\vartheta\big(\mathfrak{d}^2_{\mathrm{ist}}(\widehat{\theta}^{\widehat{m}}_\bullet, \theta_\bullet)\big) \leqslant \mathrm{K}\mathrm{R}^{\circ}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}})$ with $\mathrm{K} := \mathrm{C}_1\mathrm{C}_3 + \mathrm{C}_2\mathrm{C}_4$.

§14.01 **Remark**. In oder to establish a feasible method we have to select an upper bound M. Let us briefly descibe heuristically the strategy we eventually apply. For each $\theta_\bullet \in \Theta$ and $n, m \in \mathbb{N}$ let $\mathfrak{v}\mathrm{ar}_{n,m}(\mathfrak{d}_{\mathrm{ist}}) = \mathfrak{v}\mathrm{ar}_{n,m}(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}})$ do not depend on $\theta_\bullet \in \Theta$ and moreover let $\mathfrak{v}\mathrm{ar}_{n,\bullet}(\mathfrak{d}_{\mathrm{ist}}) = (\mathfrak{v}\mathrm{ar}_{n,m}(\mathfrak{d}_{\mathrm{ist}}))_{m\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}_{\uparrow\infty}$ be known in advance. Furthermore, for all $n \in \mathbb{N}$ let $\mathfrak{v}\mathrm{ar}_{n,1}(\mathfrak{d}_{\mathrm{ist}}) \leqslant \mathrm{C}_{\mathfrak{d}_{\mathrm{ist}}}$ for some constant $\mathrm{C}_{\mathfrak{d}_{\mathrm{ist}}} \in \mathbb{R}^+_{\setminus 0}$, which is evidently also known in advance. Consequently, the defining set of $\mathrm{M}_n := \max\big\{m \in \mathbb{N} \colon \mathfrak{v}\mathrm{ar}_{n,m}(\mathfrak{d}_{\mathrm{ist}}) \leqslant \mathrm{C}_{\mathfrak{d}_{\mathrm{ist}}}\big\}$ is not empty and finite. For all $n \in \mathbb{N}$ with $\mathrm{R}^{\circ}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) \leqslant \mathrm{C}_{\mathfrak{d}_{\mathrm{ist}}}$ follows then $m^\circ_n \in [\![\mathrm{M}_n]\!]$ since $\mathrm{C}_{\mathfrak{d}_{\mathrm{ist}}} \geqslant \mathrm{R}^{\circ}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) \geqslant \mathfrak{v}\mathrm{ar}_{n,m^\circ_n}(\mathfrak{d}_{\mathrm{ist}})$. In other words the feasible upper bound M_n satisfies the first condition in assumption (14.04). □

§14.02 **GSSM (§10.21 continued)**. Considering $\ell_2 = \mathbb{L}_2(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ and a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{H}, \ell_2)$, which is fixed and presumed to be known in advance, we illustrate the different data-driven procedures in a Gaussian sequence space model §10.21. Here the observable stochastic process $\dot{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet$ is a noisy version of $\theta_\bullet = \mathrm{U}\theta \in \ell_2$ and $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}^{\otimes\mathbb{N}}_{(0,1)}$. Consequently, $\dot{\theta}_\bullet$ admits a N^n_ϑ-distribution belonging to the family $\mathrm{N}^n_\Theta := (\mathrm{N}^n_\vartheta)_{\vartheta\in\Theta}$. Summarising the observations satisfy a statistical product experiment $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\otimes\mathbb{N}}, \mathrm{N}^n_\Theta)$ where $\Theta \subseteq \ell_2$. □

§14|02 Model selection

Given a noisy version $\dot{\theta}_\bullet \sim \mathbb{P}^n_\vartheta$ in a *statistical direct problem* as in Definition §10.19. and a collection of admissible models $[\![\mathrm{M}]\!]$ for some $\mathrm{M} \in \mathbb{N}$ we seek to minimise the global $\mathfrak{v}$-risk within the family $(\widehat{\theta}^m_\bullet = \dot{\theta}_\bullet \mathbb{1}^m_\bullet)_{m\in[\![M]\!]}$ of OPE's defined in Definition §12.04. Here and subsequently, let Assumption §12.07, $\mathfrak{v}_\bullet \in \mathscr{J}_{\setminus 0}$, $\theta_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$ and $\mathbb{1}^m_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$ for all $m \in \mathbb{N}$ be satisfied.

§14.03 **Reminder**. For all $n, m \in \mathbb{N}$ we set

$$\mathrm{R}^m_n(\theta_\bullet, \mathfrak{v}_\bullet) := \|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + n^{-1}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}, \quad m^\circ_n := \arg\min\big\{\mathrm{R}^m_n(\theta_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^{\circ}_n(\theta_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}^{m^\circ_n}_n(\theta_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}^m_n(\theta_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\} \tag{14.05}$$

Since $\theta^m_\bullet = \theta_\bullet \mathbb{1}^m_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$ (Property §11.09) and $\widehat{\theta}^m_\bullet = \dot{\theta}_\bullet \mathbb{1}^m_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$ $\mathbb{P}^n_\vartheta$-a.s. (Comment §12.08), for each $a_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)\mathbb{1}^m_\bullet$ applying the Cauchy-Schwarz inequality we have

$$\mathfrak{v}^2_\bullet\nu(|a_\bullet\widehat{\theta}_\bullet|) = \mathfrak{v}^2_\bullet\nu(|a_\bullet\dot{\theta}_\bullet\mathbb{1}^m_\bullet|) \leqslant \|a_\bullet\|_{\mathfrak{v}}\|\widehat{\theta}^m_\bullet\|_{\mathfrak{v}} \in \mathbb{R}^+,$$

and hence $a_\bullet\widehat{\theta}_\bullet \in \mathbb{L}_1(\mathfrak{v}^2_\bullet\nu)$ $\mathbb{P}^n_\vartheta$-a.s.. □

The first selection method is inspired by the work of Barron et al. [1999] and for an extensive overview of model selection by penalised contrast, the reader may refer to Massart [2007]. Let us introduce a contrast function

$$\Upsilon : \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu) \supseteq \bigcup_{m\in\mathbb{N}} \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)\mathbb{1}^m_\bullet \to \mathbb{R} \text{ with } a_\bullet \mapsto \Upsilon(a_\bullet) := \|a_\bullet\|^2_{\mathfrak{v}} - 2\mathfrak{v}^2_\bullet\nu(a_\bullet\widehat{\theta}_\bullet) \tag{14.06}$$

where for each $m \in \mathbb{N}$ the OPE $\widehat{\theta}^m_\bullet = \dot{\theta}_\bullet\mathbb{1}^m_\bullet$ and $a_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)\mathbb{1}^m_\bullet \subseteq \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)$ satisfy

$$\Upsilon(a_\bullet) = \|a_\bullet\|^2_{\mathfrak{v}} - 2\mathfrak{v}^2_\bullet\nu(a_\bullet\widehat{\theta}_\bullet) = \|a_\bullet\|^2_{\mathfrak{v}} - 2\langle a_\bullet, \widehat{\theta}^m_\bullet\rangle_{\mathfrak{v}} = \|a_\bullet - \widehat{\theta}^m_\bullet\|^2_{\mathfrak{v}} - \|\widehat{\theta}^m_\bullet\|^2_{\mathfrak{v}}.$$

Consequently, for each $m \in \mathbb{N}$ the OPE $\widehat{\theta}^m_\bullet = \dot{\theta}_\bullet\mathbb{1}^m_\bullet$ minimises the contrast function, that is,

$$-\|\widehat{\theta}^m_\bullet\|^2_{\mathfrak{v}} = \Upsilon(\widehat{\theta}^m_\bullet) = \inf\big\{\Upsilon(a_\bullet) : a_\bullet \in \mathbb{L}_2(\mathfrak{v}^2_\bullet\nu)\mathbb{1}^m_\bullet\big\}. \tag{14.07}$$

Given an upper bound $\mathrm{M} \in \mathbb{N}$ and penalties $\mathfrak{pen}_\bullet = (\mathfrak{pen}_m)_{m\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ we select a dimension among the collection of admissible values $[\![\mathrm{M}]\!]$ as minimiser of a penalised contrast criterion, that is

$$\widehat{m} := \arg\min\big\{\Upsilon(\widehat{\theta}^m_\bullet) + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\}. \tag{14.08}$$

The data-driven estimator of $\theta_\bullet$ is now given by $\widehat{\theta}^{\widehat{m}}_\bullet$ and below we derive an upper bound for its global $\mathfrak{v}$-risk $\mathbb{P}^n_q\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_\mathfrak{v}\big)$. The construction of the penalty sequence $\mathfrak{pen}_\bullet$ and the upper bound M given below is motivated by the following key arguments used in the proof of the risk bound which we present first. Moreover, both $\mathfrak{pen}_\bullet$ and M will depend, among others, on the noise level n, however, for sake of simplicity we will omit an additional subscript. The key argument for our reasoning is the next assertion. For $a \in \mathbb{R}$ we write $(a)_+ := a \vee 0$ shortly.

§14.04 **Lemma** *(key argument).* *If* $\mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ *then for all* $\mathrm{M} \in \mathbb{N}$ *and* $m \in [\![\mathrm{M}]\!]$ *we have*

$$\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_\mathfrak{v} \leqslant 3\|\theta^m_\bullet - \theta_\bullet\|^2_\mathfrak{v} + 4\mathfrak{pen}_m + 8\max\Big\{\big(\|\widehat{\theta}^j_\bullet - \theta^j_\bullet\|^2_\mathfrak{v} - \mathfrak{pen}_j/4\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\}.$$

§14.05 **Proof** of Lemma §14.04. is given in the lecture. □

Similar to m°_n as in (14.05), which realises by construction a statistical-error-squared-bias compromise, let us fix a dimension $m^\diamond \in [\![\mathrm{M}]\!]$ to be specified below. Due to the last assertion for each $\theta \in \Theta$ we have

$$\begin{aligned}\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_\mathfrak{v}\big) \leqslant 3\|\theta^{m^\diamond}_\bullet - \theta_\bullet\|^2_\mathfrak{v} + 4\mathfrak{pen}_{m^\diamond} \\ + 8\mathbb{P}^n_\theta\Big(\max\Big\{\big(\|\widehat{\theta}^j_\bullet - \theta^j_\bullet\|^2_\mathfrak{v} - \mathfrak{pen}_j/4\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\Big\}\Big)\end{aligned} \tag{14.09}$$

Keeping in mind that $m^\diamond \in [\![\mathrm{M}]\!]$ in contrast to $m^\circ_n \in \mathbb{N}$ eventually realises an optimal statistical-error-squared-bias trade-off among the collection of admissable values $[\![\mathrm{M}]\!]$ rather than $\mathbb{N}$, we wish the upper bound M to be as large as possible. In contrast, in order to control the remainder term, the last term in (14.09), we are eventually a forced to use a rather small upper bound M. However, we bound the remainder term by imposing the following assumption, which though holds true for a wide range of solutions $\theta_\bullet \in \Theta$ under reasonable model assumptions.

§14.06 **Assumption**. There exists a constant $\mathrm{C} := \mathrm{C}(\theta_\bullet) \in \mathbb{R}^+_{\setminus 0}$ possibly depending on the parameter $\theta_\bullet \in \Theta$ and $(\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}_\bullet))_{n\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ such that for each $n \in \mathbb{N}$ the upper bound $\mathrm{M} \in \mathbb{N}$ and $m^\diamond \in [\![\mathrm{M}]\!]$ satisfy

$$\mathbb{P}^n_\theta\Big(\max\Big\{\big(\|\widehat{\theta}^j_\bullet - \theta^j_\bullet\|^2_\mathfrak{v} - \mathfrak{pen}_j/4\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\Big\}\Big) \leqslant \mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}_\bullet).$$

The next assertion provides an upper bound for the $\mathfrak{v}$-risk of the estimator $\widehat{\theta}^{\widehat{m}}$ with data-driven choice $\widehat{m}$ given by (14.08).

§14.07 **Proposition**. *Let* $m^\diamond \in [\![\mathrm{M}]\!]$ *satisfy the Assumption §14.06 then we have*

$$\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_\mathfrak{v}\big) \leqslant 3\|\theta^{m^\diamond}_\bullet - \theta_\bullet\|^2_\mathfrak{v} + 4\mathfrak{pen}_{m^\diamond} + 8\,\mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}_\bullet).$$

§14.08 **Proof** of Proposition §14.07. is given in the lecture. □

§14.09 **Corollary**. *If* $m^\diamond = \arg\min\big\{\|\theta^m_\bullet - \theta_\bullet\|^2_\mathfrak{v} + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\}$ *satisfies Assumption §14.06, then*

$$\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_\mathfrak{v}\big) \leqslant 4\min\big\{\|\theta^m_\bullet - \theta_\bullet\|^2_\mathfrak{v} + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\} + 8\,\mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}_\bullet).$$

§14.10 **Proof** of Corollary §14.09. is given in the lecture. □

§14.11 **Remark.** Considering the $\mathfrak{v}$-risk bound of the estimator $\widehat{\theta}^{m^\diamond}_\bullet$ with dimension parameter $m^\diamond$ the first rhs. term in the upper risk-bound given in Proposition §14.07 is strongly reminiscent of the variance-squared-bias upper bound $\mathrm{R}^{m^\diamond}_n(\theta_\bullet, \mathfrak{v}_\bullet) = \|\theta^m_\bullet - \theta_\bullet\|^2_{\mathfrak{v}} + \mathfrak{v}\mathrm{ar}_{n,m^\diamond}(\theta_\bullet, \mathfrak{v}_\bullet)$ as given in (14.02). Indeed, in many cases the penalty term $\mathfrak{p}\mathrm{en}_{m^\diamond}$ is in the same order as the statistical error $\mathfrak{v}\mathrm{ar}_{n,m^\diamond}(\theta_\bullet, \mathfrak{v}_\bullet)$. Consequently, provided the reminder term $\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}_\bullet)$ is negligble compared to $\mathrm{R}^{m^\diamond}_n(\theta_\bullet, \mathfrak{v}_\bullet)$, the upper risk bound of the data-driven estimator is given by $\mathrm{R}^{m^\diamond}_n(\theta_\bullet, \mathfrak{v}_\bullet)$ (up to a constant). Moreover, since $m^\diamond$ realises an optimal trade-off between squared-bias and statistical error among the admissable values M, in many cases $\mathrm{R}^{m^\diamond}_n(\theta_\bullet, \mathfrak{v}_\bullet)$ is of optimal orcale order $\mathrm{R}^\circ_n(\theta_\bullet, \mathfrak{v}_\bullet)$. □

We eventually are in a situation where the sequence of penalties $\mathfrak{p}\mathrm{en}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ satisfying the Assumption §14.06 still depends on characteristics of the unknown parameter θ and thus it is only partially known in advance. Assuming a sequence of estimators $\widehat{\mathfrak{p}}\mathrm{en}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ we select similar to (14.08) the dimension

$$\widehat{m} := \arg\min\big\{\Upsilon(\widehat{\theta}^m_\bullet) + \widehat{\mathfrak{p}}\mathrm{en}_m : m \in [\![\mathrm{M}]\!]\big\}. \tag{14.10}$$

The data-driven estimator of $\theta_\bullet$ is now given by $\widehat{\theta}^{\widehat{m}}_\bullet$ and below we derive an upper bound for its global $\mathfrak{v}$-risk $\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}\big)$. The key argument for our reasoning is the next assertion. Its proof follows along the lines of the Proof §14.05.

§14.12 **Lemma** *(key argument). If* $\widehat{\mathfrak{p}}\mathrm{en}_\bullet, \mathfrak{p}\mathrm{en}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ *then for all* $\mathrm{M} \in \mathbb{N}$ *and* $m \in [\![\mathrm{M}]\!]$ *we have*

$$\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}} \leqslant 3\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + 2\mathfrak{p}\mathrm{en}_m + 2\widehat{\mathfrak{p}}\mathrm{en}_m + 2\big(\mathfrak{p}\mathrm{en}_{\widehat{m}} - \widehat{\mathfrak{p}}\mathrm{en}_{\widehat{m}}\big)_+ + 8\max\Big\{\big(\|\widehat{\theta}^j_\bullet - \theta^j_\bullet\|^2_{\mathfrak{v}} - \mathfrak{p}\mathrm{en}_j/4\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\}.$$

§14.13 **Proof** of Lemma §14.12. is given in the lecture. □

Similar to m°_n as in (14.02), which realises by construction a statistical-error-squared-bias compromise, let us fix a dimension $m^\diamond \in [\![\mathrm{M}]\!]$ to be specified below (analogously to (14.09)). Due to the last assertion for each $\theta \in \Theta$ we have

$$\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}\big) \leqslant 3\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + 2\mathfrak{p}\mathrm{en}_m + 8\mathbb{P}^n_\theta\Big(\max\Big\{\big(\|\widehat{\theta}^j_\bullet - \theta^j_\bullet\|^2_{\mathfrak{v}} - \mathfrak{p}\mathrm{en}_j/4\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\}\Big) + 2\mathbb{P}^n_\theta\big(\widehat{\mathfrak{p}}\mathrm{en}_m\big) + 2\mathbb{P}^n_\theta\Big(\big(\mathfrak{p}\mathrm{en}_{\widehat{m}} - \widehat{\mathfrak{p}}\mathrm{en}_{\widehat{m}}\big)_+\Big). \tag{14.11}$$

We bound the first remainder term by imposing Assumption §14.06, which though hold true for a wide range of solutions $\theta_\bullet = \mathrm{U}\theta \in \Theta$ under reasonable model assumptions.

§14.14 **Proposition.** *If* $m^\diamond \in [\![\mathrm{M}]\!]$ *satisfies the Assumption §14.06 then we have*

$$\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}\big) \leqslant 3\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + 2\mathfrak{p}\mathrm{en}_m + 8\mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{v}) + 2\mathbb{P}^n_\theta\big(\widehat{\mathfrak{p}}\mathrm{en}_m\big) + 2\mathbb{P}^n_\theta\Big(\big(\mathfrak{p}\mathrm{en}_{\widehat{m}} - \widehat{\mathfrak{p}}\mathrm{en}_{\widehat{m}}\big)_+\Big).$$

§14.15 **Proof** of Proposition §14.14. is given in the lecture. □

§14.16 **Corollary.** *If* $m^\diamond = \arg\min\big\{\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + \mathfrak{p}\mathrm{en}_m : m \in [\![\mathrm{M}]\!]\big\}$ *satisfies Assumption §14.06 with constant* $\mathrm{C} \in [1, \infty)$, $\mathbb{P}^n_\theta(\widehat{\mathfrak{p}}\mathrm{en}_{m^\diamond}) \leqslant \mathrm{K}_1 \mathfrak{p}\mathrm{en}_{m^\diamond}$ *and* $\mathbb{P}^n_\theta\big(\big(\mathfrak{p}\mathrm{en}_{\widehat{m}} - \widehat{\mathfrak{p}}\mathrm{en}_{\widehat{m}}\big)_+\big) \leqslant \mathrm{K}_2\,\mathrm{R}^{\mathrm{re}}_n(\theta)$ *for some* $\mathrm{K}_1, \mathrm{K}_2 \in [1, \infty)$, *then*

$$\mathbb{P}^n_\theta\big(\|\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet\|^2_{\mathfrak{v}}\big) \leqslant 4\mathrm{K}_1 \min\big\{\|\theta_\bullet \mathbb{1}^{m|\perp}_\bullet\|^2_{\mathfrak{v}} + \mathfrak{p}\mathrm{en}_m : m \in [\![\mathrm{M}]\!]\big\} + (2\mathrm{K}_2 + 8\mathrm{C})\mathrm{R}^{\mathrm{re}}_n(\theta).$$

§14.17 **Proof** of Corollary §14.16. is given in the lecture. □

§14|03 GSSM: data-driven global estimation

Let us first state some elementary inequalities for Gaussian random variables. There exist several results for tail bounds of sums of independent squared Gaussian random variables and we present next a version which is due to Birgé [2001] and the formulation (14.12) can be found in Lemma 1 in Laurent and Massart [2000].

§14.18 **Lemma.** *Let* $a_\bullet \in (\mathbb{R})^{\mathbb{N}}$ *and* $\dot{\mathrm{B}}_\bullet = (\dot{\mathrm{B}}_j)_{j\in\mathbb{N}} \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. *For all* $\eta \in \mathbb{R}^+$ *and* $m \in \mathbb{N}$ *we have*

$$\mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}\big(\|a_\bullet\dot{\mathrm{B}}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2 - \|a_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2 \geqslant 2\|a_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_2}\sqrt{\eta} + 2\|a_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}\eta\big) \leqslant \exp(-\eta). \tag{14.12}$$

which for all $\zeta \in \mathbb{R}^+$ *setting* $\eta := \zeta(\zeta\wedge 1)\|a_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2/(4\|a_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}) \in \mathbb{R}^+$ *implies*

$$\mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}\big(\|a_\bullet\dot{\mathrm{B}}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2 \geqslant (1+3\zeta/2)\|a_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2\big) \leqslant \exp(-\eta). \tag{14.13}$$

Moreover, for any $\xi \in [1,\infty)$ *we have*

$$\mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}\big(\big(\|a_\bullet\dot{\mathrm{B}}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2 - (1+3\xi/2)\|a_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2\big)_+\big) \leqslant 6\|a_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}\exp\big(-(\xi/4)\|a_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2\|a_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}^{-1}\big). \tag{14.14}$$

§14.19 **Proof** of Lemma §14.18. Exercise. □

§14|03|01 Global $\mathfrak{v}$-risk

§14.20 **Reminder** *(Global oracle* $\mathfrak{v}$*-risk in* GSSM *§14.02).* Given Model §14.02 we consider an OPE as in Section §12. Here the observable noisy version $\widehat{\theta}_\bullet$ admits a N_θ^n-distribution belonging to the family $\mathrm{N}_\Theta^n := (\mathrm{N}_\theta^n)_{\theta\in\Theta}$, $\Theta \subseteq \ell_2$. Let us recall (12.04) in Proposition §12.12 where for $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\theta_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ and $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}_n^m(\theta_\bullet,\mathfrak{v}_\bullet) := \|\theta_\bullet\mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(\theta_\bullet,\mathfrak{v}_\bullet) : m\in\mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(\theta_\bullet,\mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\circ}(\theta_\bullet,\mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\theta_\bullet,\mathfrak{v}_\bullet) : m\in\mathbb{N}\big\}. \tag{14.15}$$

Due to Corollary §12.17 the (infeasible) OPE $\widehat{\theta}_\bullet^{m_n^\circ} = \widehat{\theta}_\bullet\mathbb{1}_\bullet^{m_n^\circ} \in \ell_2\mathbb{1}_\bullet^{m_n^\circ} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with oracle dimension m_n° as in (14.15) satisfies

$$\mathrm{N}_\theta^n\big(\|\widehat{\theta}_\bullet^{m_n^\circ} - \theta_\bullet\|_{\mathfrak{v}}^2\big) = \mathrm{R}_n^\circ(\theta_\bullet,\mathfrak{v}_\bullet) = \inf_{m\in\mathbb{N}}\mathrm{N}_\theta^n\big(\|\widehat{\theta}_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2\big),$$

and hence it is *oracle optimal* (with constant 1). □

§14.21 **Assumption.** Let $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ satisfy

$$\mathrm{C}_{\mathfrak{v}} := \sum_{m\in\mathbb{N}} 4\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}\exp\big(-\|\mathfrak{v}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2/(4\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty})\big) \in \mathbb{R}^+. \qquad \square$$

§14.22 **Comment.** Since $\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty} \in \mathbb{R}_{\setminus 0}^+$ and $\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty} \geqslant \mathfrak{v}_1^2$ for all $m \in \mathbb{N}$, the Assumption §14.21 implies $\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}\|\mathfrak{v}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^{-2} = \mathrm{o}(1)$ and $\|\mathfrak{v}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^{-2} = \mathrm{o}(1)$ as $m\to\infty$. □

§14.23 **Illustration.** Consider $\mathfrak{v}_\bullet^2 = (j^a)_{j\in\mathbb{N}}$ for $a \in \mathbb{R}_{\setminus 0}^+$. Then $\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty} = m^a$ and $\|\mathfrak{v}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2 \simeq m^{a+1}$. Consequently, we have

$$\sum_{m\in\mathbb{N}} 4\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty}\exp\big(-\|\mathfrak{v}_\bullet\mathbb{1}_\bullet^m\|_{\ell_2}^2/(4\|\mathfrak{v}_\bullet^2\mathbb{1}_\bullet^m\|_{\ell_\infty})\big) \simeq \sum_{m\in\mathbb{N}} m^a\exp\big(-m\big) \in \mathbb{R}_{\setminus 0}^+.$$

and the assumption Assumption §14.21 is satisfied. On the contrary if $\mathfrak{v}_\bullet^2 = (\exp(j^a))_{j\in\mathbb{N}}$ for $a > 1$, then $\|\mathfrak{v}_\bullet^2 \mathbb{1}_\bullet^m\|_{\ell_\infty} = \exp(m^a)$ and $\|\mathfrak{v}_\bullet \mathbb{1}_\bullet^m\|_{\ell_2}^2 \simeq \exp(m^a)$. Consequently, we have

$$\sum_{m\in\mathbb{N}} 4\|\mathfrak{v}_\bullet^2 \mathbb{1}_\bullet^m\|_{\ell_\infty} \exp\big(-\|\mathfrak{v}_\bullet \mathbb{1}_\bullet^m\|_{\ell_2}^2/(4\|\mathfrak{v}_\bullet^2 \mathbb{1}_\bullet^m\|_{\ell_\infty})\big) \simeq \sum_{m\in\mathbb{N}} \exp(m^a) \simeq \infty$$

and the assumption Assumption §14.21 is not satisfied. □

§14.24 **Corollary**. *Under Assumption §14.21 we have*

$$\mathrm{N}_{\mathfrak{q}}^n\big(\max\big\{\big(\|\widehat{\theta}_\bullet^j - \theta_\bullet^j\|_{\mathfrak{v}}^2 - \tfrac{5}{2}\|\mathbb{1}_\bullet^j\|_{\mathfrak{v}}^2 n^{-1}\big)_+ : j \in [\![\mathrm{M}]\!]\big\}\big) \leqslant (3/2)\mathrm{C}_{\mathfrak{v}} n^{-1} \quad \textit{for all } \mathrm{M} \in \mathbb{N}. \tag{14.16}$$

§14.25 **Proof** of Corollary §14.24. is given in the lecture. □

§14.26 **Notation**. Consider a sequence of penalties $\mathfrak{pen}_\bullet^{\mathfrak{v}} = (\mathfrak{pen}_m^{\mathfrak{v}})_{m\in\mathbb{N}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ given by

$$\mathfrak{pen}_m^{\mathfrak{v}} := 10\, n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad \text{for each } m \in \mathbb{N} \tag{14.17}$$

which is obviously known in advance. Considering the data-driven OSE $\widehat{\theta}_\bullet^{\widehat{m}} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter $\widehat{m}$ selected as in (14.08) with penalty sequence $\mathfrak{pen}_\bullet^{\mathfrak{v}}$ given in (14.17) and arbitrary but fixed upper bound $\mathrm{M} \in \mathbb{N}$ we derive below an upper bound for its global $\mathfrak{v}$-risk, $\mathrm{N}_{\mathfrak{q}}^n\big(\|\widehat{\theta}_\bullet^{\widehat{m}} - \theta_\bullet\|_{\mathfrak{v}}^2\big)$. □

§14.27 **Proposition** *(GSSM (§14.02 continued)). Let $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\mathfrak{q}}^n$ as in Model §14.02 where $\theta_\bullet \in \ell_2$ and $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Given $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M} \in \mathbb{N}$ and $\mathfrak{pen}_\bullet^{\mathfrak{v}}$ as in (14.17) consider a data-driven OPE $\widehat{\theta}_\bullet^{\widehat{m}} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2 \mathbb{1}_\bullet^{\widehat{m}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ of $\theta_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ with*

$$\widehat{m} := \arg\min\big\{-\|\widehat{\theta}_\bullet^m\|_{\mathfrak{v}} + \mathfrak{pen}_m^{\mathfrak{v}} : m \in [\![\mathrm{M}]\!]\big\}. \tag{14.18}$$

If Assumption §14.21 is satisfied with $\mathrm{C}_{\mathfrak{v}} \in \mathbb{R}_{\setminus 0}^+$, then for all $n, \mathrm{M} \in \mathbb{N}$ we have

$$\mathrm{N}_{\mathfrak{q}}^n\big(\|\widehat{\theta}_\bullet^{\widehat{m}} - \theta_\bullet\|_{\mathfrak{v}}^2\big) \leqslant 40 \min\big\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) : m \in [\![\mathrm{M}]\!]\big\} + 12\,\mathrm{C}_{\mathfrak{v}}\, n^{-1} \tag{14.19}$$

where $\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) := \|\theta_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$ is defined as in (14.15).

§14.28 **Proof** of Proposition §14.27. is given in the lecture. □

§14.29 **Comment**. The oracle bound $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) = \mathrm{R}_n^{m_n^\circ}(\theta_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\theta_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$ (for details see Reminder §14.20) satisfies $n\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet) \geqslant \|\mathbb{1}_\bullet^{m_n^\circ}\|_{\mathfrak{v}}^2 \geqslant \mathfrak{v}_1^2$. Consequently, the last upper bound in (14.19) and the oracle bound $\mathrm{R}_n^\circ(\theta_\bullet, \mathfrak{v}_\bullet)$ coincide up to a constant $(40 + 12\,\mathrm{C}_{\mathfrak{v}}\mathfrak{v}_1^{-2})$ provided the oracle dimension fulfils $m_n^\circ \in [\![\mathrm{M}]\!]$. Therefore, we wish the upper bound M to be as large as possible. The next assertion shows that

$$\mathrm{M}^{\mathfrak{v}} := \max\big\{m \in \mathbb{N} : \|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \leqslant n\mathfrak{v}_1^2\big\} \in \mathbb{N} \tag{14.20}$$

is a suitable choice for the upper bound, where the defining set is not empty and finite since $\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^{-2} = \mathrm{o}(1)$ as $m \to \infty$. □

§14.30 **Corollary** *(GSSM (§14.02 continued)). Given $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ as in (14.20) and $\mathfrak{pen}_\bullet^{\mathfrak{v}}$ as in (14.17) consider a data-driven OPE $\widehat{\theta}_\bullet^{\widehat{m}} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2 \mathbb{1}_\bullet^{\widehat{m}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with*

$$\widehat{m} := \arg\min\big\{-\|\widehat{\theta}_\bullet^m\|_{\mathfrak{v}} + \mathfrak{pen}_m^{\mathfrak{v}} : m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\big\}. \tag{14.21}$$

Under the assumptions of Proposition §14.27 for each $n \in \mathbb{N}$ such that $\mathrm{R}^{\circ}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) \leqslant \mathfrak{v}_{1}^{2}$ we have

$$\mathrm{N}^{n}_{\theta}\big(\|\widehat{\theta}^{\widehat{m}}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}}\big) \leqslant 40\,\mathrm{R}^{\circ}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) + 12\,\mathrm{C}_{\mathfrak{v}}\,n^{-1} \leqslant \mathrm{C}\mathrm{R}^{\circ}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) \tag{14.22}$$

and, hence up to the constant $\mathrm{C} := 40 + 12\,\mathrm{C}_{\mathfrak{v}}\mathfrak{v}_{1}^{-2}$ the feasible data-driven estimator $\widehat{\theta}^{\widehat{m}}_{\bullet}$ is oracle optimal.

§14.31 **Proof** of Corollary §14.30. is given in the lecture. □

§14.32 **Remark**. If Assumption §14.21 is not satisfied (see Illustration §14.23), then we can't make use of Corollary §14.24. In this situation let $\mathfrak{v}_{\bullet} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $\delta_{\bullet} \in ([1, \infty))^{\mathbb{N}}$ satisfy

$$\mathrm{C}_{\mathfrak{v},\delta} := \sum_{m \in \mathbb{N}} 4\|\mathfrak{v}^{2}_{\bullet}\mathbb{1}^{m}_{\bullet}\|_{\ell_{\infty}} \exp\big(-\delta_{m}\|\mathfrak{v}_{\bullet}\mathbb{1}^{m}_{\bullet}\|^{2}_{\ell_{2}}/(4\|\mathfrak{v}^{2}_{\bullet}\mathbb{1}^{m}_{\bullet}\|_{\ell_{\infty}})\big) \in \mathbb{R}^{+}. \tag{14.23}$$

Consider a sequence of penalties $\mathfrak{pen}^{\mathfrak{v},\delta}_{\bullet} = (\mathfrak{pen}^{\mathfrak{v},\delta}_{m})_{m \in \mathbb{N}} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ given by

$$\mathfrak{pen}^{\mathfrak{v},\delta}_{m} := 4(1 + 3\delta_{m}/2)\,n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}}, \qquad \text{for each } m \in \mathbb{N} \tag{14.24}$$

which is obviously known in advance. Similar to Corollary §14.24 due to (14.23) for each $\mathrm{M} \in \mathbb{N}$ we obtain

$$\begin{aligned} n\mathrm{N}^{n}_{\theta}\big(\max\big\{\big(\|\widehat{\theta}^{j}_{\bullet} - \theta^{j}_{\bullet}\|^{2}_{\mathfrak{v}} - (1 + 3\delta_{j}/2)\|\mathbb{1}^{j}_{\bullet}\|^{2}_{\mathfrak{v}}n^{-1}\big)_{+} : j \in [\![\mathrm{M}]\!]\big\}\big) \\ \leqslant \sum_{m \in [\![\mathrm{M}]\!]} 6\|\mathfrak{v}^{2}_{\bullet}\mathbb{1}^{m}_{\bullet}\|_{\ell_{\infty}} \exp\big(-\delta_{m}\|\mathfrak{v}_{\bullet}\mathbb{1}^{m}_{\bullet}\|^{2}_{\ell_{2}}/(4\|\mathfrak{v}^{2}_{\bullet}\mathbb{1}^{m}_{\bullet}\|_{\ell_{\infty}})\big) = (3/2)\mathrm{C}_{\mathfrak{v},\delta}. \end{aligned}$$

Thus, the sequence of penalties $\mathfrak{pen}^{\mathfrak{v},\delta}_{\bullet} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ given in (14.24) satisfies the Assumption §14.06 with $\mathrm{C} = (3/2)\mathrm{C}_{\mathfrak{v},\delta}$ and $\mathrm{R}^{\mathrm{re}}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) = n^{-1}$. Consequently, the data-driven OPE $\widehat{\theta}^{\widehat{m}}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_{2}\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \ell_{2}(\mathfrak{v}^{2}_{\bullet})$ with

$$\widehat{m} := \arg\min\big\{-\|\widehat{\theta}^{m}_{\bullet}\|_{\mathfrak{v}} + \mathfrak{pen}^{\mathfrak{v},\delta}_{m} : m \in [\![\mathrm{M}]\!]\big\} \tag{14.25}$$

due to Proposition §14.07 for all $\theta_{\bullet} \in \ell_{2}(\mathfrak{v}^{2}_{\bullet})$ and $n, \mathrm{M} \in \mathbb{N}$ fulfils

$$\begin{aligned} \mathrm{N}^{n}_{\theta}\big(\|\widehat{\theta}^{\widehat{m}}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}}\big) &\leqslant \min\big\{3\|\theta^{m}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}} + 4\mathfrak{pen}^{\mathfrak{v},\delta}_{m} : m \in [\![\mathrm{M}]\!]\big\} + 8(3/2)\mathrm{C}_{\mathfrak{v},\delta}n^{-1} \\ &\leqslant 40\min\big\{\|\theta^{m}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}} + n^{-1}\delta_{m}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}} : m \in [\![\mathrm{M}]\!]\big\} + 12\mathrm{C}_{\mathfrak{v},\delta}n^{-1} \end{aligned}$$

Introduce $\mathrm{R}^{\diamond}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) := \min\big\{\|\theta^{m}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}} + n^{-1}\delta_{m}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}} : m \in \mathbb{N}\big\}$ where $\mathrm{R}^{\diamond}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) \geqslant \mathrm{R}^{\circ}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet}) \geqslant n^{-1}\mathfrak{v}_{1}^{2}$ since in general $n^{-1}\delta_{m}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}} \geqslant n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}}$ for all $m \in \mathbb{N}$. Consequently, if the upper bound $\mathrm{M} \in \mathbb{N}$ satisfies $\arg\min\big\{\|\theta^{m}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}} + n^{-1}\delta_{m}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}} : m \in \mathbb{N}\big\} =: m^{\diamond} \in [\![\mathrm{M}]\!]$ then we obtain $\mathrm{N}^{n}_{\theta}\big(\|\widehat{\theta}^{\widehat{m}}_{\bullet} - \theta_{\bullet}\|^{2}_{\mathfrak{v}}\big) \leqslant \mathrm{C}\mathrm{R}^{\diamond}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet})$ with $\mathrm{C} := 40 + 12\mathrm{C}_{\mathfrak{v},\delta}$. However, the upper bound $\mathrm{R}^{\diamond}_{n}(\theta_{\bullet}, \mathfrak{v}_{\bullet})$ faces a deterioration by the factor $\delta_{\bullet}$ and thus it is generally not an oracle bound. □

§14|03|02 Maximal global $\mathfrak{v}$-risk

§14.33 **Reminder** *(Maximal global $\mathfrak{v}$-risk in* GSSM *§14.02).* Given Model §14.02 we consider an OPE as in Section §12. Here the observable noisy version $\widehat{\theta}_{\bullet}$ admits a N^{n}_{θ}-distribution belonging to the family $\mathrm{N}^{n}_{\Theta} := (\mathrm{N}^{n}_{\theta})_{\theta \in \Theta}$, $\Theta \subseteq \ell_{2}$. Under Assumption §11.12 in Corollary §12.24 an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE is shown. More precisely, the performance of the OPE

$\widehat{\theta}^{m}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{m}_{\bullet} \in \ell_2\mathbb{1}^{m}_{\bullet} \subseteq \ell_2(\mathfrak{v}^2_{\bullet})$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}_{\bullet}$-risk over the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}^{\mathfrak{v}}_{n}[\,\widehat{\theta}^{m}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathrm{N}^{n}_{\theta}\big(\|\widehat{\theta}^{m}_{\bullet} - \theta_{\bullet}\|^2_{\mathfrak{v}}\big)\colon \theta_{\bullet} \in \ell_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.06) where for $n, m \in \mathbb{N}$ we have defined $(\mathfrak{a}\mathfrak{v})^2_{(m)} := \|(\mathfrak{a}\mathfrak{v})^2_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|_{\ell_\infty}$ and

$$\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) := (\mathfrak{a}\mathfrak{v})^2_{(m)} \vee n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\mathfrak{v}}, \quad m^{\star}_{n} := \arg\min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) := \mathrm{R}^{m^{\star}_{n}}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) = \min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) : m \in \mathbb{N}\big\}. \quad (14.26)$$

By Corollary §12.24 under Assumption §11.12 the maximal global $\mathfrak{v}$-risk of an OPE $\widehat{\theta}^{m^{\star}_{n}}_{\bullet}$ with optimally choosen dimension $m^{\star}_{n}$ as in (14.26) satisfies

$$\mathcal{R}^{\mathfrak{v}}_{n}[\,\widehat{\theta}^{m^{\star}_{n}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})$$

with $\mathrm{C} = 1 + \mathrm{r}^2$. Moreover, under Assumption §13.35 due to Proposition §13.37 $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})$ provides (up to a constant) also a lower bound of the maximal global $\mathfrak{v}$-risk over the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})$ is a minimax bound and $\widehat{\theta}^{m^{\star}_{n}}_{\bullet}$ is minimax optimal. However, the optimal dimension $m^{\star}_{n}$ depends on $\mathfrak{a}_{\bullet} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ characterising the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$. □

§14.34 **Proposition** *(GSSM (§14.02 continued)).* *Let $\widehat{\theta}_{\bullet} = \theta_{\bullet} + n^{-1/2}\dot{\mathrm{B}}_{\bullet} \sim \mathrm{N}^{n}_{\theta}$ as in Model §14.02 where $\theta_{\bullet} \in \ell_2$ and $\dot{\mathrm{B}}_{\bullet} \sim \mathrm{N}^{\otimes\mathbb{N}}_{(0,1)}$. Given $\mathfrak{v}_{\bullet} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M} \in \mathbb{N}$ and $\mathfrak{pen}^{\mathfrak{v}}_{\bullet}$ as in (14.17) consider a data-driven OPE $\widehat{\theta}^{\widehat{m}}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_2\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \ell_2(\mathfrak{v}^2_{\bullet})$ with $\widehat{m}$ as in (14.18). If Assumptions §11.12 and §14.21 (with $\mathrm{C}_{\mathfrak{v}} \in \mathbb{R}^{+}_{\setminus 0}$) are satisfied, then for all $n, \mathrm{M} \in \mathbb{N}$ we have*

$$\mathcal{R}^{\mathfrak{v}}_{n}[\,\widehat{\theta}^{\widehat{m}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant (3\mathrm{r}^2 + 40)\min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})\colon m \in [\![\mathrm{M}]\!]\big\} + 12\,\mathrm{C}_{\mathfrak{v}}\,n^{-1} \quad (14.27)$$

where $\mathrm{R}^{m}_{n}(\theta_{\bullet},\mathfrak{v}_{\bullet}) := (\mathfrak{a}\mathfrak{v})^2_{(m)} \vee n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\mathfrak{v}}$ is defined as in (14.26).

§14.35 **Proof** of Proposition §14.34. is given in the lecture. □

§14.36 **Comment**. The minimax bound $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) = \mathrm{R}^{m^{\star}_{n}}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) = \min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})\colon m \in \mathbb{N}\big\}$ (for details see Reminder §14.33) satisfies $n\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) \geqslant \|\mathbb{1}^{m^{\star}_{n}}_{\bullet}\|^2_{\mathfrak{v}} \geqslant \mathfrak{v}_1^2$. Consequently, the last upper bound in (14.27) and the minimax bound $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet})$ coincide up to a constant $(3\mathrm{r}^2 + 40 + 12\,\mathrm{C}_{\mathfrak{v}}\mathfrak{v}_1^{-2})$ provided the minimax dimension fulfils $m^{\star}_{n} \in [\![\mathrm{M}]\!]$. Therefore, we wish the upper bound M to be as large as possible. The next assertion shows that $\mathrm{M}^{\mathfrak{v}}$ as in (14.20) is a suitable choice for the upper bound. □

§14.37 **Corollary** *(GSSM (§14.02 continued)).* *Given $\mathfrak{v}_{\bullet} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ as in (14.20) and $\mathfrak{pen}^{\mathfrak{v}}_{\bullet}$ as in (14.17) consider a data-driven OPE $\widehat{\theta}^{\widehat{m}}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_2\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \ell_2(\mathfrak{v}^2_{\bullet})$ with $\widehat{m}$ as in (14.21). Under the assumptions of Proposition §14.34 for each $n \in \mathbb{N}$ such that $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) \leqslant \mathfrak{v}_1^2$ we have*

$$\mathcal{R}^{\mathfrak{v}}_{n}[\,\widehat{\theta}^{\widehat{m}^{\star}_{n}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant (3\mathrm{r}^2 + 40)\,\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) + 12\,\mathrm{C}_{\mathfrak{v}}\,n^{-1} \leqslant \mathrm{C}\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\mathfrak{v}_{\bullet}) \quad (14.28)$$

and, hence up to the constant $\mathrm{C} := 3\mathrm{r}^2 + 40 + 12\,\mathrm{C}_{\mathfrak{v}}\mathfrak{v}_1^{-2}$ the feasible data-driven estimator $\widehat{\theta}^{\widehat{m}}_{\bullet}$ is minimax optimal.

§14.38 **Proof** of Corollary §14.37. is given in the lecture. □

§14|04 Goldenshluger and Lepskij's method

The next selection method is inspired by a bandwidth selection method in kernel density estimation proposed in Goldenshluger and Lepskij [2011]. Let us consider a probability measure $\mathbb{P}_\theta^n$ for some $\theta_\bullet \in \Theta$. We shall measure the accuracy of the estimator $\widehat{\theta}_\bullet^m$ of $\theta_\bullet$ by its risk $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^m, \theta_\bullet)\big)$ where $\mathfrak{d}_{\text{ist}}(\cdot,\cdot)$ is a certain semi metric to be specified below. Inspired by Lepskij's method (which appeared in a series of papers by Lepskij [1990, 1991, 1992a,b]) given an integer $\mathrm{M} \in \mathbb{N}$ and a sequence $\mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ of penalties we define a contrast $\mathfrak{contr}_\bullet \in (\mathbb{R}^+)^{[\![\mathrm{M}]\!]}$ by

$$\begin{aligned}\mathfrak{contr}_m &:= \max\Big\{\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^j, \widehat{\theta}_\bullet^m) - \mathfrak{pen}_j - \mathfrak{pen}_m\big)_+ : j \in]\!]m, \mathrm{M}]\!]\Big\} \\ &= \max\Big\{\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{m\vee j}, \widehat{\theta}_\bullet^{m\wedge j}) - \mathfrak{pen}_{m\vee j} - \mathfrak{pen}_{m\wedge j}\big)_+ : j \in]\!]m, \mathrm{M}]\!]\Big\}, \quad m \in [\![\mathrm{M}]\!]. \qquad (14.29)\end{aligned}$$

In the spirit of Goldenshluger and Lepskij [2011] combining the contrast given in (14.29) and the penalisation approach of model selection in Subsection §14|02 we select the dimension

$$\widehat{m} := \arg\min\big\{\mathfrak{contr}_m + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\}. \qquad (14.30)$$

The data-driven estimator of $\theta_\bullet$ is now given by $\widehat{\theta}_\bullet^{\widehat{m}}$ and below we derive an upper bound for its risk $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big)$. The construction of the penalty sequence $\mathfrak{pen}_\bullet$ and the upper bound M given below is motivated by the following key argument used in the proof of the risk bound which we present first. Moreover, both $\mathfrak{pen}_\bullet$ and M will depend, among others, on the noise level n, however, for sake of simplicity we will omit an additional subscript. The key argument for our reasoning is the next assertion.

§14.39 **Lemma** *(key argument).* *Let* $\mathfrak{bias}_\bullet(\theta_\bullet, \mathfrak{d}_{\text{ist}}) = \big(\mathfrak{bias}_m(\theta_\bullet, \mathfrak{d}_{\text{ist}})\big)_{m\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ *be defined by*

$$\mathfrak{bias}_m(\theta_\bullet, \mathfrak{d}_{\text{ist}}) := \sup\big\{\mathfrak{d}_{\text{ist}}(\theta_\bullet^j, \theta_\bullet^m) : j \in [\![m, \infty]\!] := \mathbb{N} \cap [m, \infty) \cup \{\infty\}\big\}, \quad \forall m \in \mathbb{N}. \qquad (14.31)$$

If $\mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ *then for all* $\mathrm{M} \in \mathbb{N}$ *and* $m \in [\![\mathrm{M}]\!]$ *we have*

$$\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet) \leqslant 16\mathfrak{bias}_m^2(\theta_\bullet, \mathfrak{d}_{\text{ist}}) + \tfrac{16}{3}\mathfrak{pen}_m + 28\max\Big\{\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^j, \theta_\bullet^j) - \mathfrak{pen}_j/3\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\}.$$

§14.40 **Proof** of Lemma §14.39. is given in the lecture. □

Similar to m_n° as in (14.02), which realises by construction a statistical-error-squared-bias compromise, let us fix a dimension $m^\diamond \in [\![\mathrm{M}]\!]$ to be specified below. Due to the last assertion for each $\theta \in \Theta$ we have

$$\begin{aligned}\mathbb{P}_\theta^n\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) &\leqslant 16\mathfrak{bias}_{m^\diamond}^2(\theta_\bullet, \mathfrak{d}_{\text{ist}}) + \tfrac{16}{3}\mathfrak{pen}_{m^\diamond} \\ &\quad + 28\mathbb{P}_\theta^n\big(\max\big\{\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^j, \theta_\bullet^j) - \mathfrak{pen}_j/3\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\big\}\big). \qquad (14.32)\end{aligned}$$

Keeping in mind that $m^\diamond$ in contrast to m_n° eventually realises an optimal statistical-error-squared-bias trade-off among the collection of admissable values $[\![\mathrm{M}]\!]$ rather than $\mathbb{N}$, we wish the upper bound M to be as large as possible. In contrast, in order to control the remainder term, the last term in (14.32), we a forced to use a rather small upper bound M. However, we bound the remainder term by imposing a condition similar to Assumption §14.06, which though holds true for a wide range of solutions $\theta_\bullet = \mathrm{U}\theta \in \Theta$ under reasonable model assumptions.

§14.41 **Assumption**. There exists a constant $\mathrm{C} := \mathrm{C}(\theta_\bullet) \in \mathbb{R}_{\setminus 0}^+$ and $(\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}))_{n\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ possibly depending on the parameter $\theta_\bullet = \mathrm{U}\theta \in \Theta$ such that for each $n \in \mathbb{N}$ the upper bound $\mathrm{M} \in \mathbb{N}$ and $m^\diamond \in [\![\mathrm{M}]\!]$ satisfy

$$\mathbb{P}_\theta^n\big(\max\big\{\big(\mathfrak{d}_{\text{ist}}^2(\widehat{\theta}_\bullet^m, \theta_\bullet^m) - \mathfrak{pen}_m/3\big)_+ : m \in [\![m^\diamond, \mathrm{M}]\!]\big\}\big) \leqslant \mathrm{C}\,\mathrm{R}_n^{\text{re}}(\theta_\bullet, \mathfrak{d}_{\text{ist}}). \qquad (14.33)$$

The next assertion provides an upper bound for the risk of the estimator $\widehat{\theta}^{\widehat{m}}$ with data-driven choice $\widehat{m}$ given by (14.30).

§14.42 **Proposition.** *Let $m^\diamond \in [\![\mathrm{M}]\!]$ satisfy the Assumption §14.41 then we have*

$$\mathbb{P}_\theta^n\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) \leqslant 16\mathfrak{bias}_{m^\diamond}^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \tfrac{16}{3}\mathfrak{pen}_{m^\diamond} + 28\mathrm{C}\,\mathrm{R}_n^{\mathrm{re}}(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}).$$

§14.43 **Proof** of Proposition §14.42. is given in the lecture. □

§14.44 **Corollary.** *If $m^\diamond = \arg\min\big\{\mathfrak{bias}_m^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\}$ satisfies Assumption §14.41, then*

$$\mathbb{P}_\theta^n\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) \leqslant 16\min\big\{\mathfrak{bias}_m^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\} + 28\,\mathrm{C}\,\mathrm{R}_n^{\mathrm{re}}(\theta).$$

§14.45 **Proof** of Corollary §14.44. is given in the lecture. □

§14.46 **Comment.** Considering a global $\mathfrak{v}$-error we note that $\mathfrak{bias}_m^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) = \|\theta_\bullet^m - \theta_\bullet\|_{\mathfrak{v}}^2$ for all $m \in \mathbb{N}$, and hence the upper bound in Corollary §14.44 equals up to the numerical constants the upper bound in Corollary §14.09 using a model selection approach (Subsection §14|02). Consequently, when globally estimating the parameter in a *GSSM* with a Goldenshluger and Lepskij method rather than a model selection approach as in Subsection §14|03 we eventually obtain the same upper bounds (up to the numerical constants). However, we shall stress that in opposite to model selection the method by Goldenshluger and Lepskij does not require, that the estimator minimises a contrast function. □

We eventually are in a situation where the sequence of penalties $\mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ satisfying the Assumption §14.41 still depends on characteristics of the unknown parameter θ and thus it is only partially known in advance. Assuming a sequence of estimators $\widehat{\mathfrak{pen}}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ we define an estimated contrast $\widehat{\mathfrak{contr}}_\bullet \in (\mathbb{R}^+)^{[\![\mathrm{M}]\!]}$ by

$$\begin{aligned}\widehat{\mathfrak{contr}}_m &:= \max\Big\{\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^j, \widehat{\theta}_\bullet^m) - \widehat{\mathfrak{pen}}_j - \widehat{\mathfrak{pen}}_m\big)_+ : j \in]\!]m, \mathrm{M}]\!]\Big\} \\ &= \max\Big\{\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{m\vee j}, \widehat{\theta}_\bullet^{m\wedge j}) - \widehat{\mathfrak{pen}}_{m\vee j} - \widehat{\mathfrak{pen}}_{m\wedge j}\big)_+ : j \in]\!]m, \mathrm{M}]\!]\Big\}, \quad m \in [\![\mathrm{M}]\!] \qquad (14.34)\end{aligned}$$

and similar to (14.30) we select the dimension

$$\widehat{m} := \arg\min\big\{\widehat{\mathfrak{contr}}_m + \widehat{\mathfrak{pen}}_m : m \in [\![\mathrm{M}]\!]\big\}. \qquad (14.35)$$

The data-driven estimator of $\theta_\bullet$ is now given by $\widehat{\theta}_\bullet^{\widehat{m}}$ and below we derive an upper bound for its risk $\mathbb{P}_\theta^n\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big)$. The key argument for our reasoning is the next assertion. Its proof follows along the lines of the Proof §14.40.

§14.47 **Lemma** *(key argument). Let $\mathfrak{bias}_\bullet(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) = (\mathfrak{bias}_m(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}))_{m\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ be defined as in (14.31) (Lemma §14.39). If $\widehat{\mathfrak{pen}}_\bullet, \mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ then for all $\mathrm{M} \in \mathbb{N}$ and $m \in [\![\mathrm{M}]\!]$ we have*

$$\begin{aligned}\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet) \leqslant 16\mathfrak{bias}_m^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \tfrac{4}{3}\mathfrak{pen}_m + 28\max\Big\{\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^j, \theta_\bullet^j) - \mathfrak{pen}_j/3\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\} \\ + 8\max\Big\{\big(\mathfrak{pen}_j - \widehat{\mathfrak{pen}}_j\big)_+ : j \in [\![m, \mathrm{M}]\!]\Big\} + 4\widehat{\mathfrak{pen}}_m.\end{aligned}$$

§14.48 **Proof** of Lemma §14.47. is given in the lecture. □

Similar to m_n° as in (14.02), which realises by construction a statistical-error-squared-bias compromise, let us fix a dimension $m^\diamond \in [\![\mathrm{M}]\!]$ to be specified below (analogously to (14.32)). Due to the last assertion for each $\theta \in \Theta$ we have

$$\begin{aligned}\mathbb{P}_\theta^n\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^{\widehat{m}}, \theta_\bullet)\big) \leqslant 16\mathfrak{bias}_{m^\diamond}^2(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \tfrac{4}{3}\mathfrak{pen}_{m^\diamond} + 28\mathbb{P}_\theta^n\Big(\max\Big\{\big(\mathfrak{d}_{\mathrm{ist}}^2(\widehat{\theta}_\bullet^j, \theta_\bullet^j) - \mathfrak{pen}_j/3\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\Big\}\Big) \\ + 8\mathbb{P}_\theta^n\Big(\max\Big\{\big(\mathfrak{pen}_j - \widehat{\mathfrak{pen}}_j\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\Big\}\Big) + 4\mathbb{P}_\theta^n\big(\widehat{\mathfrak{pen}}_{m^\diamond}\big). \qquad (14.36)\end{aligned}$$

We bound the remainder terms by imposing conditions including Assumption §14.41, which though hold true for a wide range of solutions $\theta_\bullet = \mathrm{U}\theta \in \Theta$ under reasonable model assumptions.

§14.49 **Assumption**. There exists a constant $\mathrm{C} := \mathrm{C}(\theta_\bullet) \in \mathbb{R}^+_{\setminus 0}$ and $(\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}))_{n\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ possibly depending on the parameter $\theta_\bullet = \mathrm{U}\theta \in \Theta$ such that for each $n \in \mathbb{N}$ the penatlties $\widehat{\mathfrak{pen}}_\bullet, \mathfrak{pen}_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$, the upper bound $\mathrm{M} \in \mathbb{N}$ and $m^\diamond \in [\![\mathrm{M}]\!]$ satisfy (14.33) in Assumption §14.41 and in addition

$$\mathbb{P}^n_\theta\big(\max\big\{\big(\mathfrak{pen}_j - \widehat{\mathfrak{pen}}_j\big)_+ : j \in [\![m^\diamond, \mathrm{M}]\!]\big\}\big) \leqslant \mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}).$$

The next assertion provides an upper bound for the risk of the estimator $\widehat{\theta}^{\widehat{m}}$ with data-driven choice $\widehat{m}$ given by (14.35).

§14.50 **Proposition**. *If $m^\diamond \in [\![\mathrm{M}]\!]$ satisfies the Assumption §14.49 then we have*

$$\mathbb{P}^n_\theta\big(\mathfrak{d}^2_{\mathrm{ist}}(\widehat{\theta}^{\widehat{m}}_\bullet, \theta_\bullet)\big) \leqslant 16\mathfrak{bias}^2_{m^\diamond}(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \tfrac{4}{3}\mathfrak{pen}_{m^\diamond} + 4\mathbb{P}^n_\theta(\widehat{\mathfrak{pen}}_{m^\diamond}) + 36\mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}).$$

§14.51 **Proof** of Proposition §14.50. is given in the lecture. □

§14.52 **Corollary**. *If $m^\diamond = \arg\min\big\{\mathfrak{bias}^2_m(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\}$ satisfies Assumption §14.49 and $\mathbb{P}^n_\theta(\widehat{\mathfrak{pen}}_{m^\diamond}) \leqslant \mathrm{K}\mathfrak{pen}_{m^\diamond}$ for some $\mathrm{K} \in [1, \infty)$, then*

$$\mathbb{P}^n_\theta\big(\mathfrak{d}^2_{\mathrm{ist}}(\widehat{\theta}^{\widehat{m}}_\bullet, \theta_\bullet)\big) \leqslant (16 \vee 6\mathrm{K}) \min\big\{\mathfrak{bias}^2_m(\theta_\bullet, \mathfrak{d}_{\mathrm{ist}}) + \mathfrak{pen}_m : m \in [\![\mathrm{M}]\!]\big\} + 36\,\mathrm{C}\,\mathrm{R}^{\mathrm{re}}_n(\theta).$$

§14.53 **Proof** of Corollary §14.52. is given in the lecture. □

§14|05 GSSM: data-driven local estimation

Lemma §14.18 in Subsection §14|03 presents tail bounds of sums of independent squared Gaussian random variables. We state next an elementary tail bound and a concentration inequality of a single Gaussian random variable.

§14.54 **Lemma**. *Let $Z \sim \mathrm{N}_{(0,1)}$. For all $\eta \in \mathbb{R}^+_{\setminus 0}$ and $\zeta, \mathrm{K} \in [1, \infty)$ we have*

$$\mathrm{N}_{(0,1)}(Z > \eta) \leqslant (2\pi\eta^2)^{-1/2}\exp(-\eta^2/2) \text{ and } \mathrm{N}_{(0,1)}\big(\big(Z^2 - 2\zeta(1 + \log \mathrm{K})\big)_+\big) \leqslant \mathrm{K}^{-\zeta}. \quad (14.37)$$

§14.55 **Proof** of Lemma §14.54. Exercise. □

§14|05|01 Local ϕ-risk

§14.56 **Reminder** *(Local oracle ϕ-risk in GSSM §14.02).* Given Model §14.02 we consider an OPE as in Section §12. Here the observable noisy version $\widehat{\theta}_\bullet$ admits a N^n_θ-distribution belonging to the family $\mathrm{N}^n_\Theta := (\mathrm{N}^n_\theta)_{\theta\in\Theta}$, $\Theta \subseteq \ell_2$. Let us recall (12.11) in Proposition §12.32 where $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\theta_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ and $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) := |\phi\nu_{\mathbb{N}}(\theta_\bullet \mathbb{1}^{m|\perp}_\bullet)|^2 + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi, \quad m^\circ_n := \arg\min\big\{\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet) := \mathrm{R}^{m^\circ_n}_n(\theta_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (14.38)$$

Due to Corollary §12.38 the (infeasible) OPE $\widehat{\theta}^{m^\circ_n}_\bullet = \widehat{\theta}_\bullet \mathbb{1}^{m^\circ_n}_\bullet \in \ell_2 \mathbb{1}^{m^\circ_n}_\bullet \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with oracle dimension m°_n as in (14.38) satisfies

$$\mathrm{N}^n_\theta(|\phi\nu_{\mathbb{N}}(\widehat{\theta}^{m^\circ_n}_\bullet - \theta_\bullet)|^2) = \mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet) = \inf_{m\in\mathbb{N}} \mathrm{N}^n_\theta(|\phi\nu_{\mathbb{N}}(\widehat{\theta}^m_\bullet - \theta_\bullet)|^2),$$

and hence it is *oracle optimal* (with constant 1). □

§14.57 **Corollary**. *For* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *and* $n, \mathrm{M} \in \mathbb{N}$ *setting* $\mathrm{K}_m := (\|\mathbb{1}_\bullet^m\|_\phi^2 \vee 1)m^2 \geqslant 1$, $m \in \mathbb{N}$, *we have*

$$\mathrm{N}_{\mathfrak{q}}^n\big(\max\big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^m - \theta_\bullet^m)|^2 - 2(1+\log \mathrm{K}_m)n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ : m \in [\![\mathrm{M}]\!]\big\}\big) \leqslant 2n^{-1}. \tag{14.39}$$

§14.58 **Proof** of Corollary §14.57. is given in the lecture. □

§14.59 **Notation**. Consider a sequence of penalties $\mathfrak{pen}_\bullet^\phi = (\mathfrak{pen}_m^\phi)_{m\in\mathbb{N}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ given by

$$\mathfrak{pen}_m^\phi := 6\big(1 + \big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + 2\log m\big)n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad \text{for each } m \in \mathbb{N} \tag{14.40}$$

which is obviously known in advance. Moreover, studying a ϕ-error the bias term introduced in (14.31) becomes

$$\mathfrak{bias}_m(\theta_\bullet, \phi_\bullet) = \sup\big\{|\phi\nu_{\mathbb{N}}(\theta_\bullet^j - \theta_\bullet^m)| = |\phi\nu_{\mathbb{N}}(\theta_\bullet \mathbb{1}_\bullet^{]\!]m,j]\!]})|\colon j \in [\![m,\infty]\!]\big\} \quad \forall m \in \mathbb{N}.$$

If $\theta_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ and hence $\nu_{\mathbb{N}}(|\phi_\bullet\theta_\bullet|) \in \mathbb{R}$ then $\mathfrak{bias}_m(\theta_\bullet, \phi_\bullet) \leqslant \nu_{\mathbb{N}}(|\phi_\bullet\theta_\bullet|\mathbb{1}_\bullet^{m|\perp}) = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence. Considering the data-driven OSE $\widehat{\theta}_\bullet^{\widehat{m}} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter $\widehat{m}$ selected as in (14.30) with penalty sequence $\mathfrak{pen}_\bullet^\phi$ given in (14.40) and arbitrary but fixed upper bound $\mathrm{M} \in \mathbb{N}$ we derive below an upper bound for its local ϕ-risk, $\mathrm{N}_{\mathfrak{q}}^n\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^{\widehat{m}} - \theta_\bullet^m)|^2\big)$. □

§14.60 **Proposition** *(GSSM (§14.02 continued)). Let* $\widehat{\theta}_\bullet = \theta_\bullet + n^{-1/2}\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{\mathfrak{q}}^n$ *as in Model §14.02 where* $\theta_\bullet \in \ell_2$ *and* $\dot{\mathrm{B}}_\bullet \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. *Given* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M} \in \mathbb{N}$ *and* $\mathfrak{pen}_\bullet^\phi$ *as in* (14.40) *consider a data-driven OPE* $\widehat{\theta}_\bullet^{\widehat{m}} = \widehat{\theta}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2\mathbb{1}_\bullet^{\widehat{m}} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $\theta_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with*

$$\widehat{m} := \arg\min\big\{\mathfrak{contr}_m^\phi + \mathfrak{pen}_m^\phi \colon m \in [\![\mathrm{M}]\!]\big\} \quad \textit{and}$$
$$\mathfrak{contr}_m^\phi := \max\big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^j - \widehat{\theta}_\bullet^m)|^2 - \mathfrak{pen}_j^\phi - \mathfrak{pen}_m^\phi\big)_+ \colon j \in]\!]m, \mathrm{M}]\!]\big\},\ m \in [\![\mathrm{M}]\!]. \tag{14.41}$$

Then for all $n, \mathrm{M} \in \mathbb{N}$ *we have*

$$\mathrm{N}_{\mathfrak{q}}^n\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^{\widehat{m}} - \theta_\bullet)|^2\big) \leqslant 64\min\big\{\mathfrak{bias}_m^2(\theta_\bullet, \phi_\bullet) + \big(1 + \big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + \log m\big)n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2 : m \in [\![\mathrm{M}]\!]\big\} + 56\,n^{-1}. \tag{14.42}$$

§14.61 **Proof** of Proposition §14.60. is given in the lecture. □

§14.62 **Comment**. Let us compare the dominating part of the upper bound given in (14.42), that is

$$\min\big\{\mathfrak{bias}_m^2(\theta_\bullet, \phi_\bullet) + \big(1 + \big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + \log m\big)n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2 : m \in [\![\mathrm{M}]\!]\big\} \tag{14.43}$$

with the oracle bound $\mathrm{R}_n^\circ(\theta_\bullet, \phi_\bullet) = \min\big\{|\phi\nu_{\mathbb{N}}(\theta_\bullet^m - \theta_\bullet)|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2 : m \in \mathbb{N}\big\}$ (for details see Reminder §14.56). In (14.43) we face eventually a deterioration by three sources. First, we generally have $\mathfrak{bias}_m(\theta_\bullet, \phi_\bullet) \geqslant |\phi\nu_{\mathbb{N}}(\theta_\bullet^m - \theta_\bullet)|$, but note that for $\theta_\bullet\phi_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ equality holds, that is

$$\mathfrak{bias}_m(\theta_\bullet, \phi_\bullet) = \sup\big\{\nu_{\mathbb{N}}(\phi_\bullet\theta_\bullet\mathbb{1}_\bullet^{]\!]m,j]\!]}) \colon j \in [\![m,\infty]\!]\big\} = \nu_{\mathbb{N}}(\phi_\bullet\theta_\bullet\mathbb{1}_\bullet^{m|\perp}) = |\phi\nu_{\mathbb{N}}(\theta_\bullet^m - \theta_\bullet)|$$

for all $m \in \mathbb{N}$. Secondly, the variance term features an additional factor $1 + \big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + \log m$, and finally the upper bound M might impose an additional deterioration. We note that the oracle bound $\mathrm{R}_n^\circ(\theta_\bullet, \phi_\bullet)$ is parametric, i.e. $n\mathrm{R}_n^\circ(\theta_\bullet, \phi_\bullet) = \mathrm{O}(1)$ as $n \to \infty$, if $\phi_\bullet \in \ell_2$ (case **(p)** in Illustration §12.40). In the sequel we consider only the case $\phi_\bullet \notin \ell_2$, i.e. $\nu_{\mathbb{N}}(|\phi_\bullet|^2) = \infty$. We set

$$\mathrm{M}^\phi := \max\big\{m \in \mathbb{N} \colon \|\mathbb{1}_\bullet^m\|_\phi^2 \leqslant n\phi_1^2\big\} \in \mathbb{N} \tag{14.44}$$

where the defining set is not empty and finite since $\|\phi_\bullet\|_{\ell_2}^2 = \infty$. The next assertion shows that this is a suitable choice for the upper bound. Moreover, we estimate the bias term by $\mathfrak{bias}_m(\theta_\bullet, \phi_\bullet) \leqslant \nu(|\phi_\bullet\theta_\bullet|\mathbb{1}_\bullet^{m|\perp})$ where equality holds whenever $\theta_\bullet\phi_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$. □

§14.63 **Corollary** *(GSSM (§14.02 continued)).* *Given* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *with* $\phi_\bullet \notin \ell_2$, $\mathrm{M}^\phi \in \mathbb{N}$ *as in* (14.44) *and* $\mathfrak{pen}^\phi_\bullet$ *as in* (14.40) *consider a data-driven OPE* $\widehat{\theta}^{\widehat{m}}_\bullet = \widehat{\theta}_\bullet \mathbb{1}^{\widehat{m}}_\bullet \in \ell_2 \mathbb{1}^{\widehat{m}}_\bullet \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $\theta_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with*

$$\widehat{m} := \arg\min\big\{\mathfrak{contr}^\phi_m + \mathfrak{pen}^\phi_m : m \in [\![\mathrm{M}^\phi]\!]\big\} \quad \textit{and}$$
$$\mathfrak{contr}^\phi_m := \max\Big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}^j_\bullet - \widehat{\theta}^m_\bullet)|^2 - \mathfrak{pen}^\phi_j - \mathfrak{pen}^\phi_m\big)_+ : j \in]\!]m, \mathrm{M}^\phi]\!]\Big\},\ m \in [\![\mathrm{M}^\phi]\!]. \quad (14.45)$$

For $n, m \in \mathbb{N}$ *we set*

$$\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) := \big(\nu_{\mathbb{N}}(|\phi_\bullet\theta_\bullet|\mathbb{1}^{m|\perp}_\bullet)\big)^2 + \big(1 + \big(\log\|\mathbb{1}^m_\bullet\|^2_\phi\big)_+ + \log m\big)\, n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi,$$
$$m^\diamond := \arg\min\big\{\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\} \quad \textit{and}$$
$$\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet) := \mathrm{R}^{m^\diamond}_n(\theta_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}^m_n(\theta_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (14.46)$$

Under the assumptions of Proposition §14.60 for each $n \in \mathbb{N}$ *such that* $\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet) \leqslant \phi_1^2$ *we have*

$$\mathrm{N}^n_\theta\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}^{\widehat{m}}_\bullet - \theta_\bullet)|^2\big) \leqslant 64\,\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet) + 56n^{-1} \leqslant (64 + 56\phi_1^{-2})\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet). \quad (14.47)$$

§14.64 **Proof** of Proof §14.64. is given in the lecture. □

§14.65 **Comment**. The data-driven bound $\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet)$ compared to the oracle bound $\mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet)$ features a deterioration of the variance term at least by a logarithmic factor. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§14.66 **Illustration**. We illustrate the last results considering the two specifications **(o)** and **(s)** given in Table 03 [§12] (Illustration §12.40). We restrict ourselves to the case $\phi_\bullet \notin \ell_2$ only.

Table 01 [§14]

Order of the oracle rate $\mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$ $(a \in \mathbb{R}^+_{\setminus 0})$		(squarred bias)	(variance)				
$\phi_j = j^{v-1/2}$	θ_j	$(\nu_{\mathbb{N}}(\lvert\phi_\bullet\theta_\bullet\rvert\mathbb{1}^{m\lvert\perp}_\bullet))^2$	$\lVert\mathbb{1}^m_\bullet\rVert^2_\phi$	M^ϕ	$m^\diamond$	$\mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet)$	$\mathrm{R}^\diamond_n(\theta_\bullet, \phi_\bullet)$
(o) $v \in (0, a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2v}}$	$\big(\frac{n}{\log n}\big)^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$	$\big(\frac{\log n}{n}\big)^{\frac{(a-v)}{a}}$
$v = 0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	e^n	$\big(\frac{n}{(\log n)^2}\big)^{\frac{1}{2a}}$	$\frac{\log n}{n}$	$\frac{(\log n)^2}{n}$
(s) $v \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+}e^{-2m^{2a}}$	m^{2v}	$n^{\frac{1}{2v}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}}(\log\log n)}{n}$
$v = 0$	$e^{-j^{2a}}$	$m^{(1-2a)_+}e^{-2m^{2a}}$	$\log m$	e^n	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 01 [§14] the order of the oracle rate $\mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}^\circ_n(\theta_\bullet, \phi_\bullet)$ is depict for $v \geqslant 0$ only. In case $v < 0$ we have $\phi_\bullet \in \ell_2$ and thus Corollary §14.63 is not applicable. □

§14|05|02 Maximal local ϕ-risk

§14.67 **Reminder** *(Maximal local ϕ-risk in* GSSM *§14.02).* Given Model §14.02 we consider an OPE as in Section §12. Here the observable noisy version $\widehat{\theta}_\bullet$ admits a N^n_θ-distribution belonging to the family $\mathrm{N}^n_\Theta := (\mathrm{N}^n_\theta)_{\theta\in\Theta}$, $\Theta \subseteq \ell_2$. Under Assumption §11.25 in Corollary §12.45 an upper bound

for a maximal local ϕ-risk of an OPE is shown. More precisely, the performance of the OPE $\widehat{\theta}^{m}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{m}_{\bullet} \in \ell_2\mathbb{1}^{m}_{\bullet} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m \in \mathbb{N}$ is measured by its maximal local ϕ-risk over the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}_n^{\phi}[\,\widehat{\theta}^{m}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathrm{N}_{\theta}^{n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}^{m}_{\bullet} - \theta_{\bullet})|^2\big) : \theta_{\bullet} \in \ell_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.13) where for $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) := \|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi} + n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}, \quad m_n^{\star} := \arg\min\big\{\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) := \mathrm{R}_n^{m_n^{\star}}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) = \min\big\{\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) : m \in \mathbb{N}\big\}. \quad (14.48)$$

By Corollary §12.45 under Assumption §11.25 the maximal local ϕ-risk of an OPE $\widehat{\theta}^{m_n^{\star}}_{\bullet}$ with optimally choosen dimension $m_n^{\star}$ as in (14.48) satisfies

$$\mathcal{R}_n^{\phi}[\,\widehat{\theta}^{m_n^{\star}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$$

with $\mathrm{C} = 1 \vee \mathrm{r}^2$. Moreover, under Assumption §13.24 due to Proposition §13.26 $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ provides (up to a constant) also a lower bound of the maximal local ϕ-risk over the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ is a minimax bound and $\widehat{\theta}^{m_n^{\star}}_{\bullet}$ is minimax optimal. However, the optimal dimension $m_n^{\star}$ depends on $\mathfrak{a}_{\bullet} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ characterising the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}}$. □

§14.68 **Proposition** *(GSSM (§14.02 continued)). Let $\widehat{\theta}_{\bullet} = \theta_{\bullet} + n^{-1/2}\dot{\mathrm{B}}_{\bullet} \sim \mathrm{N}_{\theta}^{n}$ as in Model §14.02 where $\theta_{\bullet} \in \ell_2$ and $\dot{\mathrm{B}}_{\bullet} \sim \mathrm{N}_{(0,1)}^{\otimes\mathbb{N}}$. Given $\phi_{\bullet} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, $\mathrm{M} \in \mathbb{N}$ and $\mathfrak{pen}^{\phi}_{\bullet}$ as in (14.40) consider a data-driven OPE $\widehat{\theta}^{\widehat{m}}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_2\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with $\widehat{m}$ as in (14.41). If Assumption §11.25 is satisfied, then for all $n, \mathrm{M} \in \mathbb{N}$ we have*

$$\mathcal{R}_n^{\phi}[\,\widehat{\theta}^{\widehat{m}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,]$$
$$\leqslant (16\mathrm{r}^2 \vee 64) \min\Big\{\|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi} + \big(1 + \big(\log\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}\big)_{+} + \log m\big)n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi} : m \in [\![\mathrm{M}]\!]\Big\}$$
$$+ 56\, n^{-1}. \quad (14.49)$$

§14.69 **Proof** of Proposition §14.68. is given in the lecture. □

§14.70 **Corollary** *(GSSM (§14.02 continued)). Given $\phi_{\bullet} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\phi_{\bullet} \notin \ell_2$, $\mathrm{M}^{\phi} \in \mathbb{N}$ as in (14.44) and $\mathfrak{pen}^{\phi}_{\bullet}$ as in (14.40) consider a data-driven OPE $\widehat{\theta}^{\widehat{m}}_{\bullet} = \widehat{\theta}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_2\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with $\widehat{m}$ as in (14.45). For $n, m \in \mathbb{N}$ we set*

$$\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) := \|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi} + \big(1 + \big(\log\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}\big)_{+} + \log m\big)\, n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi},$$
$$m^{\diamond} := \arg\min\big\{\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) : m \in \mathbb{N}\big\} \quad \textit{and}$$
$$\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) := \mathrm{R}_n^{m^{\diamond}}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) = \min\big\{\mathrm{R}_n^{m}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) : m \in \mathbb{N}\big\}. \quad (14.50)$$

Under Assumption §11.25 for each $n \in \mathbb{N}$ such that $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) \leqslant \phi_1^2$ we have

$$\mathcal{R}_n^{\phi}[\,\widehat{\theta}^{\widehat{m}}_{\bullet} \,|\, \ell_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant (16\mathrm{r}^2 \vee 64)\,\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet}) + 56n^{-1} \leqslant (16\mathrm{r}^2 \vee 64 + 56\phi_1^{-2})\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet}). \quad (14.51)$$

§14.71 **Proof** of Proof §14.71. is given in the lecture. □

§14.72 **Comment**. The data-driven bound $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ compared to the minimax bound $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ features a deterioration of the variance term at least by a factor $\log n$. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§14.73 **Illustration**. We illustrate the last results considering usual behaviour for $\mathfrak{a}_\bullet, \phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$. As in Illustration §12.47 we distinguish again the following two cases **(p)** $\phi_\bullet \in \ell_2$, and **(np)** $\phi_\bullet \notin \ell_2$. Interestingly, in case **(p)** the minimax bound $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ in Proposition §12.42 is parametric, that is, $n\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) = \mathrm{O}(1)$, in case **(np)** the bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) = \infty$. In case **(np)** consider the following two specifications:

Table 02 [§14]

Order of minimax rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet)$ as $n \to \infty$

	$(j \in \mathbb{N})$ $\phi_j = j^{v-1/2}$	$(a \in \mathbb{R}_{\setminus 0}^+)$ $\mathfrak{a}_j^2$	(squarred bias) $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m\|\perp}\|_\phi^2$	(variance) $\|\mathbb{1}_\bullet^m\|_\phi^2$	M^ϕ	$m^\diamond$	$\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$	$\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet)$
(o)	$v \in (0, a)$	j^{-2a}	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2v}}$	$\left(\frac{n}{\log n}\right)^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$	$\left(\frac{\log n}{n}\right)^{\frac{a-v}{a}}$
	$v = 0$	j^{-2a}	m^{-2a}	$\log m$	e^n	$\left(\frac{n}{(\log n)^2}\right)^{\frac{1}{2a}}$	$\frac{\log n}{n}$	$\frac{(\log n)^2}{n}$
(s)	$v \in \mathbb{R}_{\setminus 0}^+$	$e^{-j^{2a}}$	$m^{2(v-a)_+} e^{-m^{2a}}$	m^{2v}	$n^{\frac{1}{2v}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}}(\log\log n)}{n}$
	$v = 0$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	e^n	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 02 [§14] the order of the minimax rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for $v \geqslant 0$ only. For $v < 0$ we have $\phi_\bullet \in \ell_2$ and thus Corollary §14.70 is not applicable. □

Chapter 4

Nonparametric density estimation

This chapter presents nonparametric density estimation along the lines of the textbooks by **Tsybakov** *[2009] and* **Comte** *[2015] where far more details, examples and further discussions can be found.*

Overview

§15 Noisy density coefficients

§15.01 **Notation** *(Reminder).* Consider the measure space $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$ where $\lambda_{[0,1]}$ denotes the restriction of the Lebesgue measure to the Borel-σ-algebra $\mathscr{B}_{[0,1]}$ over $[0,1]$, and the Hilbert space $\mathbb{L}_2(\lambda_{[0,1]}) := \mathbb{L}_2([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$ of square Lebesgue-integrable functions endowed with its usual inner product $\langle h_1, h_2 \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = \lambda_{[0,1]}(h_1 h_2)$ for all $h_1, h_2 \in \mathbb{L}_2(\lambda_{[0,1]})$. Let $\mathbb{D}_2$ be a set of square-integrable Lebesgue densities on $([0,1], \mathscr{B}_{[0,1]})$, and hence $\mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})(=: \mathbb{H})$ as in Model §10.23. We denote for each density $\mathfrak{p} \in \mathbb{D}_2$ by $\mathbb{P}_\mathfrak{p} := \mathfrak{p}\lambda_{[0,1]}$ and $\mathbb{E}_\mathfrak{p}$ the associated probability measure and expectation, respectively. Keep in mind, that we identify equivalence classes and their representatives. □

§15.02 **Assumption.** We consider the statistical product experiment $\big([0,1]^n, \mathscr{B}_{[0,1]}^{\otimes n}, \mathbb{P}_{\mathbb{D}_2}^{\otimes n} := (\mathbb{P}_\mathfrak{p}^{\otimes n})_{\mathfrak{p} \in \mathbb{D}_2}\big)$ of size $n \in \mathbb{N}$ and for $\mathfrak{p} \in \mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ we denote by $(X_i)_{i \in [\![n]\!]} \sim \mathbb{P}_\mathfrak{p}^{\otimes n}$ an iid. sample of $X \sim \mathbb{P}_\mathfrak{p}$.

§15.03 **Notation** *(Reminder).* Consider an *orthonormal system* $(\mathrm{u}_j)_{j \in \mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$. Then $\mathrm{U} : \mathbb{L}_2(\lambda_{[0,1]}) \to \ell_2$ with $h \mapsto \mathrm{U}h := h_\bullet = (h_j := \langle h, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})})_{j \in \mathbb{N}}$ is a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$. Its adjoint operator $\mathrm{U}^\star \in \mathbb{L}(\ell_2, \mathbb{L}_2(\lambda_{[0,1]}))$ satisfies $\mathrm{U}^\star a_\bullet = \sum_{j \in \mathbb{N}} a_j \mathrm{u}_j =: \nu_\mathbb{N}(a_\bullet \mathrm{u}_\bullet)$ for all $a_\bullet \in \ell_2$. We call $h_\bullet = (h_j)_{j \in \mathbb{N}}$ *(generalised) Fourier coefficients* and U *(generalised) Fourier series transform.* □

§15.04 **Remark.** Let $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$ be a generalised Fourier series transform as in Notation §15.03 where $\mathbb{L}_2(\lambda_{[0,1]}) = \ker(\mathrm{U}) \oplus \mathrm{ran}(\mathrm{U}^\star)$ and $\mathrm{ran}(\mathrm{U}^\star) = \big\{\mathrm{U}^\star a_\bullet = \nu_\mathbb{N}(a_\bullet \mathrm{u}_\bullet) \colon a_\bullet \in \ell_2\big\}$. If U is not injective, then there exists $\mathcal{K} \subseteq \mathbb{N}$ and an orthonormal basis $(\mathrm{v}_j)_{j \in \mathcal{K}}$ of $\ker(\mathrm{U})$, and each $h \in \mathbb{L}_2(\lambda_{[0,1]})$ with $h_\bullet := \mathrm{U}h$ admits an expansion $h = \mathrm{U}^\star h_\bullet + \sum_{j \in \mathcal{K}} \langle h, \mathrm{v}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} \mathrm{v}_j$. We denote by $\mathbb{1}_{[0,1]} \in \mathscr{B}_{[0,1]}$ the constant function with $x \mapsto \mathbb{1}_{[0,1]}(x) := 1$. If $\mathbb{1}_{[0,1]} \in \ker(\mathrm{U})$ then we have $\langle h, \mathbb{1}_{[0,1]} \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = 0$ for all $h \in \mathrm{ran}(\mathrm{U}^\star)$, or in equal $\langle \nu_\mathbb{N}(a_\bullet \mathrm{u}_\bullet), \mathbb{1}_{[0,1]} \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = 0$ for all $a_\bullet \in \ell_2$, and in particular $\langle \mathrm{u}_j, \mathbb{1}_{[0,1]} \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} =$

0 for all $j \in \mathbb{N}$. For each density $\mathfrak{p} \in \mathbb{L}_2(\lambda_{[0,1]})$ we have $\langle \mathfrak{p}, \mathbb{1}_{[0,1]} \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = \lambda_{[0,1]}(\mathfrak{p}) = 1$. In other words the coefficient $\langle \mathfrak{p}, \mathbb{1}_{[0,1]} \rangle_{\mathbb{L}_2(\lambda_{[0,1]})}$ is always known. Therefore, we assume here and subsequently $\mathbb{1}_{[0,1]} \in \ker(\mathrm{U})$. Moreover we have $\mathbb{L}_2(\lambda_{[0,1]}) \subseteq \mathbb{L}_1([0,1], \mathscr{B}_{[0,1]}, \mathbb{P}_\mathfrak{p}) =: \mathbb{L}_1(\mathbb{P}_\mathfrak{p})$. Indeed, $h \in \mathbb{L}_2(\lambda_{[0,1]})$ satisfies $\mathbb{P}_\mathfrak{p}(|h|) = \mathfrak{p}\lambda_{[0,1]}(|h|) = \langle \mathfrak{p}, |h| \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} \leqslant \|\mathfrak{p}\|_{\mathbb{L}_2(\lambda_{[0,1]})} \|h\|_{\mathbb{L}_2(\lambda_{[0,1]})} \in \mathbb{R}^+$, and hence $h \in \mathbb{L}_1(\mathbb{P}_\mathfrak{p})$. Evidently, we have $\mathrm{u}_j \in \mathbb{L}_1(\mathbb{P}_\mathfrak{p})$ for all $j \in \mathbb{N}$ and the Fourier coefficients $\mathfrak{p}_\bullet = (\mathfrak{p}_j)_{j\in\mathbb{N}} = \mathrm{U}\,\mathfrak{p} \in \ell_2$ of $\mathfrak{p} \in \mathbb{L}_2(\lambda_{[0,1]})$ fulfil $\mathfrak{p}_j = \langle \mathfrak{p}, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = \lambda_{[0,1]}(\mathfrak{p}\,\mathrm{u}_j) = \mathfrak{p}\lambda_{[0,1]}(\mathrm{u}_j) = \mathbb{P}_\mathfrak{p}(\mathrm{u}_j)$ for all $j \in \mathbb{N}$. In addition we assume that for each $\mathfrak{p} \in \mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ the orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs also to $\mathbb{L}_2(\mathbb{P}_\mathfrak{p}) := \mathbb{L}_2([0,1], \mathscr{B}_{[0,1]}, \mathbb{P}_\mathfrak{p})$, i.e. $\mathrm{u}_j \in \mathbb{L}_2(\mathbb{P}_\mathfrak{p})$ for all $j \in \mathbb{N}$. □

§15.05 **Assumption**. The *orthonormal system* $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$, and its associated *generalised Fourier series transform* $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$ with $h \mapsto \mathrm{U}h := h_\bullet = (h_j := \langle h, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})})_{j\in\mathbb{N}}$, is fixed and known in advance. U is a partial isometry with **(os1)** $\mathbb{1}_{[0,1]} \in \ker(\mathrm{U})$. **(os2)** For all $\mathfrak{p} \in \mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ the orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs to $\mathbb{L}_2(\mathbb{P}_\mathfrak{p})$. □

§15.06 **Remark**. If in addition $\mathbb{D}_2 \subseteq \mathbb{L}_\infty(\lambda_{[0,1]})$ then for each $\mathfrak{p} \in \mathbb{D}_2$ we have $\mathbb{L}_2(\lambda_{[0,1]}) \subseteq \mathbb{L}_2(\mathbb{P}_\mathfrak{p})$. Indeed, $h \in \mathbb{L}_2(\lambda_{[0,1]})$ satisfies $\mathbb{P}_\mathfrak{p}(|h|^2) = \lambda_{[0,1]}(\mathfrak{p}\,|h|^2) \leqslant \|\mathfrak{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \|h\|^2_{\mathbb{L}_2(\lambda_{[0,1]})} \in \mathbb{R}^+$. Consequently, any orthonormal system in $\mathbb{L}_2(\lambda_{[0,1]})$ belongs also to $\mathbb{L}_2(\mathbb{P}_\mathfrak{p})$, and **(os2)** in Assumption §15.05 is satisfied. Alternatively, **(os2)** is fulfilled for arbitrary $\mathbb{D}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ if $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs also to $\mathbb{L}_\infty(\lambda_{[0,1]})$. □

§15.07 **Notation** *(Reminder)*. Similar to an Empirical mean model §10.07 for each $j \in \mathbb{N}$ we define $\widehat{\mathfrak{p}}_j := \widehat{\mathbb{P}}_n(\mathrm{u}_j) \in \mathscr{B}^{\otimes n}_{[0,1]}$ with $x^n \mapsto (\widehat{\mathbb{P}}_n(\mathrm{u}_j))(x^n) = n^{-1}\sum_{i\in[\![n]\!]} \mathrm{u}_j(x^n_i)$. Since the stochastic process $\mathrm{u}_\bullet = (\mathrm{u}_j)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ is $\mathscr{B}_{[0,1]} \otimes 2^{\mathbb{N}}$-$\mathscr{B}$-measurable, the stochastic process $\widehat{\mathfrak{p}}_\bullet = (\widehat{\mathfrak{p}}_j := \widehat{\mathbb{P}}_n(\mathrm{u}_j))_{j\in\mathbb{N}}$ on $([0,1]^n, \mathscr{B}^{\otimes n}_{[0,1]})$ is $\mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$-$\mathscr{B}$-measurable, $\widehat{\mathfrak{p}}_\bullet \in \mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$ for short. By construction $\mathfrak{p}_\bullet = (\mathfrak{p}_j = \mathbb{P}_\mathfrak{p}(\mathrm{u}_j))_{j\in\mathbb{N}} \in 2^{\mathbb{N}}$ is the ℓ_2-mean of $\widehat{\mathfrak{p}}_\bullet$. For each $j \in \mathbb{N}$ the statistic $\varepsilon_j := n^{1/2}(\widehat{\mathbb{P}}_n(\mathrm{u}_j) - \mathbb{P}_\mathfrak{p}(\mathrm{u}_j)) \in \mathscr{B}^{\otimes n}_{[0,1]}$ is centred, i.e. $\varepsilon_j \in \mathbb{L}_1([0,1]^n, \mathscr{B}^{\otimes n}_{[0,1]}, \mathbb{P}^{\otimes n}_\mathfrak{p}) =: \mathbb{L}_1(\mathbb{P}^{\otimes n}_\mathfrak{p})$ with $\mathbb{P}^{\otimes n}_\mathfrak{p}(\varepsilon_j) = 0$, and $\varepsilon_\bullet = (\varepsilon_j)_{j\in\mathbb{N}} \in \mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$. Since $\widehat{\mathfrak{p}}_j = \mathfrak{p}_j + n^{-1/2}\varepsilon_j$ for each $j \in \mathbb{N}$ by construction $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ is a noisy version of $\mathfrak{p}_\bullet$ (see Definition §10.19). Moreover, under Assumption §15.05 $\widehat{\mathfrak{p}}_\bullet$ admits a covariance function $\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet} \in \mathbb{R}^{\mathbb{N}^2}$ given for $j, j_o \in \mathbb{N}$ by

$$n\,\mathbb{C}\mathrm{ov}(\widehat{\mathfrak{p}}_j, \widehat{\mathfrak{p}}_{j_o}) = \mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_{j_o}) = \mathbb{P}^{\otimes n}_\mathfrak{p}(\varepsilon_j\varepsilon_{j_o}) = \mathbb{P}_\mathfrak{p}(\mathrm{u}_j\mathrm{u}_{j_o}) - \mathbb{P}_\mathfrak{p}(\mathrm{u}_j)\mathbb{P}_\mathfrak{p}(\mathrm{u}_{j_o}) = \mathbb{P}_\mathfrak{p}(\mathrm{u}_j\mathrm{u}_{j_o}) - \mathfrak{p}_j\,\mathfrak{p}_{j_o} =: \mathrm{cov}^\mathfrak{p}_{j,j_o}.$$

Consequently, we have $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet})}$ and $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\varepsilon_\bullet \sim \mathrm{P}_{(\mathfrak{p}, n^{-1}\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet})}$ (see Definition §10.19). □

§15.08 **Noisy density coefficients**. Under Assumptions §15.02 and §15.05 the stochastic process $\varepsilon_\bullet = (\varepsilon_j := n^{1/2}(\widehat{\mathbb{P}}_n(\mathrm{u}_j) - \mathbb{P}_\mathfrak{p}(\mathrm{u}_j)))_{j\in\mathbb{N}}$ satisfies Assumption §10.04, i.e. $\varepsilon_\bullet \in \mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$, and $\varepsilon_\bullet$ has mean zero under $\mathbb{P}^{\otimes n}_\mathfrak{p}$. The stochastic process $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ with ℓ_2-mean $\mathfrak{p}_\bullet$ is called a *noisy version* of the density coefficients $\mathfrak{p}_\bullet = \mathrm{U}\,\mathfrak{p} \in \ell_2$, or *noisy density coefficients* for short. Moreover $\varepsilon_\bullet$ admits under $\mathbb{P}^{\otimes n}_\mathfrak{p}$ a covariance function $\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet} \in \mathbb{R}^{\mathbb{N}^2}$ given for $j, j_o \in \mathbb{N}$ by $\mathrm{cov}^\mathfrak{p}_{j,j_o} = \mathbb{P}_\mathfrak{p}(\mathrm{u}_j\mathrm{u}_{j_o}) - \mathbb{P}_\mathfrak{p}(\mathrm{u}_j)\mathbb{P}_\mathfrak{p}(\mathrm{u}_{j_o})$. We eventually write $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet})}$ and $\widehat{\mathfrak{p}}_\bullet \sim \mathrm{P}_{(\mathfrak{p}, n^{-1}\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet})}$. If in addition $\varepsilon_\bullet$ admits a covariance operator $\Gamma_\mathfrak{p} \in \mathbb{L}^{\geqslant}(\ell_2)$ then we write $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma_\mathfrak{p})}$ and $\widehat{\mathfrak{p}}_\bullet \sim \mathrm{P}_{(\mathfrak{p}, n^{-1}\Gamma_\mathfrak{p})}$ for short. □

§15.09 **Remark**. The centred stochastic process $\varepsilon_\bullet := (\varepsilon_j)_{j\in\mathbb{N}}$ of error terms in Definition §15.08 is in general not a white noise process. □

§15.10 **Lemma**. *Under Assumptions §15.02 and §15.05 consider the stochastic process $\varepsilon_\bullet \in \mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$ as in Definition §15.08.*

(i) *If $\mathfrak{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then under $\mathbb{P}^{\otimes n}_\mathfrak{p}$, $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^\mathfrak{p}_{\bullet,\bullet})}$ admits a covariance operator $\Gamma_\mathfrak{p} \in \mathbb{L}^{\geqslant}(\ell_2)$ given by*

$$a_\bullet \mapsto \Gamma_\mathfrak{p} a_\bullet = (\nu_{\mathbb{N}}(\mathrm{cov}^\mathfrak{p}_{j,\bullet} a_\bullet) = \sum_{j_o\in\mathbb{N}} \mathrm{cov}^\mathfrak{p}_{j,j_o} a_{j_o})_{j\in\mathbb{N}}$$

where $\|\Gamma_{\mathbb{p}}\|_{\mathbb{L}(\ell_2)} \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}$.

(ii) *If* $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *and* $\mathbb{p}^{-1} := \frac{1}{\mathbb{p}} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *then* $\Gamma_{\mathbb{p}} \in \mathbb{L}^{\geqslant}(\ell_2)$ *is invertible with inverse* $\Gamma_{\mathbb{p}}^{-1} \in \mathbb{L}(\ell_2)$ *where* $\|\Gamma_{\mathbb{p}}^{-1}\|_{\mathbb{L}(\ell_2)} \leqslant \|\mathbb{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}$.

Consequently, if $\mathbb{v}_{\mathbb{p}} := \max(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathbb{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ *then for all* $a_\bullet \in \ell_2$ *we have*

$$\mathbb{v}_{\mathbb{p}}^{-1}\|a_\bullet\|_{\ell_2}^2 \leqslant \|a_\bullet\|_{\Gamma_{\mathbb{p}}}^2 = \langle \Gamma_{\mathbb{p}} a_\bullet, a_\bullet\rangle_{\ell_2} \leqslant \mathbb{v}_{\mathbb{p}}\|a_\bullet\|_{\ell_2}^2.$$

§15.11 **Proof** of Lemma §15.10. is given in the lecture. □

§15.12 **Remark**. For each $j \in \mathbb{N}$ consider $\mathbb{1}_\bullet^{\{j\}} = (\mathbb{1}_{\{j\}}(l))_{l\in\mathbb{N}} \in (\mathbb{R})^{\mathbb{N}}$ where $(\mathbb{1}_\bullet^{\{j\}})_{j\in\mathbb{N}}$ forms an orthonormal basis in ℓ_2. If $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ from Lemma §15.10 (i) for each $j \in \mathbb{N}$ we obtain

$$\mathbb{P}_{\mathbb{p}}^{\otimes n}\left(\varepsilon_j^2\right) = \mathbb{P}_{\mathbb{p}}^{\otimes n}\left(|\nu_{\mathbb{N}}(\mathbb{1}_\bullet^{\{j\}}\varepsilon_\bullet)|^2\right) = \langle \Gamma_{\mathbb{p}}\mathbb{1}_\bullet^{\{j\}}, \mathbb{1}_\bullet^{\{j\}}\rangle_{\ell_2} \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\|\mathbb{1}_\bullet^{\{j\}}\|_{\ell_2}^2 = \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}$$

Keeping the last identities in mind if $\mathbb{v}_{\mathbb{p}} := \max(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathbb{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ then due to Lemma §15.10 for all $j \in \mathbb{N}$ we have $\mathbb{v}_{\mathbb{p}}^{-1} \leqslant \mathbb{P}_{\mathbb{p}}^{\otimes n}\left(\varepsilon_j^2\right) \leqslant \mathbb{v}_{\mathbb{p}}$. □

§16 Projection density estimator

§16.01 **Notation** *(Reminder)*. Consider the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ as in Notation §10.11. For $w_\bullet \in \mathbb{R}^{\mathbb{N}}$ define the multiplication map $\mathrm{M}_{w_\bullet} : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ with $a_\bullet \mapsto \mathrm{M}_{w_\bullet}a_\bullet := w_\bullet a_\bullet$. Note that each $w_\bullet \in \mathbb{R}^{\mathbb{N}}$ is $2^{\mathbb{N}}$-$\mathscr{B}$-measurable. We denote by $\mathbb{M}_{\mathbb{R}^{\mathbb{N}}}$ the set of all multiplication maps defined on $\mathbb{R}^{\mathbb{N}}$. If in addition $w_\bullet \in \ell_\infty = \mathbb{L}_\infty(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ then we have also $\mathrm{M}_{w_\bullet} : \ell_2 \to \ell_2$. We set $\mathbb{L}^{\mathrm{M}}(\ell_2) = \left\{\mathrm{M}_{w_\bullet} \in \mathbb{M}_{\mathbb{R}^{\mathbb{N}}} : w_\bullet \in \ell_\infty\right\} \subseteq \mathbb{L}(\ell_2)$ noting that $\|\mathrm{M}_{w_\bullet}\|_{\mathbb{L}(\ell_2)} = \sup\left\{\|w_\bullet a_\bullet\|_{\ell_2} : \|a_\bullet\|_{\ell_2} \leqslant 1\right\} \leqslant \|w_\bullet\|_{\ell_\infty}$ for each $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\ell_2)$. □

§16.02 **Reminder**. If $w_\bullet \in \ell_\infty$ then $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\ell_2)$, and $\mathrm{M}_{w_\bullet^\dagger} : \ell_2 \supseteq \mathrm{dom}(\mathrm{M}_{w_\bullet^\dagger}) \to \ell_2$. Moreover, we have $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \ell_2$, $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \ell_2 w_\bullet$ and $\ker(\mathrm{M}_{w_\bullet}) = \ell_2\mathbb{1}_\bullet^{\mathcal{N}_w}$ with $\mathcal{N}_w = \left\{j \in \mathbb{N} : w_j = 0\right\} \in 2^{\mathbb{N}}$ (see Property §11.03), and $\mathrm{dom}(\mathrm{M}_{w_\bullet^\dagger}) = \ell_2 w_\bullet \oplus \ell_2\mathbb{1}_\bullet^{\mathcal{N}_w}$ (see Property §11.05). Consequently, if in addition $\nu_{\mathbb{N}}(\mathcal{N}_w) = 0$ or in equal $w_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, then $w_\bullet^\dagger = w_\bullet^{-1} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, hence $w_\bullet^{2|\dagger} = w_\bullet^{-2} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$, and $\ell_2^w = \mathrm{dom}(\mathrm{M}_{w_\bullet^{-1}}) = \ell_2 w_\bullet = \mathbb{L}_2(w_\bullet^{-2}\nu_{\mathbb{N}}) =: \ell_2(w_\bullet^{-2})$. For each $m \in \mathbb{N}$ we write $\mathbb{1}_\bullet^m = (\mathbb{1}_j^m)_{j\in\mathbb{N}} := \mathbb{1}_\bullet^{[\![m]\!]}$ and $\mathbb{1}_\bullet^{m|\perp} := \mathbb{1}_\bullet - \mathbb{1}_\bullet^m$ with $[\![m]\!] := [-m, m] \cap \mathbb{N}$. Consequently, $\mathrm{M}_{\mathbb{1}_\bullet^m} \in \mathbb{L}^{\geqslant}(\ell_2)$ and $\mathrm{M}_{\mathbb{1}_\bullet^{m|\perp}} \in \mathbb{L}^{\geqslant}(\ell_2)$ is the *orthogonal projection* onto the linear subspace $\ell_2\mathbb{1}_\bullet^m \subseteq \ell_2$ and its orthogonal complement $\ell_2\mathbb{1}_\bullet^{m|\perp} = (\ell_2\mathbb{1}_\bullet^m)^\perp \subseteq \ell_2$, respectively, that is $\ell_2 = \ell_2\mathbb{1}_\bullet^m \oplus \ell_2\mathbb{1}_\bullet^{m|\perp}$ (see Property §11.07). Finally, given $h_\bullet = \mathrm{U}h \in \ell_2$ for $h \in \mathbb{L}_2(\lambda_{[0,1]})$ we consider the orthogonal projections $h_\bullet^m = h_\bullet\mathbb{1}_\bullet^m \in \ell_2\mathbb{1}_\bullet^m$ and $h^m := \mathrm{U}^\star h_\bullet^m \in \mathbb{L}_2(\lambda_{[0,1]})$ (Definition §11.08). □

§16.03 **Notation** *(Reminder)*. Consider the stochastic processes $\varepsilon_\bullet = (\varepsilon_j := n^{1/2}(\widehat{\mathbb{P}}_n(\mathrm{u}_j) - \mathbb{P}_{\mathbb{p}}(\mathrm{u}_j)))_{j\in\mathbb{N}}$ given in Definition §15.08. The observable noisy density coefficients $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficents $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19). Under Assumptions §15.02 and §15.05 $\varepsilon_\bullet$ is centred and admits a covariance function $\mathrm{cov}_{\bullet,\bullet}^{\mathbb{p}} \in \mathbb{R}^{\mathbb{N}^2}$ given in Definition §15.08, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}_{\bullet,\bullet}^{\mathbb{p}})}$ and $\widehat{\mathbb{p}}_\bullet \sim \mathrm{P}_{(\mathbb{p}_\bullet, n^{-1}\mathrm{cov}_{\bullet,\bullet}^{\mathbb{p}})}$. If in addition $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then $\varepsilon_\bullet$ admits a covariance operator $\Gamma_{\mathbb{p}} \in \mathbb{L}^{\geqslant}(\ell_2)$ given in Lemma §15.10, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma_{\mathbb{p}})}$ and $\widehat{\mathbb{p}}_\bullet \sim \mathrm{P}_{(\mathbb{p}_\bullet, n^{-1}\Gamma_{\mathbb{p}})}$ □

§16.04 **Definition**. Given a noisy version $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficients $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2$ for each $m \in \mathbb{N}$ we call $\widehat{\mathbb{p}}_\bullet^m := \widehat{\mathbb{p}}_\bullet\mathbb{1}_\bullet^m$ *orthogonal projection estimator (OPE)* of $\mathbb{p}_\bullet$. □

§16.05 **Remark.** Under Assumptions §15.02 and §15.05 we consider the function (with random coefficents) $\widehat{\mathbb{p}}^m := \mathbb{1}_{[0,1]} + \mathrm{U}^\star \widehat{\mathbb{p}}^m_\bullet$ which belongs to $\mathbb{L}_2(\lambda_{[0,1]})$, integrates to one, but may take on negative values. Fortunately, there is a simple remedy — its $\mathbb{L}_2(\lambda_{[0,1]})$-projection onto a class of nonnegative densities,

$$\widehat{\mathbb{p}} := \big(\widehat{\mathbb{p}}^m - c\big)_+ \quad \text{with } c \in \mathbb{R}^+ \text{ such that } \lambda_{[0,1]}(\widehat{\mathbb{p}}) = 1.$$

We call $\widehat{\mathbb{p}}^m$ an *orthogonal projection density estimator* of $\mathbb{p}$. If $\mathbb{p} = \mathbb{1}_{[0,1]} + \mathrm{U}^\star \mathbb{p}_\bullet$ (for example $\ker(\mathrm{U})$ is spanned by $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ or in equal $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ is an orthonormal basis of $\mathbb{L}_2(\lambda_{[0,1]})$), then we have

$$\|\widehat{\mathbb{p}}^m - \mathbb{p}\|^2_{\mathbb{L}_2(\lambda_{[0,1]})} = \|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet\|^2_{\ell_2}.$$

In this situation all results for the OPE $\widehat{\mathbb{p}}^m_\bullet$ of the density coefficients immediately transfer onto the orthogonal projection density estimator $\widehat{\mathbb{p}}^m$ of the density $\mathbb{p}$. □

§16|01 Global and maximal global $\mathfrak{v}$-risk

We measure first the accuracy of the OPE $\widehat{\mathbb{p}}^m_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^m_\bullet$ of $\mathbb{p}^m_\bullet = \mathbb{p}_\bullet \mathbb{1}^m_\bullet \in \ell_2 \mathbb{1}^m_\bullet$ with $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2$ by a global mean-$\mathfrak{v}$-error, i.e. $\mathfrak{v}$-risk.

§16.06 **Reminder.** If $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $\mathbb{p}_\bullet \in \ell_2(\mathfrak{v}^2_\bullet)$ then we have $\mathbb{p}^m_\bullet = \mathbb{p}_\bullet \mathbb{1}^m_\bullet \in \ell_2(\mathfrak{v}^2_\bullet)$ too and $\|\mathbb{p}^m_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}} = o(1)$ as $m \to \infty$ (Property §11.09). Moreover, $\boldsymbol{\varepsilon}_\bullet \in \mathscr{B}^{\otimes n}_{[0,1]} \otimes 2^{\mathbb{N}}$ given in Definition §15.08 satisfies $\mathfrak{v}_\bullet \boldsymbol{\varepsilon}_\bullet \mathbb{1}^m_\bullet \in \ell_2$ (note that $\mathbb{1}^m_\bullet \in \ell_2$ and $\mathfrak{v}_\bullet \mathbb{1}^m_\bullet, \boldsymbol{\varepsilon}_\bullet \mathbb{1}^m_\bullet \in \ell_\infty$) and thus also

$$n^{-1/2} \mathfrak{v}_\bullet \boldsymbol{\varepsilon}_\bullet \mathbb{1}^m_\bullet + \mathfrak{v}_\bullet \mathbb{p}^m_\bullet = \mathfrak{v}_\bullet \widehat{\mathbb{p}}^m_\bullet \in \ell_2. \tag{16.01}$$

Finally, under Assumptions §15.02 and §15.05 and $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ due to Lemma §15.10 we have $\mathbb{P}^{\otimes n}_{\mathbb{p}}(\boldsymbol{\varepsilon}^2_\bullet) \in \ell_\infty$, more precisely, $\|\mathbb{P}^{\otimes n}_{\mathbb{p}}(\boldsymbol{\varepsilon}^2_\bullet)\|_{\ell_\infty} \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}$ (see Remark §15.12). □

§16|01|01 Global $\mathfrak{v}$-risk

§16.07 **Proposition** *(Upper bound).* *Let Assumptions §15.02 and §15.05, $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $\mathbb{p}_\bullet \in \ell_2(\mathfrak{v}^2_\bullet)$ be satisfied and for all $n, m \in \mathbb{N}$ set*

$$\mathrm{R}^m_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) := \|\mathbb{p}_\bullet \mathbb{1}^{m\perp}_\bullet\|^2_{\mathfrak{v}} + n^{-1}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}, \quad m^\circ_n := \arg\min\big\{\mathrm{R}^m_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}^{m^\circ_n}_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}^m_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \tag{16.02}$$

If $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then we have $\mathbb{P}^{\otimes n}_{\mathbb{p}}(\|\widehat{\mathbb{p}}^{m^\circ_n}_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}}) \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$.

§16.08 **Proof** of Proposition §16.07. is given in the lecture. □

§16.09 **Oracle inequality.** *Under Assumptions §15.02 and §15.05 let $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $\mathbb{p}_\bullet \in \ell_2(\mathfrak{v}^2_\bullet)$. If in addition $\mathbb{v}_{\mathbb{p}} := \max(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathbb{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ then $\mathbb{v}^{-1}_{\mathbb{p}} \leqslant \mathbb{v}_j^{\mathbb{p}} := \mathbb{P}^{\otimes n}_{\mathbb{p}}(\boldsymbol{\varepsilon}^2_j) \leqslant \mathbb{v}_{\mathbb{p}}$ for all $j \in \mathbb{N}$ (see Remark §15.12), and hence Property §12.15 implies*

$$\mathbb{v}^{-1}_{\mathbb{p}} \mathrm{R}^m_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{P}^{\otimes n}_{\mathbb{p}}(\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}}) = n^{-1}\nu_{\mathbb{N}}(\mathbb{v}^{\mathbb{p}}_\bullet \mathfrak{v}^2_\bullet \mathbb{1}^m_\bullet) + \|\mathbb{p}_\bullet \mathbb{1}^{m\perp}_\bullet\|^2_{\mathfrak{v}}$$
$$\leqslant \mathbb{v}_{\mathbb{p}} \mathrm{R}^m_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) \quad \textit{for all } m, n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality (see Definition §12.14)

$$\mathbb{v}^{-1}_{\mathbb{p}} \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}^{\otimes n}_{\mathbb{p}}(\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}}) \leqslant \mathbb{P}^{\otimes n}_{\mathbb{p}}(\|\widehat{\mathbb{p}}^{m^\circ_n}_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}})$$
$$\leqslant \mathbb{v}_{\mathbb{p}} \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{v}^2_{\mathbb{p}} \inf_{m\in\mathbb{N}} \mathbb{P}^{\otimes n}_{\mathbb{p}}(\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet\|^2_{\mathfrak{v}}), \tag{16.03}$$

and, hence $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$, m_n° *and the statistic* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$, *respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant* $\mathbb{v}_{\mathbb{p}}^2$*). We observe that* $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) = \mathrm{o}(1)$ *as* $n \to \infty$ *(Remark §12.16), and thus,* $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$ *is an oracle rate. However, note that the oracle dimension* $m_n^\circ = m_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$ *depends on the unknown density coefficients* $\mathbb{p}_\bullet$*, and thus also the oracle optimal statistic* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$*. In other words* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$ *is not a feasible estimator.* □

§16.10 **Illustration.** We illustrate the last results considering usual behaviour for the bias and variance term. We distinguish the following two cases

(p) $\mathfrak{v}_\bullet \in \ell_2$ or there is $m \in \mathbb{N}$ with $\|\mathbb{p}_\bullet^m - \mathbb{p}_\bullet\|_{\mathfrak{v}}^2 = 0$,

(np) $\mathfrak{v}_\bullet \notin \ell_2$ and for all $m \in \mathbb{N}$ holds $\|\mathbb{p}_\bullet^m - \mathbb{p}_\bullet\|_{\mathfrak{v}}^2 \in \mathbb{R}_{\setminus 0}^+$.

Interestingly, in case **(p)** the oracle bound is parametric, that is, $n\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) = \mathrm{O}(1)$, in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet) = \infty$. In case **(np)** consider the following two specifications:

Table 01 [§16]

Order of the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$	$(a \in \mathbb{R}_{\setminus 0}^+)$	(squarred bias)	(variance)		
$\mathfrak{v}_j^2 = j^{2v}$	$\mathbb{p}_j^2$	$\|\mathbb{p}_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\mathfrak{v}}^2$	$\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$	m_n°	$\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$
(o) $v \in (-1/2, a)$	j^{-2a-1}	$m^{-2(a-v)}$	m^{2v+1}	$n^{\frac{1}{2a+1}}$	$n^{-\frac{2(a-v)}{2a+1}}$
$v = -1/2$	j^{-2a-1}	m^{-2a-1}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a+1}}$	$\frac{\log n}{n}$
(s) $v + 1/2 \in \mathbb{R}_{\setminus 0}^+$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+} e^{-m^{2a}}$	m^{2v+1}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{2v+1}{2a}}}{n}$
$v = -1/2$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 01 [§16] the order of the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$ is depict for $v \geqslant -1/2$ only. In case $v < -1/2$ the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \mathfrak{v}_\bullet)$ is parametric. □

§16|01|02 Maximal global $\mathfrak{v}$-risk

§16.11 **Assumption.** Consider weights $\mathfrak{a}_\bullet, \mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\mathfrak{v})_\bullet := (\mathfrak{a}_j\mathfrak{v}_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\mathfrak{v}_\bullet \in \ell_\infty$. We write $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet^2 = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2')** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2\mathrm{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} =: \mathcal{T}_{\mathfrak{a},\mathrm{u}}^2 \in \mathbb{R}^+$. □

§16.12 **Reminder.** Under Assumption §16.11 we have $\ell_2^{\mathfrak{a}} = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}^{-1}}) = \ell_2\mathfrak{a}_\bullet \subseteq \ell_2$ and the three measures $\nu_{\mathbb{N}}$, $\mathfrak{a}_\bullet^{-2}\nu_{\mathbb{N}}$ and $\mathfrak{v}_\bullet^2\nu_{\mathbb{N}}$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\ell_2^{\mathfrak{a}}$ endowed with $\|\cdot\|_{\mathfrak{a}^{-1}} = \|\mathrm{M}_{\mathfrak{a}^{-1}}\cdot\|_{\ell_2}$ and given a constant $\mathrm{r} \in \mathbb{R}_{\setminus 0}^+$ the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}} := \{a_\bullet \in \ell_2^{\mathfrak{a}} : \|a_\bullet\|_{\mathfrak{a}^{-1}} \leqslant \mathrm{r}\} \subseteq \ell_2^{\mathfrak{a}}$. Since $(\mathfrak{a}\mathfrak{v})_\bullet \in \ell_\infty$, and hence $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ we have $\ell_2^{\mathfrak{a}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ (Property §11.15), and $\|a_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\mathfrak{v}} \leqslant \mathrm{r}\,(\mathfrak{a}\mathfrak{v})_{(m)}$ for all $a_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.17). □

§16.13 **Remark.** We replace Assumption §15.05 **(os1)** and **(os2)**, respectively, by the stronger Assumption §16.11 **(os1')** and **(os2')**. Indeed, under **(os1')** we have **(os1)** $\mathbb{1}_{[0,1]} \in \ker(\mathrm{U})$. Furthermore, $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs to $\mathbb{L}_\infty(\lambda_{[0,1]})$ due to **(os2')** (and $\mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$), and hence **(os2)** is fulfilled (see also Remark §15.06). □

§16.14 **Lemma.** *Under Assumption §16.11 set*

$$\mathbb{D}_2^{\mathfrak{a},\mathrm{r}} := \big\{\mathbb{p} \in \mathbb{L}_2(\lambda_{[0,1]}) \colon \mathbb{p} \text{ is a density and } \mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2^{\mathfrak{a},\mathrm{r}}\big\}. \tag{16.04}$$

Then we have $\sup\{\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}: \mathbb{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\} \leqslant 1 + \mathrm{r}\tau_{\mathfrak{a},\mathfrak{u}}$.

§16.15 **Proof** of Lemma §16.14. is given in the lecture. □

§16.16 **Proposition** *(Upper bound). Let Assumptions §15.02 and §16.11 be satisfied. For* $n \in \mathbb{N}$ *considering* $m_n^\star \in \mathbb{N}$ *and* $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \in \mathbb{R}^+$ *as in* (12.06) *(Proposition §12.21) we have*

$$\sup\{\mathbb{P}_{\mathbb{p}}^{\otimes n}(\|\widehat{\mathbb{p}}_\bullet^{m_n^\star} - \mathbb{p}_\bullet\|_{\mathfrak{v}}^2): \mathbb{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet).$$

with constant $\mathrm{C} = 1 + \mathrm{r}\tau_{\mathfrak{a},\mathfrak{u}} + \mathrm{r}^2$.

§16.17 **Proof** of Proposition §16.16. is given in the lecture. □

§16.18 **Illustration**. The *trigonometric basis* given for $x \in [0,1]$ by

$$\mathfrak{u}_0 := \mathbb{1}_{[0,1]},\ \mathfrak{u}_{2k}(x) := \sqrt{2}\cos(2\pi kt),\ \mathfrak{u}_{2k-1}(x) := \sqrt{2}\sin(2\pi kt),\ k \in \mathbb{N},$$

is an orthonormal basis of $\mathbb{L}_2(\lambda_{[0,1]})$. It satisfies Assumption §16.11 **(os1')**, since $\|\mathfrak{u}_j^2\|_{\mathbb{L}_2(\lambda_{[0,1]})} = 2$ for all $j \in \mathbb{N}$. Consequently, for all $\mathfrak{a}_\bullet \in \ell_2$ also the Assumption §16.11 **(os2')** is satisfied with $\tau_{\mathfrak{a},\mathfrak{u}}^2 \leqslant 2\|\mathfrak{a}_\bullet\|_{\ell_2}^2$.

(o) If $\mathfrak{a}_{2j} = \mathfrak{a}_{2j-1} = j^{-a}$, $a \in \mathbb{N}$, $j \in \mathbb{N}$, then $\{h \in \mathbb{L}_2(\lambda_{[0,1]}): \mathrm{U}h \in \ell_2^{\mathfrak{a}}\}$ is a subset of the *Sobolev space* of a-times differentiable periodic functions. Moreover, up to a constant, for any function $h \in \mathbb{L}_2(\lambda_{[0,1]})$ the weighted norm $\|h_\bullet\|_{\mathfrak{a}^{-1}}^2$ equals the $\mathbb{L}_2$-norm of its a-th weak derivative $h^{(a)}$ (Tsybakov [2009]).

(s) If $\mathfrak{a}_j = \exp(-j^{2a})$, $a > 1/2$, $j \in \mathbb{N}$, then $\{h \in \mathbb{L}_2(\lambda_{[0,1]}): \mathrm{U}h \in \ell_2^{\mathfrak{a}}\}$ is a *class of analytic functions* (Kawata [1972]).

In Table 02 [§12] (Illustration §12.26) the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is depict for the two cases **(o)** and **(s)**. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^+$. □

§16|02 Local and maximal local ϕ-risk

We measure secondly the accuracy of the OPE $\widehat{\mathbb{p}}_\bullet^m = \widehat{\mathbb{p}}_\bullet \mathbb{1}_\bullet^m$ of $\mathbb{p}_\bullet^m = \mathbb{p}_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m$ with $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2$ by a local mean-ϕ-error, i.e. ϕ-risk.

§16.19 **Reminder**. If $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}}) := \{a_\bullet \in \ell_2: \phi_\bullet a_\bullet \in \ell_1\}$, then we have $\mathbb{p}_\bullet^m = \mathbb{p}_\bullet \mathbb{1}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ too and $|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet - \mathbb{p}_\bullet^m)| = \mathrm{o}(1)$ as $m \to \infty$ (Property §11.22). Moreover, $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ given in Definition §15.08 satisfies $\varepsilon_\bullet \mathbb{1}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ (note that $\phi_\bullet \mathbb{1}_\bullet^m, \varepsilon_\bullet \mathbb{1}_\bullet^m \in \ell_2$) and thus also

$$n^{-1/2}\varepsilon_\bullet \mathbb{1}_\bullet^m + \mathbb{p}_\bullet^m = \widehat{\mathbb{p}}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}}). \tag{16.05}$$

Finally, under Assumptions §15.02 and §15.05 and $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ due to Lemma §15.10 (i) the process $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ admits a covariance operator $\Gamma_{\mathbb{p}} \in \mathbb{L}(\ell_2)$, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma_{\mathbb{p}})}$, satisfying $\|\Gamma_{\mathbb{p}}\|_{\mathbb{L}(\ell_2)} \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}$. □

§16|02|01 Local ϕ-risk

§16.20 **Proposition** *(Upper bound). Let Assumptions §15.02 and §15.05,* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *and* $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *be satisfied and for all* $n, m \in \mathbb{N}$ *set*

$$\mathrm{R}_n^m(\mathbb{p}_\bullet, \phi_\bullet) := |\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\circ := \arg\min\{\mathrm{R}_n^m(\mathbb{p}_\bullet, \phi_\bullet): m \in \mathbb{N}\}$$
$$\textit{and} \quad \mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\circ}(\mathbb{p}_\bullet, \phi_\bullet) := \min\{\mathrm{R}_n^m(\mathbb{p}_\bullet, \phi_\bullet): m \in \mathbb{N}\}. \tag{16.06}$$

If $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *then we have* $\mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^{m_n^\circ} - \mathbb{p}_\bullet)|^2) \leqslant \|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\,\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$.

§16.21 **Proof** of Proposition §16.20. is given in the lecture. □

§16.22 **Oracle inequality**. *Under Assumptions §15.02 and §15.05 let* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *and* $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$. *If in addition* $\mathbb{v}_{\mathbb{p}} := \max(\|\mathbb{p}_\bullet\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathbb{p}_\bullet^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ *then* $\max(\|\Gamma_{\mathbb{p}}\|_{\mathbb{L}(\ell_2)}, \|\Gamma_{\mathbb{p}}^{-1}\|_{\mathbb{L}(\ell_2)}) \leqslant \mathbb{v}_{\mathbb{p}}$ *(see Lemma §15.10), and hence Property §12.36 implies*

$$\mathbb{v}_{\mathbb{p}}^{-1}\mathrm{R}_n^m(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^m - \mathbb{p}_\bullet)|^2) = n^{-1}\|\phi_\bullet \mathbb{1}_\bullet^m\|_{\Gamma_{\mathbb{p}}}^2 + |\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}_\bullet^{m|\perp})|^2$$
$$\leqslant \mathbb{v}_{\mathbb{p}}\mathrm{R}_n^m(\mathbb{p}_\bullet, \phi_\bullet) \quad \textit{for all } m, n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality (see Definition §12.34)

$$\mathbb{v}_{\mathbb{p}}^{-1}\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^m - \mathbb{p}_\bullet)|^2) \leqslant \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^{m_n^\circ} - \mathbb{p}_\bullet)|^2)$$
$$\leqslant \mathbb{v}_{\mathbb{p}}\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \mathbb{v}_{\mathbb{p}}^2 \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^m - \mathbb{p}_\bullet)|^2), \quad (16.07)$$

and hence, $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$, m_n° *and the statistic* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$, *respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant* $\mathbb{v}_{\mathbb{p}}^2$*). We observe that* $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) = \mathrm{o}(1)$ *as* $n \to \infty$ *(Remark §12.37), and thus,* $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$ *is an oracle rate. However, note that the oracle dimension* $m_n^\circ = m_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$ *depends on the unknown density coefficients* $\mathbb{p}_\bullet$, *and thus also the oracle optimal statistic* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$. *In other words* $\widehat{\mathbb{p}}_\bullet^{m_n^\circ}$ *is not a feasible estimator.* □

§16.23 **Illustration**. We illustrate the last results considering usual behaviour for both the variance and the bias term. Similar to the two cases **(p)** and **(np)** in Illustration §16.10 we distinguish here the following two cases

(p) $\phi_\bullet \in \ell_2$ or there is $K \in \mathbb{N}$ with $\sup\{|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m \in \mathbb{N} \cap [K, \infty)\} = 0$,

(np) $\phi_\bullet \notin \ell_2$ and for all $m \in \mathbb{N}$ holds $\sup\{|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m \in \mathbb{N} \cap [K, \infty)\} \in \mathbb{R}^+_{\setminus 0}$.

In case **(p)** the oracle bound is again parametric, i.e. $n\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) = \mathrm{O}(1)$, while in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet) = \infty$. In case **(np)** consider the following two specifications

Table 02 [§16]

Order of the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$ as $n \to \infty$

	$(j \in \mathbb{N})$ $\phi_j = j^{v-1/2}$	$(a \in \mathbb{R}^+_{\setminus 0})$ $\mathbb{p}_j$	(squarred bias) $\lvert\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}_\bullet^{m\lvert\perp})\rvert^2$	(variance) $\lVert\mathbb{1}_\bullet^m\rVert_\phi^2$	m_n°	$\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$
(o)	$v \in (0, a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$
	$v = 0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a}}$	$\frac{\log n}{n}$
(s)	$v \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+}e^{-2m^{2a}}$	m^{2v}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$
	$v = 0$	$e^{-j^{2a}}$	$m^{(1-2a)_+}e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 02 [§16] the order of the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$ is depict for $v \geqslant 0$ only. For $v < 0$ the oracle rate $\mathrm{R}_n^\circ(\mathbb{p}_\bullet, \phi_\bullet)$ is parametric. □

§16|02|02 Maximal local ϕ-risk

§16.24 **Assumption**. Consider $\phi_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\phi)_\bullet := (\mathfrak{a}_j\phi_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\phi_\bullet \in \ell_2$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet^2 = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2')** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2\mathrm{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \tau_{\mathfrak{a},\mathrm{u}}^2$ for $\tau_{\mathfrak{a},\mathrm{u}} \in [1, \infty)$. □

§16.25 **Reminder**. Under Assumption §16.24 we have $\ell_2^{\mathfrak{a}} = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}^{-1}}) = \ell_2\mathfrak{a}_\bullet \subseteq \ell_2$ and the three measures $\nu_{\mathbb{N}}$, $\mathfrak{a}_\bullet^{-2}\nu_{\mathbb{N}}$ and $|\phi_\bullet|\nu_{\mathbb{N}}$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\ell_2^{\mathfrak{a}_\bullet}$ endowed with $\|\cdot\|_{\mathfrak{a}^{-1}} = \|\mathrm{M}_{\mathfrak{a}^{-1}_\bullet}\cdot\|_{\ell_2}$ and given a constant $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}} := \{a_\bullet \in \ell_2^{\mathfrak{a}_\bullet} : \|a_\bullet\|_{\mathfrak{a}^{-1}} \leqslant \mathrm{r}\} \subseteq \ell_2^{\mathfrak{a}}$. Since $(\mathfrak{a}\phi)_\bullet \in \ell_2$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ ($\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence) we have $\ell_2^{\mathfrak{a}} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ (Property §11.27), and $|\phi\nu_{\mathbb{N}}(\mathfrak{p}_\bullet \mathbb{1}_\bullet^{m|\perp})| \leqslant \mathrm{r}\,\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi$ for all $\mathfrak{p}_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.29). □

§16.26 **Remark**. We replace Assumption §15.05 **(os1)** and **(os2)**, respectively, by the stronger Assumption §16.24 **(os1')** and **(os2')** (see Remark §16.13). Moreover, considering the set $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ of densities in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (16.04) we have $\|\mathfrak{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant 1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}$ for all $\mathfrak{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ due to Lemma §16.14. □

§16.27 **Proposition** *(Upper bound). Let Assumptions §15.02 and §16.24 be satisfied. For $n \in \mathbb{N}$ considering $m_n^\star \in \mathbb{N}$ and $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \in \mathbb{R}^+$ as in (12.13) (Proposition §12.42) we have*

$$\sup\big\{\mathbb{P}_{\mathfrak{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathfrak{p}}_\bullet^{m} - \mathfrak{p}_\bullet)|^2) \colon \mathfrak{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\big\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet).$$

with constant $\mathrm{C} = (1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}) \vee \mathrm{r}^2$.

§16.28 **Proof** of Proposition §16.27. is given in the lecture. □

§16.29 **Illustration**. Consider the *trigonometric basis* as in Illustration §16.18 which satisfies Assumption §16.24 for all $\mathfrak{a}_\bullet \in \ell_2$. In Table 04 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §12.47. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}^+_{\setminus 0}$. □

§17 Minimax optimal density estimation

§17|01 Maximal local ϕ-risk

§17.01 **Reminder** *(Maximal local ϕ-risk).* Under Assumptions §15.02 and §16.24 the observable noisy density coefficients $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\boldsymbol{\varepsilon}_\bullet$ of the density coefficients $\mathfrak{p}_\bullet = \mathrm{U}\,\mathfrak{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\boldsymbol{\varepsilon}_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08. Under Assumptions §15.02 and §16.24 in Proposition §16.27 an upper bound for a maximal local ϕ-risk of an OPE over the class $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ of densities in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (16.04) is shown. More precisely, the performance of the OPE $\widehat{\mathfrak{p}}_\bullet^{m} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{m} \in \ell_2\mathbb{1}_\bullet^{m} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m \in \mathbb{N}$ is measured by its maximal local ϕ-risk, that is

$$\mathcal{R}_n^{\phi}[\,\widehat{\mathfrak{p}}_\bullet^{m} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathbb{P}_{\mathfrak{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathfrak{p}}_\bullet^{m} - \mathfrak{p}_\bullet)|^2\big) \colon \mathfrak{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.13) (Proposition §12.42) where for $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (17.01)$$

By Proposition §16.27 under Assumptions §15.02 and §16.24 the maximal local ϕ-risk of an OPE $\widehat{\mathfrak{p}}_\bullet^{m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (17.01) satisfies

$$\mathcal{R}_n^{\phi}[\,\widehat{\mathfrak{p}}_\bullet^{m_n^\star} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$$

with $\mathrm{C} = (1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}) \vee \mathrm{r}^2$. □

§17.02 **Lemma** *(Lower bound based on two hypotheses).* *If there are* $\mathbb{p}^0, \mathbb{p}^1 \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ *with associated probability measures* $\mathbb{P}_0 := \mathbb{P}_{\mathbb{p}^0}$ *and* $\mathbb{P}_1 := \mathbb{P}_{\mathbb{p}^1}$ *such that* $\mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leqslant 2n^{-1}$ *then for all* $n \geqslant 2$ *we have*

$$\inf_{\widetilde{\mathbb{p}}_\bullet} \mathcal{R}_n^{\phi}[\,\widetilde{\mathbb{p}}_\bullet \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant \frac{1}{64} |\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^0 - \mathbb{p}_\bullet^1)|^2.$$

where the infimum is taken over all possible estimators.

§17.03 **Proof** of Lemma §17.02. is given in the lecture. □

§17.04 **Remark**. If we consider furthermore candidate densities $\mathbb{p}^0 := \mathbb{1}_{[0,1]} + \mathrm{U}^\star \mathbb{p}_\bullet^*$ and $\mathbb{p}^1 = \mathbb{1}_{[0,1]} - \mathrm{U}^\star \mathbb{p}_\bullet^*$ for some $\mathbb{p}_\bullet^* \in \ell_2^{\mathfrak{a},\mathrm{r}}$, and hence by definition $\mathbb{p}^0, \mathbb{p}^1 \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$, then trivially $|\phi\nu(\mathbb{p}^0 - \mathbb{p}^1)|^2 = 4|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^*)|^2$. If the associated probability measures $\mathbb{P}_0 := \mathbb{P}_{\mathbb{p}^0}$ and $\mathbb{P}_1 := \mathbb{P}_{\mathbb{p}^1}$ satisfy $\mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leqslant 2n^{-1}$ then due to Lemma §17.02 for all $n \geqslant 2$ we have

$$\inf_{\widetilde{\mathbb{p}}_\bullet} \mathcal{R}_n^{\phi}[\,\widetilde{\mathbb{p}}_\bullet \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant \frac{1}{16} |\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^*)|^2. \tag{17.02}$$

We find a minimax-optimal lower bound by constructing a candidate $\mathbb{p}_\bullet^* \in \ell_2^{\mathfrak{a},\mathrm{r}}$ that has the largest possible $|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^*)|^2$-value but $\mathbb{P}_{\mathbb{p}^0}^{\otimes n}$ and $\mathbb{P}_{\mathbb{p}^1}^{\otimes n}$ are still statistically indistinguishable in the sense that $\mathrm{H}^2(\mathbb{P}_{\mathbb{p}^0}, \mathbb{P}_{\mathbb{p}^1}) \leqslant 2n^{-1}$. □

§17.05 **Lemma**. *Under Assumption §16.24 let* $\mathbb{p}_\bullet^* \in \ell_2^{\mathfrak{a},\mathrm{r}}$ *with* $\|\mathbb{p}_\bullet^*\|_{\mathfrak{a}^{-1}} \leqslant 1/(2\tau_{\mathfrak{a},\mathrm{u}})$. *Then* $\mathbb{p}^0 := \mathbb{1}_{[0,1]} + \mathrm{U}^\star \mathbb{p}_\bullet^*$ *and* $\mathbb{p}^1 := \mathbb{1}_{[0,1]} - \mathrm{U}^\star \mathbb{p}_\bullet^*$ *belong to* $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$, *and the associated probability measures* $\mathbb{P}_0 := \mathbb{P}_{\mathbb{p}^0}$ *and* $\mathbb{P}_1 := \mathbb{P}_{\mathbb{p}^1}$ *satisfy* $\mathrm{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leqslant 2\|\mathbb{p}_\bullet^*\|_{\ell_2}^2$.

§17.06 **Proof** of Lemma §17.05. is given in the lecture. □

§17.07 **Reminder**. Under Assumption §16.24 let in addition $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\searrow 0}^+)_{\downarrow 0}^{\mathbb{N}}$ (see Notation §13.23), then Assumption §13.24 is satisfied. If $\mathfrak{a}_2^2 > n^{-1}$ then exploiting the definition (17.01) of $m_n^\star$ we have $\mathfrak{a}_{m_n^\star}^2 > n^{-1} \geqslant \mathfrak{a}_{m_n^\star+1}^2$ (see Comment §13.25) which we use in the next proof. □

§17.08 **Proposition** *(Lower bound).* *Let Assumptions §15.02 and §16.24 be satisfied. If* $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\searrow 0}^+)_{\downarrow 0}^{\mathbb{N}}$ *then for all* $n \in \mathbb{N} \cap (1 \vee \mathfrak{a}_2^{-2}, \infty)$ *we have*

$$\inf_{\widetilde{\mathbb{p}}_\bullet} \mathcal{R}_n^{\phi}[\,\widetilde{\mathbb{p}}_\bullet \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant \mathrm{C}\, \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \tag{17.03}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{r}^2 \wedge 1/(4\tau_{\mathfrak{a},\mathrm{u}}^2) \wedge 1)$ *and infimum taken over all estimators.*

§17.09 **Proof** of Proposition §17.08. is given in the lecture. □

§17.10 **Illustration**. Consider the *trigonometric basis* as in Illustration §16.18 which satisfies Assumption §16.24 for all $\mathfrak{a}_\bullet \in \ell_2$ (see Illustration §16.29). In Table 04 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §16.29. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\searrow 0}^+$. In both cases the additional assumption $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\searrow 0}^+)_{\downarrow 0}^{\mathbb{N}}$ is satisfied. Consequently, due to Proposition §17.08 the Table 04 [§12] presents the order of the *minimax rate* $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{\mathbb{p}}_\bullet^{m_n^\star} = \widehat{\mathbb{p}}_\bullet \mathbb{1}_\bullet^{m_n^\star} \in \ell_2 \mathbb{1}_\bullet^{m_n^\star} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with optimally selected dimension $m_n^\star$ (Proposition §16.27). We shall stress, that the order of $m_n^\star$ given in the Table 04 [§12] depends on the parameter $a \in \mathbb{R}_{\searrow 0}^+$ characterising the (abstract) smoothness of the density of interest which is generally not known in advance. □

§17|02 Maximal global $\mathfrak{v}$-risk

§17.11 **Reminder** *(Maximal global $\mathfrak{v}$-risk).* Under Assumptions §15.02 and §16.11 the observable noisy density coefficients $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficients $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08. Under Assumptions §15.02 and §16.11 in Proposition §16.16 an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE over the class $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ of densities in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (16.04) is shown. More precisely, the performance of the OPE $\widehat{\mathbb{p}}_\bullet^{\,m} = \widehat{\mathbb{p}}_\bullet\mathbb{1}_\bullet^m \in \ell_2(\mathfrak{v}_\bullet^2)\mathbb{1}_\bullet^m \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}$-risk, that is

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\mathbb{p}}_\bullet^{\,m}\,|\,\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(\|\widehat{\mathbb{p}}_\bullet^{\,m} - \mathbb{p}_\bullet\|_{\mathfrak{v}}^2\big)\colon \mathbb{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.06) (Proposition §12.21) where for $n, m \in \mathbb{N}$ we have defined $(\mathfrak{a}\mathfrak{v})_{(m)}^2 = \|(\mathfrak{a}\mathfrak{v})_\bullet\mathbb{1}_\bullet^{m|\perp}\|_{\ell_\infty}^2$ and

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) := [(\mathfrak{a}\mathfrak{v})_{(m)}^2 \vee n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2], \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \quad (17.04)$$

By Proposition §16.16 under Assumptions §15.02 and §16.11 the maximal global $\mathfrak{v}_\bullet$-risk of an OPE $\widehat{\mathbb{p}}_\bullet^{\,m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (17.04) satisfies

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\mathbb{p}}_\bullet^{\,m_n^\star}\,|\,\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$$

with $\mathrm{C} = 1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + \mathrm{r}^2$. Furthermore, as in Notation §13.29 for $m \in \mathbb{N}$ we set $\mathcal{T}_m := \{-1,1\}^m$ and for each $\tau := (\tau_j)_{j\in[\![m]\!]} \in \mathcal{T}_m$ and $j \in [\![m]\!]$ we introduce $\tau^{(j)} \in \mathcal{T}_m$ given by $\tau_j^{(j)} := -\tau_j$ and $\tau_l^{(j)} := \tau_l$ for $l \in [\![m]\!]\backslash\{j\}$. □

§17.12 **Lemma** *(Assouad's cube technique).* *If for each $\tau \in \mathcal{T}_m$ there is $\mathbb{p}^\tau \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ with associated probability measure $\mathbb{P}_\tau := \mathbb{P}_{\mathbb{p}^\tau}$ such that for all $\tau \in \mathcal{T}_m$ and $j \in [\![m]\!]$ we have $\mathrm{H}(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 2n^{-1}$ then for all $n \geqslant 2$*

$$\inf_{\widetilde{\mathbb{p}}_\bullet} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{\mathbb{p}}_\bullet\,|\,\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant 2^{-m}\sum_{\tau\in\mathcal{T}_m}\tfrac{1}{64}\sum_{j\in[\![m]\!]}\big(\mathfrak{v}_j^2|\mathbb{p}_j^\tau - \mathbb{p}_j^{\tau^{(j)}}|^2\big)$$

where the infimum is taken over all possible estimators.

§17.13 **Proof** of Lemma §17.12. is given in the lecture. □

§17.14 **Remark.** If we assume furthermore candidate densities $\mathbb{p}^\tau := \mathbb{1}_{[0,1]} + \mathrm{U}^\star\,\mathbb{p}_\bullet^\tau$ with $\mathbb{p}_\bullet^\tau := (\tau_j\,\mathbb{p}_j^\star\mathbb{1}_j^m)_{j\in\mathbb{N}}$, $\tau \in \mathcal{T}_m$, for some $\mathbb{p}_\bullet^\star \in \ell_2^{\mathfrak{a},\mathrm{r}}$, where evidently $\mathbb{p}_\bullet^\tau \in \ell_2^{\mathfrak{a},\mathrm{r}}$ too and hence $\mathbb{p}^\tau \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$, then it is easily seen that $\sum_{j\in[\![m]\!]}\big(\mathfrak{v}_j^2|\,\mathbb{p}_j^\tau - \mathbb{p}_j^{\tau^{(j)}}|^2\big) = 4\|\,\mathbb{p}_\bullet^\star\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$. If for all $\tau \in \mathcal{T}_m$ and $j \in [\![m]\!]$ the associated probability measures $\mathbb{P}_\tau := \mathbb{P}_{\mathbb{p}^\tau}$ and $\mathbb{P}_{\tau^{(j)}} := \mathbb{P}_{\mathbb{p}^{\tau^{(j)}}}$ satisfy $\mathrm{H}(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 2n^{-1}$ then due to Lemma §17.12 for all $n \geqslant 2$ we have

$$\inf_{\widetilde{\mathbb{p}}_\bullet} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{\mathbb{p}}_\bullet\,|\,\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant 2^{-m}\sum_{\tau\in\mathcal{T}_m}\tfrac{1}{16}\|\,\mathbb{p}_\bullet^\star\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 = \tfrac{1}{16}\|\,\mathbb{p}_\bullet^\star\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2. \quad (17.05)$$

We find a minimax-optimal lower bound by choosing the parameter m and the function $\mathbb{p}^*$ that have the largest possible $\|\,\mathbb{p}_\bullet^\star\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2$-value although that the associated $\mathbb{P}_\tau$, $\tau \in \mathcal{T}_m$ are still statistically indistinguishable in the sense that $\mathrm{H}^2(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 2n^{-1}$ for all $j \in [\![m]\!]$ and $\tau \in \mathcal{T}_m$. □

§17.15 **Lemma.** *Under Assumption §16.11 let $\mathbb{p}_\bullet^* \in \ell_2^{\mathfrak{a},\mathrm{r}}$ with $\|\,\mathbb{p}_\bullet^*\|_{\mathfrak{a}_\bullet^{-1}} \leqslant 1/(2\tau_{\mathfrak{a},\mathrm{u}})$. Then for each $\tau \in \mathcal{T}_m$, $\mathbb{p}^\tau := \mathbb{1}_{[0,1]} + \mathrm{U}^\star\,\mathbb{p}_\bullet^\tau$ with $\mathbb{p}_\bullet^\tau := (\tau_j\,\mathbb{p}_j^\star\mathbb{1}_j^m)_{j\in\mathbb{N}}$ belongs to $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$, and for each $j \in [\![m]\!]$ the associated probability measures $\mathbb{P}_\tau := \mathbb{P}_{\mathbb{p}^\tau}$ and $\mathbb{P}_{\tau^{(j)}} := \mathbb{P}_{\mathbb{p}^{\tau^{(j)}}}$ satisfy $\mathrm{H}^2(\mathbb{P}_\tau, \mathbb{P}_{\tau^{(j)}}) \leqslant 2\|\,\mathbb{p}_\bullet^*\mathbb{1}_\bullet^m\|_{\ell_\infty}^2$.*

§17.16 **Proof** of Lemma §17.15. is given in the lecture. □

§17.17 **Reminder**. For $w_\bullet \in \ell_\infty$ we set $w_{(0)}^2 := \|w_\bullet^2\|_{\ell_\infty}$ and $w_{(\bullet)}^2 = (w_{(j)}^2 := \|w_\bullet^2 \mathbb{1}_\bullet^{j|\perp}\|_{\ell_\infty})_{j\in\mathbb{N}}$ (Notation §13.34) where by construction $w_{(j)}^2 = \sup\{w_i^2 \colon i \in \mathbb{N} \cap [j+1,\infty)\}$, $j \in \mathbb{N}_0$ and $w_{(\bullet)}^2 \in (\mathbb{R}_{\backslash 0}^+)_\searrow^{\mathbb{N}}$. Under Assumption §16.11 let in addition $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\backslash 0}^+)_{\downarrow 0}^{\mathbb{N}}$ and there exists $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} \in (0,1]$ such that $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\|(\mathfrak{a}\mathfrak{v})_\bullet^{-2}\mathbb{1}_\bullet^m\|_{\ell_\infty} \leqslant (\mathfrak{a}\mathfrak{v})_{(m-1)}^{-2}$ or in equal

$$(\mathfrak{a}\mathfrak{v})_{(m-1)}^2 \geqslant \min\{(\mathfrak{a}\mathfrak{v})_j^2 \colon j \in [\![m]\!]\} \geqslant \mathrm{C}_{(\mathfrak{a}\mathfrak{v})}(\mathfrak{a}\mathfrak{v})_{(m-1)}^2$$

for all $m \in \mathbb{N}$, then Assumption §13.35 is satisfied. For $m_n^\star$ and $\mathrm{R}_n^\star := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ as in (17.04) we distinguish case i) : $\mathrm{R}_n^\star = n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2 > (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2$ and case ii) : $\mathrm{R}_n^\star = (\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 \geqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$. Due to Comment §13.36 if $(\mathfrak{a}\mathfrak{v})_{(1)}^2 > n^{-1}\mathfrak{v}_1^2$ then in case i) $(\mathfrak{a}\mathfrak{v})_{(m_n^\star-1)}^2 \geqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2$, while in case ii) setting (the defining set is not empty since $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\backslash 0}^+)_{\downarrow 0}^{\mathbb{N}}$)

$$m_n^\circ := \min\{m \in \mathbb{N} \cap [m_n^\star + 1, \infty) \colon n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \geqslant (\mathfrak{a}\mathfrak{v})_{(m)}^2\} \tag{17.06}$$

we have $(\mathfrak{a}\mathfrak{v})_{(m_n^\star)}^2 = (\mathfrak{a}\mathfrak{v})_{(m_n^\circ-1)}^2 \leqslant n^{-1}\|\mathbb{1}_\bullet^{m_n^\circ}\|_{\mathfrak{v}}^2$. We use those estimates in the next proof. □

§17.18 **Proposition** *(Lower bound). Let Assumptions §15.02 and §16.11 be satisfied. If* $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\backslash 0}^+)_{\downarrow 0}^{\mathbb{N}}$ *and there exists* $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} \in (0,1]$ *such that* $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\|(\mathfrak{a}\mathfrak{v})_\bullet^{-2}\mathbb{1}_\bullet^m\|_{\ell_\infty} \leqslant (\mathfrak{a}\mathfrak{v})_{(m-1)}^{-2}$ *for all* $m \in \mathbb{N}$, *then for all* $n \in \mathbb{N} \cap (1 \vee \mathfrak{v}_1^2(\mathfrak{a}\mathfrak{v})_{(1)}^{-2}, \infty)$ *we have*

$$\inf_{\widetilde{\mathbb{p}}} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{\mathbb{p}}_\bullet \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \geqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \tag{17.07}$$

with constant $\mathrm{C} := (\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}/16)(\mathrm{r}^2 \wedge 1/(4\tau_{\mathfrak{a},\mathfrak{u}}^2) \wedge 1)$ *and infimum taken over all estimators.*

§17.19 **Proof** of Proposition §17.18. is given in the lecture. □

§17.20 **Illustration**. Consider the *trigonometric basis* as in Illustration §16.18 which satisfies Assumption §16.11 for all $\mathfrak{a}_\bullet \in \ell_2$ (see Illustration §16.18). In Table 02 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §16.18. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\backslash 0}^+$. In both cases the additional assumptions, $(\mathfrak{a}\mathfrak{v})_{(\bullet)}^2 \in (\mathbb{R}_{\backslash 0}^+)_{\downarrow 0}^{\mathbb{N}}$ and there exists $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} \in (0,1]$ such that $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\|(\mathfrak{a}\mathfrak{v})_\bullet^{-2}\mathbb{1}_\bullet^m\|_{\ell_\infty} \leqslant (\mathfrak{a}\mathfrak{v})_{(m-1)}^{-2}$ for all $m \in \mathbb{N}$, are satisfied. Consequently, due to Proposition §17.18 the Table 02 [§12] presents the order of the *minimax rate* $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{\mathbb{p}}^{m_n^\star} = \widehat{\mathbb{p}}_\bullet \mathbb{1}_\bullet^{m_n^\star} \in \ell_2\mathbb{1}_\bullet^{m_n^\star} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with optimally selected dimension $m_n^\star$ (Proposition §16.16). We shall stress, that the order of $m_n^\star$ given in the Table 02 [§12] depends on the parameter $a \in \mathbb{R}_{\backslash 0}^+$ characterising the (abstract) smoothness of the density of interest which is generally not known in advance. □

§18 Data-driven density estimation

§18|01 Data-driven global estimation by model selection

The next assertion provides our key argument in order to control the deviations of the reminder term. The inequality is due to Talagrand [1996] and in this form for example given in Klein and Rio [2005].

§18.01 **Lemma** *(Talagrand's inequality). Let* $(Z_i)_{i\in[\![n]\!]}$ *be independent* $(\mathcal{Z}, \mathscr{Z})$*-valued random variables and let* $\{\mathrm{r}_t \colon t \in \mathcal{T}\} \subseteq \mathscr{Z}$ *be a countable class of Borel-measurable functions. For* $t \in \mathcal{T}$ *setting* $\overline{\mathrm{r}}_t = n^{-1}\sum_{i\in[\![n]\!]}\{\mathrm{r}_t(Z_i) - \mathbb{E}(\mathrm{r}_t(Z_i))\}$ *we have*

$$\mathbb{E}\big(\big(\sup\{|\overline{\mathrm{r}}_t|^2 \colon t \in \mathcal{T}\} - 6\mathrm{H}^2\big)_+\big) \leqslant \mathrm{C}_{\mathrm{tal}}\Big\{\frac{\mathrm{v}}{n}\exp\Big(\frac{-n\mathrm{H}^2}{6\mathrm{v}}\Big) + \frac{\mathrm{h}^2}{n^2}\exp\Big(\frac{-n\mathrm{H}}{100\mathrm{h}}\Big)\Big\} \tag{18.01}$$

for some universal numerical constant $\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ *and where*

$$\sup\big\{|\mathrm{r}_t(z)|\colon t\in\mathcal{T}, z\in\mathcal{Z}\big\}\leqslant \mathrm{h},\ \mathbb{E}\big(\sup\big\{|\bar{\mathrm{r}}_t|\colon t\in\mathcal{T}\big\}\big)\leqslant \mathrm{H},\ \sup\big\{n\mathbb{E}\big(|\bar{\mathrm{r}}_t|^2\big)\colon t\in\mathcal{T}\big\}\leqslant \mathrm{v}. \tag{18.02}$$

§18.02 **Remark**. Let us briefly reconsider the OPE $\widehat{\mathbb{p}}^m_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^m_\bullet \in \ell_2\mathbb{1}^m_\bullet$ with dimension $m\in\mathbb{N}$ (Definition §16.04) where $\widehat{\mathbb{p}}_\bullet = \widehat{\mathbb{P}}_n\mathrm{u}_\bullet = (\widehat{\mathbb{P}}_n\mathrm{u}_j)_{j\in\mathbb{N}}$ are noisy versions (Definition §15.08) of the density coefficients $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} = \mathbb{P}_{\mathbb{p}}(\mathrm{u}_\bullet) = (\mathbb{P}_{\mathbb{p}}\mathrm{u}_j = \lambda_{[0,1]}(\mathbb{p}\,\mathrm{u}_j))_{j\in\mathbb{N}}$. For $m\in\mathbb{N}$ introduce the unit ball $\mathbb{B}_m := \big\{a_\bullet\in\ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet\colon \|a_\bullet\|_{\mathfrak{v}}\leqslant 1\big\}$ contained in the linear subspace $\ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet$ spanned by $(\mathbb{1}^{\{j\}}_\bullet)_{j\in[\![m]\!]}$. Clearly, for each $a_\bullet\in\ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet$ we have $\mathrm{r}_{a_\bullet} := \sum_{j\in[\![m]\!]}\mathfrak{v}^2_j a_j\mathrm{u}_j = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet\mathrm{u}_\bullet)\in\mathscr{B}_{[0,1]}$, i.e. it is a $\mathscr{B}_{[0,1]}$-$\mathscr{B}$-measurable function, where $\widehat{\mathbb{P}}_n(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet\widehat{\mathbb{P}}_n\mathrm{u}_\bullet) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet\widehat{\mathbb{p}}_\bullet)$, $\mathbb{P}_{\mathbb{p}}(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet\mathbb{P}_{\mathbb{p}}\mathrm{u}_\bullet) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet\mathbb{p}_\bullet)$ and hence $\bar{\mathrm{r}}_{a_\bullet} = \widehat{\mathbb{P}}_n(\mathrm{r}_{a_\bullet}) - \mathbb{P}_{\mathbb{p}}(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet(\widehat{\mathbb{p}}_\bullet - \mathbb{p}_\bullet)) = \langle\widehat{\mathbb{p}}_\bullet - \mathbb{p}_\bullet, a_\bullet\rangle_{\mathfrak{v}}$. Consequently, we obtain

$$\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}^m_\bullet\|^2_{\mathfrak{v}} = \sup\big\{|\langle\widehat{\mathbb{p}}_\bullet - \mathbb{p}_\bullet, a_\bullet\rangle_{\mathfrak{v}}|^2\colon a_\bullet\in\mathbb{B}_m\big\} = \sup\big\{|\bar{\mathrm{r}}_{a_\bullet}|^2\colon a_\bullet\in\mathbb{B}_m\big\}$$

The last identity provides the necessary argument to apply below Talagrand's inequality (§18.01). Note that, the unit ball $\mathbb{B}_m$ is not a countable set, however, it contains a countable dense subset, say $\mathcal{B}_m$, since $\ell_2(\mathfrak{v}^2_\bullet)$ is separable. Exploiting the continuity of the inner product it is straightforward to see that $\sup\big\{|\langle b_\bullet, a_\bullet\rangle_{\mathfrak{v}}|^2\colon a_\bullet\in\mathbb{B}_m\big\} = \sup\big\{|\langle b_\bullet, a_\bullet\rangle_{\mathfrak{v}}|^2\colon a_\bullet\in\mathcal{B}_m\big\}$ for all $b_\bullet\in\ell_2(\mathfrak{v}^2_\bullet)$. Consequently, provided that

$$\begin{aligned}\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}^m_\bullet\|_{\mathfrak{v}}\colon x\in[0,1]\big\} &= \sup\big\{|\mathrm{r}_{a_\bullet}(x)|\colon a_\bullet\in\mathcal{B}_m, x\in[0,1]\big\}\leqslant \mathrm{h},\\ \mathbb{P}^{\otimes n}_{\mathbb{p}}\big(\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}^m_\bullet\|^2_{\mathfrak{v}}\big) &= \mathbb{P}^{\otimes n}_{\mathbb{p}}\big(\sup\big\{|\bar{\mathrm{r}}_{a_\bullet}|^2\colon a_\bullet\in\mathcal{B}_m\big\}\big)\leqslant \mathrm{H}^2,\\ \sup\big\{\mathbb{P}_{\mathbb{p}}\big(|\nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet(\mathrm{u}_\bullet - \mathbb{P}_{\mathbb{p}}\mathrm{u}_\bullet))|^2\big)\colon a_\bullet\in\mathbb{B}_m\big\} &= \sup\big\{n\mathbb{P}^{\otimes n}_{\mathbb{p}}\big(|\bar{\mathrm{r}}_{a_\bullet}|^2\big)\colon a_\bullet\in\mathcal{B}_m\big\}\leqslant \mathrm{v}.\end{aligned} \tag{18.03}$$

due to Talagrand's inequality (§18.01) we have

$$\mathbb{P}^{\otimes n}_{\mathbb{p}}\Big(\big(\|\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}^m_\bullet\|^2_{\mathfrak{v}} - 6\mathrm{H}^2\big)_+\Big)\leqslant \mathrm{C}_{\mathrm{tal}}\Big\{\tfrac{\mathrm{v}}{n}\exp\Big(\tfrac{-n\mathrm{H}^2}{6\mathrm{v}}\Big) + \tfrac{\mathrm{h}^2}{n^2}\exp\Big(\tfrac{-n\mathrm{H}}{100\mathrm{h}}\Big)\Big\} \tag{18.04}$$

for some universal numerical constant $\mathrm{C}_{\mathrm{tal}}\in[1,\infty)$. □

§18|01|01 Global $\mathfrak{v}$-risk

§18.03 **Assumption**. The weights $\mathfrak{v}_\bullet\in(\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ satisfy

$$\forall x\in\mathbb{R}^+_{\setminus 0}:\quad \sum_{m\in\mathbb{N}}\big\{x\|\mathfrak{v}^2_\bullet\mathbb{1}^m_\bullet\|_{\ell_\infty}\exp\big(-\|\mathfrak{v}_\bullet\mathbb{1}^m_\bullet\|^2_{\ell_2}/(x\|\mathfrak{v}^2_\bullet\mathbb{1}^m_\bullet\|_{\ell_\infty})\big)\big\} =: \mathrm{C}_{\mathfrak{v}}(x)\in\mathbb{R}^+. \tag{18.05}$$

The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet = (\mathrm{u}_j)_{j\in\mathbb{N}}$ on $([0,1],\mathscr{B}_{[0,1]})$ for all $m\in\mathbb{N}$ satisfies **(os2'')** $\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}_\bullet}\colon x\in[0,1]\big\}\leqslant\tau^2_{\mathfrak{v},\mathrm{u}}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}_\bullet}\in\mathbb{R}^+$ for $\tau_{\mathfrak{v},\mathrm{u}}\in[1,\infty)$. □

§18.04 **Remark**. We replace Assumption §15.05 **(os1)** and **(os2)**, respectively, by the stronger Assumption §18.03 **(os1')** and **(os2'')**. Indeed, under **(os1')** we have **(os1)** $\mathbb{1}_{[0,1]}\in\ker(\mathrm{U})$. Furthermore, $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs to $\mathbb{L}_\infty(\lambda_{[0,1]})$ due to **(os2'')** (and $\mathfrak{v}_\bullet\in(\mathbb{R}_{\setminus 0})^{\mathbb{N}}$), and hence **(os2)** is fulfilled (see also Remark §15.06). Under Assumption §18.03 (18.05) we have $\|\mathfrak{v}_\bullet\mathbb{1}^m_\bullet\|^{-2}_{\ell_2} = \mathrm{o}(1)$ as $m\to\infty$ (Comment §14.22), see also Illustration §14.23 for an example when (18.05) is not satisfied. □

§18.05 **Reminder** *(Global oracle $\mathfrak{v}$-risk)*. Given Assumptions §15.02 and §18.03 we consider an OPE as in Definition §16.04. Here the observable noisy density coefficients $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of

the density coefficients $\mathfrak{p}_\bullet = \mathrm{U}\,\mathfrak{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08. Under Assumptions §15.02 and §18.03, (and hence Assumption §15.05 and $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ see Remark §18.04) and $\mathfrak{p}_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ in §16.09 an *oracle inequality* for the global $\mathfrak{v}$-risk of the OPE's is shown. More presicely, as in (16.02) (Proposition §16.07) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}_n^m(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) := \|\mathfrak{p}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\circ}(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \tag{18.06}$$

and assuming $\mathbb{v}_{\mathfrak{p}} := \max(\|\mathfrak{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathfrak{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}_{\setminus 0}^+$ due to Property §16.09 the (infeasible) OPE $\widehat{\mathfrak{p}}_\bullet^{m_n^\circ} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{m_n^\circ} \in \ell_2(\mathfrak{v}_\bullet^2)\mathbb{1}_\bullet^{m_n^\circ} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with oracle dimension m_n° as in (18.06) satisfies

$$\mathbb{v}_{\mathfrak{p}}^{-1}\,\mathrm{R}_n^\circ(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathfrak{p}}^{\otimes n}(\|\widehat{\mathfrak{p}}_\bullet^m - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2) \leqslant \mathbb{P}_{\mathfrak{p}}^{\otimes n}(\|\widehat{\mathfrak{p}}_\bullet^{m_n^\circ} - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2)$$
$$\leqslant \mathbb{v}_{\mathfrak{p}}\,\mathrm{R}_n^\circ(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{v}_{\mathfrak{p}}^2 \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathfrak{p}}^{\otimes n}(\|\widehat{\mathfrak{p}}_\bullet^m - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2),$$

and hence it is *oracle optimal* (with constant $\mathbb{v}_{\mathfrak{p}}^2$). □

§18.06 **Notation**. Consider a sequence of penalties $\mathfrak{pen}_\bullet^{\mathfrak{v}} = (\mathfrak{pen}_m^{\mathfrak{v}})_{m\in\mathbb{N}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ given by

$$\mathfrak{pen}_m^{\mathfrak{v}} := 24\,\tau_{\mathfrak{v},\mathrm{u}}^2 n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \qquad \text{for each } m \in \mathbb{N} \tag{18.07}$$

and the upper bound (where the defining set is not empty)

$$\mathrm{M}^{\mathfrak{v}} := \max\big\{m \in \mathbb{N}\colon \|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \leqslant n\mathfrak{v}_1^2,\ m \leqslant \exp(\tfrac{n^{1/2}}{100})\big\} \tag{18.08}$$

which are obviously known in advance. Considering the data-driven OSE $\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter

$$\widehat{m} := \arg\min\big\{-\|\widehat{\mathfrak{p}}_\bullet^m\|_{\mathfrak{v}}^2 + \mathfrak{pen}_m^{\mathfrak{v}}\colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\big\} \tag{18.09}$$

we derive below an upper bound for its global $\mathfrak{v}$-risk, $\mathbb{P}_{\mathfrak{p}}^{\otimes n}\big(\|\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2\big)$. □

§18.07 **Lemma**. *Under Assumptions §15.02 and §18.03 and* $\mathfrak{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\mathfrak{pen}_\bullet^{\mathfrak{v}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ *as in* (18.07) *and* $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ *as in* (18.08) *we have*

$$\mathbb{P}_{\mathfrak{p}}^{\otimes n}\Big(\max\Big\{\big(\|\widehat{\mathfrak{p}}_\bullet^m - \mathfrak{p}_\bullet^m\|_{\mathfrak{v}}^2 - \mathfrak{pen}_m^{\mathfrak{v}}/4\big)_+ : m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\Big\}\Big) \leqslant \mathrm{C}_{\mathrm{tal}}\tau_{\mathfrak{v},\mathrm{u}}^2\big(\mathrm{C}_{\mathfrak{v}}(x_{\mathfrak{p}}) + \mathfrak{v}_1^2\big)\,n^{-1} \tag{18.10}$$

for some universal numerical constant $\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ *and* $x_{\mathfrak{p}} := 6\|\mathfrak{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\tau_{\mathfrak{v},\mathrm{u}}^{-2} \in \mathbb{R}^+$.

§18.08 **Proof** of Lemma §18.07. is given in the lecture. □

§18.09 **Proposition** *(Upper bound). Under Assumptions §15.02 and §18.03 and* $\mathfrak{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ *as in* (18.08) *and* $\mathfrak{pen}_\bullet^{\mathfrak{v}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ *as in* (18.07) *the data-driven OPE* $\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2(\mathfrak{v}_\bullet^2)\mathbb{1}_\bullet^{\widehat{m}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ *of* $\mathfrak{p}_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ *with data-driven dimension* $\widehat{m} \in [\![\mathrm{M}^{\mathfrak{v}}]\!]$ *as in* (18.09) *satisfies*

$$\mathbb{P}_{\mathfrak{p}}^{\otimes n}\big(\|\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2\big) \leqslant 96\tau_{\mathfrak{v},\mathrm{u}}^2 \min\big\{\mathrm{R}_n^m(\mathfrak{p}_\bullet, \mathfrak{v}_\bullet)\colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\big\} + \mathrm{C}\tau_{\mathfrak{v},\mathrm{u}}^2\big(\mathrm{C}_{\mathfrak{v}}(x_{\mathfrak{p}}) + \mathfrak{v}_1^2\big)\,n^{-1} \tag{18.11}$$

for some universal numerical constant $\mathrm{C} = 8\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ *and* $x_{\mathfrak{p}} := 6\|\mathfrak{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\tau_{\mathfrak{v},\mathrm{u}}^{-2} \in \mathbb{R}^+$.

§18.10 **Proof** of Proposition §18.09. is given in the lecture. □

§18|01|02 Maximal global $\mathfrak{v}$-risk

§18.11 **Assumption.** Consider weights $\mathfrak{a}_\bullet, \mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\mathfrak{v})_\bullet := (\mathfrak{a}_j \mathfrak{v}_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet \mathfrak{v}_\bullet \in \ell_\infty$. We write $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. The weights $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ satisfy (18.05). The orthonormal system $(\mathfrak{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathfrak{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathfrak{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathfrak{u}_\bullet^2 = (\mathfrak{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2')** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2 \mathfrak{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \tau_{\mathfrak{a},\mathfrak{u}}^2$ and **(os2")** $\sup\{\|\mathfrak{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 : x \in [0,1]\} \leqslant \tau_{\mathfrak{v},\mathfrak{u}} \|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \in \mathbb{R}^+$ for $\tau_{\mathfrak{a},\mathfrak{u}}, \tau_{\mathfrak{v},\mathfrak{u}} \in [1,\infty)$. □

§18.12 **Reminder** *(Maximal global $\mathfrak{v}$-risk).* Given Assumptions §15.02 and §18.11 we consider an OPE as in Definition §16.04. Here the observable noisy density coefficients $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficients $\mathfrak{p}_\bullet = \mathrm{U}\,\mathfrak{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08. Under Assumptions §15.02 and §18.11 in Proposition §16.16 an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE is shown over the set $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ given in (16.04) (Lemma §16.14). More precisely, the performance of the OPE $\widehat{\mathfrak{p}}_\bullet^m = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^m \in \ell_{2(\mathfrak{v}_\bullet^2)}\mathbb{1}_\bullet^m \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}$-risk over the ellipsoid $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\mathfrak{p}}_\bullet^m \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\Big\{\mathbb{P}_{\mathfrak{p}}^{\otimes n}\big(\|\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} - \mathfrak{p}_\bullet\|_{\mathfrak{v}}^2\big) : \mathfrak{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\Big\}.$$

As in (12.06) for $n, m \in \mathbb{N}$ setting $(\mathfrak{a}\mathfrak{v})_{(m)}^2 := \|(\mathfrak{a}\mathfrak{v})_\bullet^2 \mathbb{1}_\bullet^{m|\perp}\|_{\ell_\infty}$ and

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := (\mathfrak{a}\mathfrak{v})_{(m)}^2 \vee n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\} \tag{18.12}$$

by Proposition §16.16 under Assumptions §15.02 and §18.03 the maximal global $\mathfrak{v}$-risk of an OPE $\widehat{\mathfrak{p}}^{m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (18.12) satisfies

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\theta}_\bullet^{m_n^\star} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$$

with $\mathrm{C} = 1 + \mathrm{r}\tau_{\mathfrak{a},\mathfrak{u}} + \mathrm{r}^2$. Moreover, due to Proposition §17.18 $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ provides (up to a constant) also a lower bound of the maximal global $\mathfrak{v}$-risk over the ellipsoid $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is a minimax bound and $\widehat{\mathfrak{p}}^{m_n^\star}$ is minimax optimal. However, the optimal dimension $m_n^\star$ depends on $\mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ characterising the ellipsoid $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$. □

§18.13 **Proposition** *(Upper bound).* *Under Assumptions §15.02 and §18.03 for $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ as in (18.08) and $\mathfrak{pen}_\bullet^{\mathfrak{v}} \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ as in (18.07) the data-driven OPE $\widehat{\mathfrak{p}}^{\widehat{m}} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_{2(\mathfrak{v}_\bullet^2)}\mathbb{1}_\bullet^{\widehat{m}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with data-driven dimension $\widehat{m} \in [\![\mathrm{M}^{\mathfrak{v}}]\!]$ as in (18.09) satisfies*

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{\mathfrak{p}}^{\widehat{m}} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant (3\mathrm{r}^2 + 96\tau_{\mathfrak{v},\mathfrak{u}}^2) \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\big\} + \mathrm{C}\tau_{\mathfrak{v},\mathfrak{u}}^2\big(\mathrm{C}_{\mathfrak{v}}(\xi) + \mathfrak{v}_1^2\big)\, n^{-1} \tag{18.13}$$

for some universal numerical constant $\mathrm{C} = 8\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ and $\xi := 6(1 + \mathrm{r}\tau_{\mathfrak{a},\mathfrak{u}})\tau_{\mathfrak{v},\mathfrak{u}}^{-2} \in \mathbb{R}^+$.

§18.14 **Proof** of Proposition §18.13. is given in the lecture. □

§18.15 **Comment.** The minimax bound $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$ (for details see Reminder §18.12) satisfies $n\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \geqslant \|\mathbb{1}_\bullet^{m_n^\star}\|_{\mathfrak{v}}^2 \geqslant \mathfrak{v}_1^2$. Consequently, the last upper bound in (18.13) and the minimax bound $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ coincide up to a constant $(3\mathrm{r}^2 + 96\tau_{\mathfrak{v},\mathfrak{u}}^2 + \mathrm{C}\tau_{\mathfrak{v},\mathfrak{u}}^2\big(\mathrm{C}_{\mathfrak{v}}(x_{\mathfrak{a}})\mathfrak{v}_1^{-2} + 1\big))$ provided the minimax dimension fulfils $m_n^\star \in [\![\mathrm{M}^{\mathfrak{v}}]\!]$. Therefore, we wish the upper bound $\mathrm{M}^{\mathfrak{v}}$ to be as large as possible. The next assertion shows that $\mathrm{M}^{\mathfrak{v}}$ as in (18.08) is a suitable choice for the upper bound. □

§18.16 **Corollary**. *Under the assumptions of Proposition §18.13 for each* $n \in \mathbb{N}$ *such that* $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \leqslant \mathfrak{v}_1^2$ *and* $m^\star_n \leqslant \exp(\frac{n^{1/2}}{100})$ *we have*

$$\mathcal{R}^{\mathfrak{v}}_n[\,\widehat{\mathbb{p}}^{\widehat{m}}_\bullet \,|\, \mathbb{D}^{\mathfrak{a},\mathrm{r}}_2\,] \leqslant (3\mathrm{r}^2 + 96\tau^2_{\mathfrak{v},\mathrm{u}}) \min\big\{\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\big\} + \mathrm{C}\tau^2_{\mathfrak{v},\mathrm{u}}\big(\mathrm{C}_{\mathfrak{v}}(x_{\mathfrak{a}}) + \mathfrak{v}_1^2\big)\, n^{-1}$$
$$\leqslant \mathrm{K}\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \quad (18.14)$$

and, hence up to the constant $\mathrm{K} := 3\mathrm{r}^2 + 96\tau^2_{\mathfrak{v},\mathrm{u}} + \mathrm{C}\tau^2_{\mathfrak{v},\mathrm{u}}\big(\mathrm{C}_{\mathfrak{v}}(\xi)\mathfrak{v}_1^{-2} + 1\big)$ *the feasible data-driven estimator* $\widehat{\mathbb{p}}^{\widehat{m}}_\bullet$ *is minimax optimal.*

§18.17 **Proof** of Corollary §18.16. is given in the lecture. □

§18.18 **Illustration**. Consider the *trigonometric basis* as in Illustration §16.18 which satisfies Assumption §18.11 **(os1')**, **(os2')** for all $\mathfrak{a}_\bullet \in \ell_2$ and **(os2")**. In Table 02 [§12] (Illustration §12.26) the order of the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is depict for the two specifications **(o)** and **(s)**. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}^+_{\setminus 0}$. The sequence $\mathfrak{v}_\bullet$ satisfies Assumption §18.11, i.e. (18.05), for $v \geqslant -1/2$. Moreover, the optimal dimension $m^\star_n$ given in Table 02 [§12] satisfies $m^\star_n \leqslant \exp(\frac{n^{1/2}}{100})$, and thus (under the above restrictions) the adaptive density estimator attains the minimax optimal rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ up to the constant given in Corollary §18.16. □

§18|02 Data-driven local estimation by Goldenshluger and Lepskij's method

The next assertion provides our key argument in order to control the deviations of the reminder term. The Bernstein inequality in the formulation (18.15) Exrcise is for example given in Comte [2015], Appendix B, Lemma B.2.

§18.19 **Lemma** *(Bernstein inequality). Let* $(Z_i)_{i\in[\![n]\!]}$ *be independent random variables with* $\mathbb{P}(Z_i) = 0$, $\mathbb{P}\big(Z_i^2\big) \leqslant \mathrm{v}^2 \in \mathbb{R}^+$ *and* $|Z_i| \leqslant 2\mathrm{b} \in \mathbb{R}^+$ *for all* $i \in [\![n]\!]$. *Then for any* $x \in \mathbb{R}^+$ *we have*

$$\mathbb{P}\big(\tfrac{1}{n}\sum_{i\in[\![n]\!]} Z_i \geqslant x\big) \leqslant \max\Big\{\exp\big(-\frac{nx^2}{4\mathrm{v}^2}\big), \exp\big(-\frac{nx}{4\mathrm{b}}\big)\Big\} \text{ and}$$
$$\mathbb{P}\big(\big|\tfrac{1}{\sqrt{n}}\sum_{i\in[\![n]\!]} Z_i\big| \geqslant x\big) \leqslant 2\max\Big\{\exp\big(-\frac{x^2}{4\mathrm{v}^2}\big), \exp\big(-\frac{n^{1/2}x}{4\mathrm{b}}\big)\Big\}. \quad (18.15)$$

Moreover, for any $\mathrm{K} \in [1, \infty)$ *we have*

$$\mathbb{P}\Big(\Big(\big|\tfrac{1}{\sqrt{n}}\sum_{i\in[\![n]\!]} Z_i\big|^2 - (4\mathrm{v}^2 + 32\mathrm{b}^2(\log \mathrm{K})n^{-1})\log \mathrm{K}\Big)_+\Big) \leqslant 8\mathrm{K}^{-1}\{\mathrm{v}^2 + 16\mathrm{b}^2 n^{-1}\}. \quad (18.16)$$

§18.20 **Proof** of Lemma §18.19. Exercise □

§18.21 **Remark**. Let us briefly reconsider the OPE $\widehat{\mathbb{p}}^m_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^m_\bullet \in \ell_2\mathbb{1}^m_\bullet$ with dimension $m \in \mathbb{N}$ (Definition §16.04) where $\widehat{\mathbb{p}}_\bullet = \widehat{\mathbb{P}}_n\mathrm{u}_\bullet = (\widehat{\mathbb{P}}_n\mathrm{u}_j)_{j\in\mathbb{N}}$ are noisy versions (Definition §15.08) of the density coefficients $\mathbb{p}_\bullet = \mathrm{U}\,\mathbb{p} = \mathbb{P}_{\mathbb{p}}\mathrm{u}_\bullet = (\mathbb{P}_{\mathbb{p}}\mathrm{u}_j = \lambda_{[0,1]}(\mathbb{p}\,\mathrm{u}_j))_{j\in\mathbb{N}}$. Clearly, $\phi\nu_{\mathbb{N}}((\mathrm{u}_\bullet - \mathbb{p}_\bullet)\mathbb{1}^m_\bullet)$ is a $\mathscr{B}_{[0,1]}$-$\mathscr{B}$-measurable function. Therefore, given $(X_i)_{i\in[\![n]\!]} \sim \mathbb{P}^{\otimes n}_{\mathbb{p}}$ for $i \in [\![n]\!]$ setting $Z_i := \phi\nu_{\mathbb{N}}((\mathrm{u}_\bullet(X_i) - \mathbb{p}_\bullet)\mathbb{1}^m_\bullet)$ we have $\mathbb{P}_{\mathbb{p}}(Z_i) = 0$ exploing $\mathbb{p}_\bullet = \mathbb{P}_{\mathbb{p}}\mathrm{u}_\bullet$ and

$$\phi\nu_{\mathbb{N}}((\widehat{\mathbb{p}}_\bullet - \mathbb{p}_\bullet)\mathbb{1}^m_\bullet) = \widehat{\mathbb{P}}_n\big(\phi\nu_{\mathbb{N}}((\mathrm{u}_\bullet - \mathbb{p}_\bullet)\mathbb{1}^m_\bullet)\big) = n^{-1}\sum_{i\in[\![n]\!]} Z_i.$$

Consequently, provided that

$$\mathbb{P}_{\mathfrak{p}}(Z_i^2) = \mathbb{P}_{\mathfrak{p}}\big(|\phi\nu_{\mathbb{N}}((\mathrm{u}_\bullet - \mathfrak{p}_\bullet)\mathbb{1}_\bullet^m)|^2\big) \leqslant \mathrm{v}_{\mathfrak{p},m}^2 \in \mathbb{R}^+,$$
$$\sup\big\{|\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m)|\colon x \in [0,1]\big\} \leqslant \mathrm{b}_m \in \mathbb{R}^+, \text{ and hence } |Z_i| \leqslant 2\mathrm{b}_m,\ \forall i \in [\![n]\!], \tag{18.17}$$

due to the Bernstein inequality (Lemma §18.19 (18.16)) we have

$$\mathbb{P}\Big(\Big(\big|n^{1/2}\phi\nu_{\mathbb{N}}((\widehat{\mathfrak{p}}_\bullet - \mathfrak{p}_\bullet)\mathbb{1}_\bullet^m)\big|^2 - (4\mathrm{v}_{\mathfrak{p},m}^2 + 32\mathrm{b}_m^2(\log \mathrm{K})n^{-1})\log \mathrm{K}\Big)_+\Big)$$
$$\leqslant 8\mathrm{K}^{-1}\{\mathrm{v}_{\mathfrak{p},m}^2 + 16\mathrm{b}_m^2 n^{-1}\}. \tag{18.18}$$

for any $\mathrm{K} \in [1,\infty)$. □

§18|02|01 Local ϕ-risk

§18.22 **Assumption**. Let $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and the orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet = (\mathrm{u}_j)_{j\in\mathbb{N}}$ on $([0,1],\mathscr{B}_{[0,1]})$ for all $m \in \mathbb{N}$ satisfies **(os2")** $\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\ell_2}^2 \colon x \in [0,1]\big\} \leqslant \tau_{\mathrm{u}}^2 m \in \mathbb{R}^+$ for $\tau_{\mathrm{u}} \in [1,\infty)$. □

§18.23 **Remark**. We replace Assumption §15.05 **(os1)** and **(os2)**, respectively, by the stronger Assumption §18.22 **(os1')** and **(os2")**. Indeed, under **(os1')** we have **(os1)** $\mathbb{1}_{[0,1]} \in \ker(\mathrm{U})$. Furthermore, $(\mathrm{u}_j)_{j\in\mathbb{N}}$ belongs to $\mathbb{L}_\infty(\lambda_{[0,1]})$ due to **(os2")**, and hence **(os2)** is fulfilled (see also Remark §15.06). We use in the sequel that under Assumption §18.22 **(os2")** for each $m \in \mathbb{N}$

$$\sup\big\{|\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet(x))\mathbb{1}_\bullet^m)|^2 \colon x \in [0,1]\big\}$$
$$\leqslant \|\mathbb{1}_\bullet^m\|_\phi^2 \sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\ell_2}^2 \colon x \in [0,1]\big\} \leqslant \tau_{\mathrm{u}}^2 m\|\mathbb{1}_\bullet^m\|_\phi^2 =: \mathrm{b}_m^2 \tag{18.19}$$

by applying the Cauchy Schwarz inequality and moreover (see Proof §15.11)

$$\mathbb{P}_{\mathfrak{p}}\big(|\phi\nu_{\mathbb{N}}((\mathrm{u}_\bullet - \mathfrak{p}_\bullet)\mathbb{1}_\bullet^m)|^2\big) \leqslant \mathbb{P}_{\mathfrak{p}}\big(|\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet\mathbb{1}_\bullet^m)|^2\big) =: \mathrm{v}_{\mathfrak{p},m}^2 \leqslant \|\,\mathfrak{p}\,\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\|\mathbb{1}_\bullet^m\|_\phi^2 \tag{18.20}$$

exploiting Lemma §15.10 (i). Combining (18.19), (18.20) and (18.18) (Remark §18.21) we obtain

$$\mathbb{P}\Big(\Big(\big|n^{1/2}\phi\nu_{\mathbb{N}}((\widehat{\mathfrak{p}}_\bullet - \mathfrak{p}_\bullet)\mathbb{1}_\bullet^m)\big|^2 - (4\mathrm{v}_{\mathfrak{p},m}^2 + 32\mathrm{b}_m^2(\log \mathrm{K})n^{-1})\log \mathrm{K}\Big)_+\Big)$$
$$\leqslant 8\mathrm{K}^{-1}\{\|\,\mathfrak{p}\,\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 16\tau_{\mathrm{u}}^2 m n^{-1}\}\|\mathbb{1}_\bullet^m\|_\phi^2 \tag{18.21}$$

for any $m \in \mathbb{N}$ and $\mathrm{K} \in [1,\infty)$. □

§18.24 **Reminder** *(Local oracle ϕ-risk)*. Given Assumptions §15.02 and §18.22 we consider an OPE as in Definition §16.04. Here the observable noisy density coefficients $\widehat{\mathfrak{p}}_\bullet = \mathfrak{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficients $\mathfrak{p}_\bullet = \mathrm{U}\,\mathfrak{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08. Under Assumptions §15.02 and §18.22, (and hence Assumption §15.05 and $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ see Remark §18.23) and $\mathfrak{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ in §16.22 an *oracle inequality* for the local ϕ-risk of the OPE's is shown. More presicely, as in (16.06) (Proposition §16.20) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}_n^m(\mathfrak{p}_\bullet,\phi_\bullet) := |\phi\nu_{\mathbb{N}}(\mathfrak{p}_\bullet\mathbb{1}_\bullet^{m|\perp})|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{p}_\bullet,\phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(\mathfrak{p}_\bullet,\phi_\bullet) := \mathrm{R}_n^{m_n^\circ}(\mathfrak{p}_\bullet,\phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{p}_\bullet,\phi_\bullet) : m \in \mathbb{N}\big\}. \tag{18.22}$$

and assuming $\mathbb{v}_{\mathbb{p}} := \max(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}, \|\mathbb{p}^{-1}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$, and hence $\max(\|\Gamma_{\mathbb{p}}\|_{\mathbb{L}(\ell_2)}, \|\Gamma_{\mathbb{p}}^{-1}\|_{\mathbb{L}(\ell_2)}) \leqslant \mathbb{v}_{\mathbb{p}}$ (see Lemma §15.10), due to Property §16.22 the (infeasible) OPE $\widehat{\mathbb{p}}^{m^\circ_n}_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^{m^\circ_n}_\bullet \in \ell_2 \mathbb{1}^{m^\circ_n}_\bullet \subseteq \mathrm{dom}(\phi\nu)$ with oracle dimension m°_n as in (18.22) satisfies

$$\mathbb{v}_{\mathbb{p}}^{-1} \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet)|^2) \leqslant \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^{m^\circ_n}_\bullet - \mathbb{p}_\bullet)|^2)$$
$$\leqslant \mathbb{v}_{\mathbb{p}} \mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \mathbb{v}_{\mathbb{p}}^2 \inf_{m\in\mathbb{N}} \mathbb{P}_{\mathbb{p}}^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}_\bullet)|^2),$$

and hence it is *oracle optimal* (with constant $\mathbb{v}_{\mathbb{p}}^2$). □

Partially known penalty sequence

§18.25 **Notation.** Consider first a sequence of penalties $\mathfrak{pen}^{\mathbb{p},\phi}_\bullet = (\mathfrak{pen}^{\mathbb{p},\phi}_m)_{m\in\mathbb{N}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given by

$$\mathfrak{pen}^{\mathbb{p},\phi}_m := 12n^{-1}\big(\mathrm{v}^2_{\mathbb{p},m} + 8\mathrm{b}^2_m(\log \mathrm{K}_m)n^{-1}\big)(\log \mathrm{K}_m) \quad \text{with} \quad \mathrm{v}^2_{\mathbb{p},m} := \mathbb{P}_{\mathbb{p}}\big(|\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet \mathbb{1}^m_\bullet)|^2\big),$$
$$\mathrm{b}^2_m := \tau_{\mathrm{u}}^2 m \|\mathbb{1}^m_\bullet\|^2_\phi, \quad \text{and} \quad \mathrm{K}_m := (1 \vee \|\mathbb{1}^m_\bullet\|^2_\phi)\, m^3 \in [1,\infty) \quad \text{for each } m \in \mathbb{N}, \tag{18.23}$$

which is obviously only *partially known* in advance, and arbitrary but fixed upper bound $\mathrm{M} \in \mathbb{N}$. Considering the data-driven OSE $\widehat{\mathbb{p}}^{\widehat{m}}_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^{\widehat{m}}_\bullet$ with dimension parameter selected by Goldenshluger and Lepskij's method

$$\widehat{m} := \arg\min\big\{\mathfrak{contr}^{\mathbb{p},\phi}_m + \mathfrak{pen}^{\mathbb{p},\phi}_m \colon m \in [\![\mathrm{M}]\!]\big\} \quad \text{and}$$
$$\mathfrak{contr}^{\mathbb{p},\phi}_m := \max\big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^j_\bullet - \widehat{\mathbb{p}}^m_\bullet)|^2 - \mathfrak{pen}^{\mathbb{p},\phi}_j - \mathfrak{pen}^{\mathbb{p},\phi}_m\big)_+ \colon j \in]\!]m,\mathrm{M}]\!]\big\}, \ m \in [\![\mathrm{M}]\!]. \tag{18.24}$$

Moreover, studying a ϕ-error the bias term introduced in (14.31) becomes

$$\mathfrak{bias}_m(\mathbb{p}_\bullet, \phi_\bullet) = \sup\big\{|\phi\nu_{\mathbb{N}}(\mathbb{p}^j_\bullet - \mathbb{p}^m_\bullet)| = |\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet \mathbb{1}^{]\!]m,j]\!]}_\bullet)| \colon j \in [\![m,\infty]\!]\big\} \quad \forall m \in \mathbb{N}.$$

If $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ and hence $\nu_{\mathbb{N}}(|\phi_\bullet \mathbb{p}_\bullet|) \in \mathbb{R}$ then $\mathfrak{bias}_m(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \nu_{\mathbb{N}}(|\phi_\bullet \mathbb{p}_\bullet| \mathbb{1}^{m|\perp}_\bullet) = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence. Considering the data-driven OSE $\widehat{\mathbb{p}}^{\widehat{m}}_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^{\widehat{m}}_\bullet$ with dimension parameter $\widehat{m}$ selected as in (18.24) with penalty sequence $\mathfrak{pen}^{\mathbb{p},\phi}_\bullet$ given in (18.23) and arbitrary upper bound $\mathrm{M} \in \mathbb{N}$ we derive below an upper bound for its local ϕ-risk, $\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^{\widehat{m}}_\bullet - \mathbb{p}_\bullet)|^2\big)$. □

§18.26 **Lemma.** *Under Assumptions §15.02 and §18.22 and* $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\mathfrak{pen}^{\mathbb{p},\mathbb{v}}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ *as in* (18.23) *and for any* $\mathrm{M} \in \mathbb{N}$ *we have*

$$\mathbb{P}_{\mathbb{p}}^{\otimes n}\Big(\max\big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^m_\bullet - \mathbb{p}^m_\bullet)|^2 - \mathfrak{pen}^{\mathbb{p},\phi}_m/3\big)_+ \colon m \in [\![\mathrm{M}]\!]\big\}\Big) \leqslant 14\{\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 16\tau_{\mathrm{u}}^2 n^{-1}\}n^{-1}. \tag{18.25}$$

§18.27 **Proof** of Lemma §18.26. is given in the lecture. □

§18.28 **Proposition** *(Upper bound).* *Under Assumptions §15.02 and §18.22 and* $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\mathfrak{pen}^{\mathbb{p},\mathbb{v}}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ *as in* (18.23) *the data-driven OPE* $\widehat{\mathbb{p}}^{\widehat{m}}_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^{\widehat{m}}_\bullet \in \ell_2 \mathbb{1}^{\widehat{m}}_\bullet \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with data-driven dimension* $\widehat{m} \in [\![\mathrm{M}]\!]$ *as in* (18.24) *satisfies for all* $n, \mathrm{M} \in \mathbb{N}$

$$\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^{\widehat{m}}_\bullet - \mathbb{p}_\bullet)|^2\big) \leqslant 64(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 8\tau_{\mathrm{u}}^2)$$
$$\times \min\big\{\mathfrak{bias}^2_m(\mathbb{p}_\bullet, \phi_\bullet) + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m)mn^{-1}) \colon m \in [\![\mathrm{M}]\!]\big\}$$
$$+ 392(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 16\tau_{\mathrm{u}}^2 n^{-1})n^{-1}. \tag{18.26}$$

§18.29 **Proof** of Proposition §18.28. is given in the lecture. □

§18.30 **Comment.** Let us compare the dominating part of the upper bound given in (18.26), that is

$$\min\big\{\mathfrak{bias}_m^2(\mathbb{p}_\bullet,\phi)+n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2(\log \mathrm{K}_m)(1\vee(\log \mathrm{K}_m)mn^{-1})\colon m\in[\![\mathrm{M}]\!]\big\} \tag{18.27}$$

with the oracle bound $\mathrm{R}_n^\circ(\theta_\bullet,\phi)=\min\big\{|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^m-\mathbb{p}_\bullet)|^2+n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2\colon m\in\mathbb{N}\big\}$ (for details see Reminder §18.24). In (18.27) we face eventually a deterioration by three sources. First, we generally have $\mathfrak{bias}_m(\mathbb{p}_\bullet,\phi_\bullet)\geqslant|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^m-\mathbb{p}_\bullet)|$, but note that for $\mathbb{p}_\bullet\phi_\bullet\in(\mathbb{R}^+)^{\mathbb{N}}$ equality holds, that is

$$\mathfrak{bias}_m(\mathbb{p}_\bullet,\phi_\bullet)=\sup\big\{\nu_{\mathbb{N}}(\phi_\bullet\mathbb{p}_\bullet\mathbb{1}_\bullet^{]\!]m,j]\!]})\colon j\in[\![m,\infty]\!]\big\}=\nu_{\mathbb{N}}(\phi_\bullet\mathbb{p}_\bullet\mathbb{1}_\bullet^{m|\perp})=|\phi\nu_{\mathbb{N}}(\mathbb{p}_\bullet^m-\mathbb{p}_\bullet)|$$

for all $m\in\mathbb{N}$. Secondly, the variance term features an additional factor $(\log \mathrm{K}_m)(1\vee(\log \mathrm{K}_m)mn^{-1})$, and finally the upper bound M might impose an additional deterioration. We note that the oracle bound $\mathrm{R}_n^\circ(\mathbb{p}_\bullet,\phi_\bullet)$ is parametric, i.e. $n\mathrm{R}_n^\circ(\mathbb{p}_\bullet,\phi_\bullet)=\mathrm{O}(1)$ as $n\to\infty$, if $\phi_\bullet\in\ell_2$ (case **(p)** in Illustration §12.40). In the sequel we consider only the case $\phi_\bullet\notin\ell_2$, i.e. $\nu_{\mathbb{N}}(|\phi_\bullet|^2)=\infty$. We set

$$\mathrm{M}^\phi:=\max\big\{m\in\mathbb{N}\colon\|\mathbb{1}_\bullet^m\|_\phi^2\leqslant n\phi_1^2\big\}\in\mathbb{N} \tag{18.28}$$

where the defining set is not empty and finite since $\|\phi_\bullet\|_{\ell_2}^2=\infty$. The next assertion shows that this is a suitable choice for the upper bound. Moreover, we estimate the bias term by $\mathfrak{bias}_m(\mathbb{p}_\bullet,\phi_\bullet)\leqslant\nu(|\phi_\bullet\mathbb{p}_\bullet|\mathbb{1}_\bullet^{m|\perp})$ where equality holds whenever $\mathbb{p}_\bullet\phi_\bullet\in(\mathbb{R}^+)^{\mathbb{N}}$. □

§18.31 **Corollary.** *Given $\phi_\bullet\in(\mathbb{R}_{\setminus0})^{\mathbb{N}}$ with $\phi_\bullet\notin\ell_2$, $\mathrm{M}^\phi\in\mathbb{N}$ as in (18.28) and $\mathfrak{pen}_\bullet^{\mathbb{p},\phi}$ as in (18.23) consider a data-driven OPE $\widehat{\mathbb{p}}_\bullet^{\widehat{m}}=\widehat{\mathbb{p}}_\bullet\mathbb{1}_\bullet^{\widehat{m}}\in\ell_2\mathbb{1}_\bullet^{\widehat{m}}\subseteq\mathrm{dom}(\phi\nu_{\mathbb{N}})$ of $\mathbb{p}_\bullet\in\mathrm{dom}(\phi\nu_{\mathbb{N}})$ with*

$$\begin{aligned}\widehat{m}&:=\arg\min\big\{\mathfrak{contr}_m^\phi+\mathfrak{pen}_m^{\mathbb{p},\phi}\colon m\in[\![\mathrm{M}^\phi]\!]\big\}\quad\textit{and}\\ \mathfrak{contr}_m^{\mathbb{p},\phi}&:=\max\big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^j-\widehat{\theta}_\bullet^m)|^2-\mathfrak{pen}_j^{\mathbb{p},\phi}-\mathfrak{pen}_m^{\mathbb{p},\phi}\big)_+\colon j\in]\!]m,\mathrm{M}^\phi]\!]\big\},\ m\in[\![\mathrm{M}^\phi]\!].\end{aligned} \tag{18.29}$$

For $n,m\in\mathbb{N}$ we set

$$\begin{aligned}\mathrm{R}_n^m(\mathbb{p}_\bullet,\phi_\bullet)&:=\big(\nu_{\mathbb{N}}(|\phi_\bullet\mathbb{p}_\bullet|\mathbb{1}_\bullet^{m|\perp})\big)^2\\&\quad+(1+\big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+\big)+\log m\big)\big(1+\big(\big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+\big)+\log m\big)mn^{-1}\big)\,n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2,\\ m^\diamond&:=\arg\min\big\{\mathrm{R}_n^m(\mathbb{p}_\bullet,\phi_\bullet)\colon m\in\mathbb{N}\big\}\quad\textit{and}\\ \mathrm{R}_n^\diamond(\mathbb{p}_\bullet,\phi_\bullet)&:=\mathrm{R}_n^{m^\diamond}(\mathbb{p}_\bullet,\phi_\bullet)=\min\big\{\mathrm{R}_n^m(\mathbb{p}_\bullet,\phi_\bullet)\colon m\in\mathbb{N}\big\}.\end{aligned} \tag{18.30}$$

Under the assumptions of Proposition §18.28 for each $n\in\mathbb{N}$ such that $\mathrm{R}_n^\diamond(\mathbb{p}_\bullet,\phi_\bullet)\leqslant\phi_1^2$ we have

$$\begin{aligned}\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^{\widehat{m}}-\mathbb{p}_\bullet)|^2\big)&\leqslant576(\|\,\mathbb{p}\,\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}+8\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathbb{p}_\bullet,\phi_\bullet)+392(\|\,\mathbb{p}\,\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}+16\tau_{\mathrm{u}}^2n^{-1})n^{-1}\\&\leqslant(576+784\phi_1^{-2})(\|\,\mathbb{p}\,\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}+8\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathbb{p}_\bullet,\phi_\bullet).\end{aligned} \tag{18.31}$$

§18.32 **Proof** of Corollary §18.31. is given in the lecture. □

§18.33 **Comment.** The data-driven bound $\mathrm{R}_n^\diamond(\mathbb{p}_\bullet,\phi_\bullet)$ compared to the oracle bound $\mathrm{R}_n^\circ(\mathbb{p}_\bullet,\phi_\bullet)$ features a deterioration of the variance term at least by a logarithmic factor. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§18.34 **Illustration**. We illustrate the last results considering the two specifications **(o)** and **(s)** given in Table 03 [§12] (Illustration §12.40). We restrict ourselves to the case $\phi_\bullet \notin \ell_2$ only.

Table 01 [§18]

Order of the oracle rate $\mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}^\diamond_n(\mathbb{p}_\bullet, \phi_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$ $\phi_\bullet = j^{v-1/2}$	$(a \in \mathbb{R}^+_{\setminus 0})$ $\mathbb{p}$	(squarred bias) $(\nu_{\mathbb{N}}(\lvert\phi_\bullet \mathbb{p} \lvert \mathbb{1}^{m\lvert\perp}_\bullet))^2$	(variance) $\lVert \mathbb{1}^m_\bullet \rVert^2_\phi$	M^ϕ	$m^\diamond$	$\mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet)$	$\mathrm{R}^\diamond_n(\mathbb{p}_\bullet, \phi_\bullet)$
(o) $v \in (0,a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2v}}$		$n^{-\frac{(a-v)}{a}}$	
$a \in (1/2, \infty)$					$\left(\frac{n}{\log n}\right)^{\frac{1}{2a}}$		$\left(\frac{\log n}{n}\right)^{\frac{(a-v)}{a}}$
$a \in (0, 1/2]$					$\left(\frac{n}{\log n}\right)^{\frac{1}{a+1/2}}$		$\left(\frac{\log n}{n}\right)^{\frac{2(a-v)}{a+1/2}}$
$v = 0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	e^n		$\frac{\log n}{n}$	
$a \in (1/2, \infty)$					$\left(\frac{n}{(\log n)^2}\right)^{\frac{1}{2a}}$		$\frac{(\log n)^2}{n}$
$a \in (0, 1/2]$					$\left(\frac{n^2}{(\log n)^3}\right)^{\frac{1}{2a+1}}$		$\left(\frac{(\log n)^3}{n^2}\right)^{\frac{a}{a+1/2}}$
(s) $v \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+} e^{-2m^{2a}}$	m^{2v}	$n^{\frac{1}{2v}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}} (\log\log n)}{n}$
$v = 0$	$e^{-j^{2a}}$	$m^{(1-2a)_+} e^{-2m^{2a}}$	$\log m$	e^n	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 01 [§18] the order of the oracle rate $\mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet)$ and the data-driven rate $\mathrm{R}^\diamond_n(\mathbb{p}_\bullet, \phi_\bullet)$ is depict for $v \geqslant 0$ only. In case $v < 0$ we have $\phi_\bullet \in \ell_2$ and thus Corollary §18.31 is not applicable. Moreover, in case **(s)** for $a \in \mathbb{R}^+_{\setminus 0}$ and **(o)** for $a \in (1/2, \infty)$ the rate $\mathrm{R}^\diamond_n(\mathbb{p}_\bullet, \phi_\bullet)$ features only an additional logarithmic factor compared with the oracle rate $\mathrm{R}^\circ_n(\mathbb{p}_\bullet, \phi_\bullet)$. □

Estimated penalty sequence

§18.35 **Notation**. The penalty sequence $\mathfrak{p}\mathrm{en}^{\mathbb{p},\mathfrak{v}}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given in (18.23) still depends on characteristics of the unknown density $\mathbb{p}$. More precisely, for $m \in \mathbb{N}$ the term $\mathfrak{p}\mathrm{en}^{\mathbb{p},\mathfrak{v}}_m$ involves the quantity $\mathrm{v}^2_{\mathbb{p},m} = \mathbb{P}_{\mathbb{p}}\big(\lvert\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet \mathbb{1}^m_\bullet)\rvert^2\big)$ which we eventually estimate without bias by $\widehat{\mathrm{v}}^2_m := \widehat{\mathbb{P}}_n\big(\lvert\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet \mathbb{1}^m_\bullet)\rvert^2\big)$. Based on this estimator let us introduce a fully data-driven sequence of penalties $\widehat{\mathfrak{p}}\mathrm{en}^\phi_\bullet = (\widehat{\mathfrak{p}}\mathrm{en}^\phi_m)_{m\in\mathbb{N}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given by

$$\widehat{\mathfrak{p}}\mathrm{en}^\phi_m := 12n^{-1}\big(2\widehat{\mathrm{v}}^2_m + 3{\times}8\mathrm{b}^2_m(\log \mathrm{K}_m)n^{-1}\big)(\log \mathrm{K}_m) \quad \text{with} \quad \widehat{\mathrm{v}}^2_m := \widehat{\mathbb{P}}_n\big(\lvert\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet \mathbb{1}^m_\bullet)\rvert^2\big),$$
$$\mathrm{b}^2_m := \tau^2_{\mathrm{u}} m \lVert \mathbb{1}^m_\bullet \rVert^2_\phi, \quad \text{and} \quad \mathrm{K}_m := (1 \vee \lVert \mathbb{1}^m_\bullet \rVert^2_\phi)\, m^3 \in [1, \infty) \quad \text{for each } m \in \mathbb{N}, \tag{18.32}$$

which is now *fully known* in advance, and arbitrary but fixed upper bound $\mathrm{M} \in \mathbb{N}$. Considering the data-driven OSE $\widehat{\mathbb{p}}^{\widehat{m}}_\bullet = \widehat{\mathbb{p}}_\bullet \mathbb{1}^{\widehat{m}}_\bullet$ with dimension parameter selected by Goldenshluger and Lepskij's method

$$\widehat{m} := \arg\min\big\{\widehat{\mathfrak{c}}\mathrm{ontr}^\phi_m + \widehat{\mathfrak{p}}\mathrm{en}^\phi_m : m \in [\![\mathrm{M}]\!]\big\} \quad \text{and}$$
$$\widehat{\mathfrak{c}}\mathrm{ontr}^\phi_m := \max\Big\{\big(\lvert\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^j_\bullet - \widehat{\mathbb{p}}^m_\bullet)\rvert^2 - \widehat{\mathfrak{p}}\mathrm{en}^\phi_j - \widehat{\mathfrak{p}}\mathrm{en}^\phi_m\big)_+ : j \in]\!]m, \mathrm{M}]\!]\Big\}, \ m \in [\![\mathrm{M}]\!] \tag{18.33}$$

we derive below an upper bound for its local ϕ-risk, $\mathbb{P}^{\otimes n}_{\mathbb{p}}\big(\lvert\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^{\widehat{m}}_\bullet - \mathbb{p}_\bullet)\rvert^2\big)$. □

§18.36 **Lemma**. *Under Assumptions §15.02 and §18.22 and* $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\mathfrak{p}\mathrm{en}^{\mathbb{p},\phi}_\bullet, \widehat{\mathfrak{p}}\mathrm{en}^\phi_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ *as in* (18.23) *and* (18.32)*, respectively, and for any* $\mathrm{M} \in \mathbb{N}$ *we have*

$$\mathbb{P}^{\otimes n}_{\mathbb{p}}\Big(\max\Big\{\big(\mathfrak{p}\mathrm{en}^{\mathbb{p},\phi}_j - \widehat{\mathfrak{p}}\mathrm{en}^\phi_j\big)_+ : j \in [\![\mathrm{M}]\!]\Big\}\Big) \leqslant 40\{\lVert \mathbb{p} \rVert_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 6\tau^2_{\mathrm{u}} n^{-1}\}n^{-1}. \tag{18.34}$$

§18.37 **Proof** of Lemma §18.36. is given in the lecture. □

§18.38 **Proposition** *(Upper bound).* *Under Assumptions §15.02 and §18.22 and* $\mathbb{p} \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\widehat{\mathfrak{p}}\mathrm{en}_\bullet^\phi \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ *as in* (18.32) *the data-driven OPE* $\widehat{\mathbb{p}}_\bullet^{\widehat{m}} = \widehat{\mathbb{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2 \mathbb{1}_\bullet^{\widehat{m}} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with data-driven dimension* $\widehat{m} \in [\![\mathrm{M}]\!]$ *as in* (18.33) *satisfies for all* $n, \mathrm{M} \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^{\widehat{m}} - \mathbb{p}_\bullet)|^2\big) &\leqslant 112(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 12\tau_{\mathrm{u}}^2) \\ &\quad \times \min\big\{\mathfrak{bias}_m^2(\mathbb{p}_\bullet, \phi_\bullet) + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m)m n^{-1}) \colon m \in [\![\mathrm{M}]\!]\big\} \\ &\quad + 1440(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 16\tau_{\mathrm{u}}^2 n^{-1})n^{-1}. \quad (18.35)\end{aligned}$$

§18.39 **Proof** of Proposition §18.38. is given in the lecture. □

§18.40 **Comment.** We shall stress that the last upper bound (18.35) in Proposition §18.38 (for the fully data-driven procedure) and the upper bound (18.26) in Proposition §18.28 (for the partially data-driven procedure) differ only in the numerical constants. Thus, thus the proof of the next results follows line by line their counterparts above. □

§18.41 **Corollary.** *Given* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *with* $\phi_\bullet \notin \ell_2$, $\mathrm{M}^\phi \in \mathbb{N}$ *as in* (18.28) *and* $\widehat{\mathfrak{p}}\mathrm{en}_\bullet^\phi$ *as in* (18.32) *consider a data-driven OPE* $\widehat{\mathbb{p}}_\bullet^{\widehat{m}} = \widehat{\mathbb{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2 \mathbb{1}_\bullet^{\widehat{m}} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $\mathbb{p}_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with*

$$\begin{aligned}&\widehat{m} := \arg\min\big\{\widehat{\mathfrak{c}}\mathrm{ontr}_m^\phi + \widehat{\mathfrak{p}}\mathrm{en}_m^\phi \colon m \in [\![\mathrm{M}^\phi]\!]\big\} \quad \textit{and} \\ &\widehat{\mathfrak{c}}\mathrm{ontr}_m^\phi := \max\big\{\big(|\phi_\bullet\nu_{\mathbb{N}}(\widehat{\theta}_\bullet^j - \widehat{\theta}_\bullet^m)|^2 - \widehat{\mathfrak{p}}\mathrm{en}_j^\phi - \widehat{\mathfrak{p}}\mathrm{en}_m^\phi\big)_+ \colon j \in]\!]m, \mathrm{M}^\phi]\!]\big\},\ m \in [\![\mathrm{M}^\phi]\!]. \quad (18.36)\end{aligned}$$

For $n, m \in \mathbb{N}$ *let* $m^\diamond$ *and* $\mathrm{R}_n^\diamond(\mathbb{p}_\bullet, \phi_\bullet)$ *defined as in* (18.30). *Under the assumptions of Proposition §18.38 for each* $n \in \mathbb{N}$ *such that* $\mathrm{R}_n^\diamond(\mathbb{p}_\bullet, \phi_\bullet) \leqslant \phi_1^2$ *we have*

$$\begin{aligned}\mathbb{P}_{\mathbb{p}}^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}_\bullet^{\widehat{m}} - \mathbb{p}_\bullet)|^2\big) &\leqslant 1008(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 8\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathbb{p}_\bullet, \phi_\bullet) + 1440(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 16\tau_{\mathrm{u}}^2 n^{-1})n^{-1} \\ &\leqslant (1008 + 1920\phi_1^{-2})(\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 12\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathbb{p}_\bullet, \phi_\bullet). \quad (18.37)\end{aligned}$$

§18.42 **Proof** of Proof §18.42. is given in the lecture. □

§18.43 **Comment.** The fullay data-driven bound $\mathrm{R}_n^\diamond(\mathbb{p}_\bullet, \phi_\bullet)$ equals exactly the bound in the partially known case. Therefore, the Comment §18.33 and the Illustration §18.34 apply here equally. □

§18|02|02 Maximal local ϕ-risk

§18.44 **Assumption.** Consider $\phi_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\phi)_\bullet := (\mathfrak{a}_j\phi_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\phi \in \ell_2$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and $\mathrm{u}_0 := \mathbb{1}_{[0,1]}$ form an **(os1')** *orthonormal basis* $(\mathrm{u}_j)_{j\in\mathbb{N}_0}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet^2 = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2')** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2 \mathrm{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \tau_{\mathfrak{a},\mathrm{u}}^2$ and **(os2'')** $\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\ell_2}^2 \colon x \in [0,1]\big\} \leqslant \tau_{\mathrm{u}} m \in \mathbb{R}^+$ for $\tau_{\mathfrak{a},\mathrm{u}}, \tau_{\mathrm{u}} \in [1, \infty)$. □

§18.45 **Remark.** Assumption §18.44 contains Assumption §18.22 and thus Assumption §15.05 **(os1)** and **(os2)** are satisfied (see Remark §18.23). Moreover, considering the set $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ of densities in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (16.04) we have $\|\mathbb{p}\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant 1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}$ for all $\mathbb{p} \in \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ due to **(os2')** which allows us to apply Lemma §16.14. Consequently, given in addition Assumption §15.02 all assumptions of Proposition §18.38 are satisfied. □

§18.46 **Reminder** *(Maximal local ϕ-risk).* Given Assumptions §15.02 and §18.44 we consider an OPE as in Definition §16.04. Here the observable noisy density coefficients $\widehat{\mathbb{p}}_\bullet = \mathbb{p}_\bullet + n^{-1/2}\varepsilon_\bullet$ of the density coefficients $\mathbb{p}_\bullet = \mathrm{U}\mathbb{p} \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in \mathscr{B}_{[0,1]}^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §15.08.

Under Assumptions §15.02 and §18.44 (and hence Assumption §16.24) in Proposition §16.27 an upper bound for a maximal local ϕ-risk of an OPE is shown over the set $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ given in (16.04) (Lemma §16.14). More presicely, as in (12.13) (Proposition §12.42) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (18.38)$$

by Proposition §16.27 under Assumptions §15.02 and §18.44 the maximal local ϕ-risk of an OPE $\widehat{\mathfrak{p}}_\bullet^{m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (18.38) satisfies

$$\mathcal{R}_n^\phi[\,\widehat{\mathfrak{p}}_\bullet^{m_n^\star} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$$

with $\mathrm{C} = (1+\mathrm{r}\tau_{\mathfrak{a},\mathrm{u}})\vee\mathrm{r}^2$. Moreover, due to Proposition §17.08 $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ provides (up to a constant) also a lower bound of the maximal global ϕ-risk over the ellipsoid $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is a minimax bound and $\widehat{\mathfrak{p}}_\bullet^{m_n^\star}$ is minimax optimal. However, the optimal dimension $m_n^\star$ depends on $\mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ characterising the ellipsoid $\mathbb{D}_2^{\mathfrak{a},\mathrm{r}}$. □

§18.47 **Proposition** *(Upper bound). Under Assumptions §15.02 and §18.44 for* $\widehat{\mathfrak{p}\mathrm{en}}_\bullet^\phi \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ *as in* (18.32) *the OPE* $\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2\mathbb{1}_\bullet^{\widehat{m}} \subseteq \mathrm{dom}(\phi\iota_{\mathbb{N}})$ *with fully data-driven dimension* $\widehat{m} \in [\![\mathrm{M}]\!]$ *as in* (18.33) *satisfies for all* $n, \mathrm{M} \in \mathbb{N}$

$$\mathcal{R}_n^\phi[\,\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant 168(\mathrm{r}^2 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + 9\tau_{\mathrm{u}}^2)$$
$$\times \min\big\{\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m)mn^{-1}) : m \in [\![\mathrm{M}]\!]\big\}$$
$$+ 1440(1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + 16\tau_{\mathrm{u}}^2 n^{-1})n^{-1}. \quad (18.39)$$

§18.48 **Proof** of Proposition §18.47. is given in the lecture. □

§18.49 **Corollary**. *Under Assumptions §15.02 and §18.44 and* $\phi_\bullet \notin \ell_2$ *given* $\mathrm{M}^\phi \in \mathbb{N}$ *as in* (18.28) *and* $\widehat{\mathfrak{p}\mathrm{en}}_\bullet^\phi \in (\mathbb{R}_{\setminus 0}^+)^{\mathbb{N}}$ *as in* (18.32) *consider a data-driven OPE* $\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} = \widehat{\mathfrak{p}}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2\mathbb{1}_\bullet^{\widehat{m}} \subseteq \mathrm{dom}(\phi\iota_{\mathbb{N}})$ *with data-driven dimension* $\widehat{m} \in [\![\mathrm{M}_n]\!]$ *as in* (18.36). *For* $n, m \in \mathbb{N}$ *we set*

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2$$
$$+ \big(1 + \big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + \log m\big)\big(1 + \big(\big(\log\|\mathbb{1}_\bullet^m\|_\phi^2\big)_+ + \log m\big)mn^{-1}\big)\, n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2,$$
$$m^\diamond := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\} \quad \textit{and}$$
$$\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m^\diamond}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (18.40)$$

For each $n \in \mathbb{N}$ *such that* $\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet) \leqslant \phi_1^2$ *we have*

$$\mathcal{R}_n^\phi[\,\widehat{\mathfrak{p}}_\bullet^{\widehat{m}} \,|\, \mathbb{D}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant 1512(\mathrm{r}^2 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + 9\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet) + 1440(1 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + 16\tau_{\mathrm{u}}^2 n^{-1})n^{-1}$$
$$\leqslant (1512 + 1440\phi_1^{-2})(\mathrm{r}^2 + \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}} + 17\tau_{\mathrm{u}}^2)\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet). \quad (18.41)$$

§18.50 **Proof** of Corollary §18.49. is given in the lecture. □

§18.51 **Comment**. The data-driven bound $\mathrm{R}_n^\diamond(\mathfrak{a}_\bullet, \phi_\bullet)$ compared to the minimax bound $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ features a deterioration of the variance term at least by a logarithmic factor. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§18.52 **Illustration**. We illustrate the last results considering the two specifications **(o)** and **(o)** given in Table 04 [§12] (Illustration §12.47). We restrict ourselves again to the case $\phi_{\bullet} \notin \ell_2$ only.

Table 02 [§18]

Order of the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ and the data-driven rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ as $n \to \infty$

	$(j \in \mathbb{N})$	$(a \in \mathbb{R}_{\setminus 0}^{+})$	(squarred bias)	(variance)				
	$\phi_j = j^{v-1/2}$	$\mathfrak{a}_j^2$	$\lVert \mathfrak{a}_{\bullet} \mathbb{1}_{\bullet}^{m\lvert\perp} \rVert_{\phi}^2$	$\lVert \mathbb{1}_{\bullet}^{m} \rVert_{\phi}^2$	M^{ϕ}	$m^{\diamond}$	$\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$	$\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$
(o)	$v \in (0, a)$	j^{-a}	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2v}}$		$n^{-\frac{(a-v)}{a}}$	
	$a \in (1/2, \infty)$					$\left(\frac{n}{\log n}\right)^{\frac{1}{2a}}$		$\left(\frac{\log n}{n}\right)^{\frac{(a-v)}{a}}$
	$a \in (0, 1/2]$					$\left(\frac{n}{\log n}\right)^{\frac{1}{a+1/2}}$		$\left(\frac{\log n}{n}\right)^{\frac{2(a-v)}{a+1/2}}$
	$v = 0$	j^{-a}	m^{-2a}	$\log m$	e^{n}		$\frac{\log n}{n}$	
	$a \in (1/2, \infty)$					$\left(\frac{n}{(\log n)^2}\right)^{\frac{1}{2a}}$		$\frac{(\log n)^2}{n}$
	$a \in (0, 1/2]$					$\left(\frac{n^2}{(\log n)^3}\right)^{\frac{1}{2a+1}}$		$\left(\frac{(\log n)^3}{n^2}\right)^{\frac{a}{a+1/2}}$
(s)	$v \in \mathbb{R}_{\setminus 0}^{+}$	$e^{-j^{2a}}$	$m^{2(v-a)_+} e^{-m^{2a}}$	m^{2v}	$n^{\frac{1}{2v}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}} (\log\log n)}{n}$
	$v = 0$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	e^{n}	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 02 [§18] the order of the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ and the data-driven rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ is depict for $v \geqslant 0$ only. In case $v < 0$ we have $\phi_{\bullet} \in \ell_2$ and thus Corollary §18.49 is not applicable. Moreover, in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^{+}$ and **(o)** for $a \in (1/2, \infty)$ the rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ features only an additional logarithmic factor compared with the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$. □

Chapter 5

Nonparametric regression

This chapter presents nonparametric regression with uniform design along the lines of the textbooks by **Tsybakov** *[2009] and* **Comte** *[2015] where far more details, examples and further discussions can be found.*

Overview

§19 Noisy regression coefficients

§19.01 **Notation** *(Reminder).* Consider the measure space $([0,1],\mathscr{B}_{[0,1]},\lambda_{[0,1]})$ where $\lambda_{[0,1]}$ denotes the restriction of the Lebesgue measure to the Borel-σ-algebra $\mathscr{B}_{[0,1]}$ over $[0,1]$, and the Hilbert space $\mathbb{L}_2(\lambda_{[0,1]}) := \mathbb{L}_2([0,1],\mathscr{B}_{[0,1]},\lambda_{[0,1]})$ of square Lebesgue-integrable functions. Let (X,Y) be a $[0,1]\times\mathbb{R}$-valued random vector. We denote by $\mathbb{P}^X \in \mathcal{W}(\mathscr{B}_{[0,1]})$ the marginal distribution of X, by $\mathbb{P}^{Y|X}$ a regular conditional distribution of Y given X, and by $\mathbb{P}^{X,Y} = \mathbb{P}^X \odot \mathbb{P}^{Y|X} \in \mathcal{W}(\mathscr{B}_{[0,1]}\otimes\mathscr{B})$ the joint distribution of (X,Y). We tactically identify X and Y with the coordinate map $\Pi_{[0,1]}$ and $\Pi_{\mathbb{R}}$, respectively, and thus (X,Y) with the identity $\mathrm{id}_{[0,1]\times\mathbb{R}}$ such that $\mathbb{P} = \mathbb{P}^{X,Y} \in \mathcal{W}(\mathscr{B}_{[0,1]}\otimes\mathscr{B})$. If in addition $Y \in \mathcal{L}_1(\mathbb{P}) = \mathcal{L}_1([0,1]\times\mathbb{R},\mathscr{B}_{[0,1]}\otimes\mathscr{B},\mathbb{P})$ then $\mathbb{P}^{Y|X}(\mathrm{id}_{\mathbb{R}}) = \mathbb{P}\big(Y\big|X\big) =: f \in \mathscr{B}_{[0,1]}$ is unique up to $\mathbb{P}^X$-a.s. equality. Moreover, we have $f \in \mathcal{L}_1(\mathbb{P}^X) = \mathcal{L}_1([0,1],\mathscr{B}_{[0,1]},\mathbb{P}^X)$ and the error term $\xi := Y - f(X)$ satisfies $\xi \in \mathcal{L}_1(\mathbb{P})$ with $\mathbb{P}(\xi) = 0$. Let us denote in this situation by $\mathbb{P}_f^{Y|X}$ and $\mathbb{P}_f := \mathbb{P}^X \odot \mathbb{P}_f^{Y|X} \in \mathcal{W}(\mathscr{B}_{[0,1]}\otimes\mathscr{B})$, respectively, a regular conditional distribution of Y given X and the joint distribution of (X,Y). Keep however in mind, that even if $f \in \mathcal{L}_1(\mathbb{P}^X)$ is fixed the conditional distribution $\mathbb{P}_f^{Y|X}$ is still not fully specified. In what follows we assume that the error term ξ has in addition a finite second moment and its distribution does not depend on the regression function, that is $\xi \sim \mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}^+_2}$ where $\mathrm{P}_{\{0\}\times\mathbb{R}^+_2} \subseteq \mathcal{W}(\mathscr{B})$ is the subset of all probability distributions over $(\mathbb{R},\mathscr{B})$ with finite second moment and mean zero. For $a\in\mathbb{R}$ denote by $\mathbb{P}_a^\xi \in \mathcal{W}(\mathscr{B})$ the distribution of $\xi + a$. If ξ and X are *independent*, which is assumed throughtout this chapter, then there exists a $\mathbb{P}^X$-null set $\mathcal{N} \in \mathscr{B}_{[0,1]}$ such that $\mathbb{P}_f^{Y|X=x}(B) = \mathbb{P}_{f(x)}^\xi(B)$ for all $B\in\mathscr{B}$ and $x \in \mathcal{N}^c$ *(Witting [1985], Satz 129, p.130)*. In other words, $(x,B)\mapsto \mathbb{P}_{f(x)}^\xi(B)$ is a version of the conditional distributions of Y given X. Evidently, if for each $B\in\mathscr{B}$ the map $\mathbb{P}_\bullet^\xi(B) : \mathbb{R}\to[0,1]$ with $a\mapsto\mathbb{P}_a^\xi(B)$ is Borel-measurable, $\mathbb{P}_\bullet^\xi(B)\in\mathscr{B}$ for short, then $\mathbb{P}_\bullet^\xi : \mathbb{R}\times\mathscr{B}\to[0,1]$ with $(a,B)\mapsto\mathbb{P}_a^\xi(B)$ is a Markov kernel from $(\mathbb{R},\mathscr{B})$ to $(\mathbb{R},\mathscr{B})$. In this situation, for any $f\in\mathscr{B}_{[0,1]}$

the map $\mathbb{P}^{\xi}_{f(X)} : [0,1] \times \mathscr{B} \to [0,1]$ with $(x, B) \mapsto \mathbb{P}^{\xi}_{f(x)}(B)$ is a Markov kernel from $([0,1], \mathscr{B}_{[0,1]})$ to $(\mathbb{R}, \mathscr{B})$, and hence it is a regular version of the conditional distribution of Y given X, in symbols $\mathbb{P}^{Y|X}_f = \mathbb{P}^{\xi}_{f(X)}$. Consequently, we have $\mathbb{P}_f = \mathbb{P}^X \odot \mathbb{P}^{\xi}_{f(X)} \in \mathcal{W}(\mathscr{B}_{[0,1]} \otimes \mathscr{B})$. We assume in what follows that $f \in \mathbb{F}_2 \subseteq \mathbb{L}_2(\mathbb{P}^X)$ identifying again equivalence classes and their representatives. □

§19.02 **Assumption**. The $[0,1] \times \mathbb{R}$-valued random vector $(X, Y) \sim \mathbb{P}_f = \mathbb{P}^X \odot \mathbb{P}^{Y|X}_f \in \mathcal{W}(\mathscr{B}_{[0,1]} \otimes \mathscr{B})$ satisfies $Y \in \mathcal{L}_1(\mathbb{P}_f)$ and $\mathbb{P}_f(Y \,|\, X) = f$ $\mathbb{P}^X$-a.s. with regression function $f \in \mathbb{F}_2 \subseteq \mathbb{L}_2(\lambda_{[0,1]})$.

(NR1) The error term $\xi = Y - f(X) \sim \mathbb{P}^{\xi} \in \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}$ has a finite second moment, mean zero and its distribution does not depend on the regression function f. We set $\sigma^2_{\xi} := \mathbb{P}^{\xi}(\mathrm{id}^2_{\mathbb{R}}) = \mathbb{P}_f(\xi^2)$.

(NR2) The error term ξ and the explanatory variable X are *independent*.

(NR3) The map $\mathbb{P}^{\xi}_{\bullet} : \mathbb{R} \times \mathscr{B} \to [0,1]$ with $(a, B) \mapsto \mathbb{P}^{\xi}_a(B)$ is a *Markov kernel* from $(\mathbb{R}, \mathscr{B})$ to $(\mathbb{R}, \mathscr{B})$. Consequently, under **(NR2)** the Markov kernel $\mathbb{P}^{\xi}_{f(X)}$ is a regular version of the conditional distribution of Y given X, i.e. $\mathbb{P}^{Y|X}_f = \mathbb{P}^{\xi}_{f(X)}$.

(NR4) The regressor X is uniformly distributed on the interval $[0,1]$, i.e. $X \sim \mathcal{U}_{[0,1]}$, and thus $\mathbb{P}^X = \mathcal{U}_{[0,1]} = \lambda_{[0,1]}$. Denote by $\mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}^{Y|X}_f$ the joint distribution of (X, Y).

Under **(NR1)**-**(NR4)** given $f \in \mathbb{F}_2$ and $\mathbb{P}^{\xi} \in \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}$ the joint distribution $\mathcal{U}_f = \mathcal{U}_{[0,1]} \odot \mathbb{P}^{\xi}_{f(X)}$ of (X, Y) is fully specified and we set $\mathcal{U}_{\mathbb{F}_2 \times \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}} := (\mathcal{U}_f = \mathcal{U}_{[0,1]} \odot \mathbb{P}^{\xi}_{f(X)})_{f \in \mathbb{F}_2, \mathbb{P}^{\xi} \in \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}}$. We consider the statistical product experiment $\big(([0,1] \times \mathbb{R})^n, (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n}, \mathcal{U}^{\otimes n}_{\mathbb{F}_2 \times \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}} := (\mathcal{U}^{\otimes n}_f)_{f \in \mathbb{F}_2, \mathbb{P}^{\xi} \in \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}}\big)$ of size $n \in \mathbb{N}$ and for $f \in \mathbb{F}_2$ and $\mathbb{P}^{\xi} \in \mathrm{P}_{\{0\} \times \mathbb{R}^+_0}$ we denote by $((X_i, Y_i))_{i \in [\![n]\!]} \sim \mathcal{U}^{\otimes n}_f$ an iid. sample of $(X, Y) \sim \mathcal{U}_f = \mathcal{U}_{[0,1]} \odot \mathbb{P}^{\xi}_{f(X)}$. □

§19.03 **Notation** *(Reminder)*. Consider an *orthonormal system* $(\mathrm{u}_j)_{j \in \mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$. Then $\mathrm{U} : \mathbb{L}_2(\lambda_{[0,1]}) \to \ell_2$ with $h \mapsto \mathrm{U}h := h_{\bullet} = (h_j := \langle h, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})})_{j \in \mathbb{N}}$ is a surjective partial isometry $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$. Its adjoint operator $\mathrm{U}^{\star} \in \mathbb{L}(\ell_2, \mathbb{L}_2(\lambda_{[0,1]}))$ satisfies $\mathrm{U}^{\star} a_{\bullet} = \sum_{j \in \mathbb{N}} a_j \mathrm{u}_j =: \nu_{\mathbb{N}}(a_{\bullet} \mathrm{u}_{\bullet})$ for all $a_{\bullet} \in \ell_2$. We call $h_{\bullet} = (h_j)_{j \in \mathbb{N}}$ *(generalised) Fourier coefficients* and U *(generalised) Fourier series transform*. □

§19.04 **Remark**. Let $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$ be a generalised Fourier series transform as in Notation §19.03. For $f, h \in \mathcal{L}_2(\mathbb{P}^X) \subseteq \mathcal{L}_1(\mathbb{P}^X)$ we have $fh \in \mathcal{L}_1(\mathbb{P}^X)$ and thus $\mathbb{P}^X(fh) \in \mathbb{R}$. Keeping in mind that X and Y equals the coordinate map $\Pi_{[0,1]}$ and $\Pi_{\mathbb{R}}$, respectively, due to Assumption §19.02 **(NR1)**, i.e., $\xi \in \mathcal{L}_2(\mathbb{P}_f)$, hence $\xi h(X) \in \mathcal{L}_1(\mathbb{P}_f)$, and **(NR2)** we have $\mathbb{P}_f(\xi h(X)) = \mathbb{P}_f(\xi)\mathbb{P}^X(h) = 0$. Consequently, we obtain $Yh(X) = (f(X) + \xi)h(X) \in \mathcal{L}_1(\mathbb{P}_f)$ and $\mathbb{P}_f(Yh(X)) = \mathbb{P}^X(fh) \in \mathbb{R}$. Moreover, if in addition $f \in \mathcal{L}_{\infty}(\mathbb{P}^X)$ then we have also $fh \in \mathcal{L}_2(\mathbb{P}^X)$ which together with $\mathbb{P}_f(\xi^2 h^2(X)) = \mathbb{P}_f(\xi^2)\mathbb{P}^X(h^2) = \sigma^2_{\xi}\mathbb{P}^X(h^2) \in \mathbb{R}^+$ implies $Yh(X) \in \mathcal{L}_2(\mathbb{P}_f)$. Since $\lambda_{[0,1]} = \mathbb{P}^X$ and $\mathcal{U}_f = \mathbb{P}_f$ under Assumption §19.02 **(NR4)** for all $f, h \in \mathbb{L}_2(\lambda_{[0,1]})$ it follows immediately $\mathcal{U}_f(Yh(X)) = \lambda_{[0,1]}(fh) = \langle f, h \rangle_{\mathbb{L}_2(\lambda_{[0,1]})}$ identifying again equivalence classes and their representatives. Evidently, we have $\mathrm{u}_j \in \mathbb{L}_2(\lambda_{[0,1]})$ for all $j \in \mathbb{N}$ and the (generalised) Fourier coefficients $f_{\bullet} = (f_j)_{j \in \mathbb{N}} = \mathrm{U}f \in \ell_2$ of $f \in \mathbb{L}_2(\lambda_{[0,1]})$ fulfil $f_j = \langle f, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})} = \mathcal{U}_f(Y \mathrm{u}_j(X))$ for all $j \in \mathbb{N}$. □

§19.05 **Assumption**. The *orthonormal system* $(\mathrm{u}_j)_{j \in \mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$, and its associated *generalised Fourier series transform* $\mathrm{U} \in \mathbb{L}(\mathbb{L}_2(\lambda_{[0,1]}), \ell_2)$ with $h \mapsto \mathrm{U}h := h_{\bullet} = (h_j := \langle h, \mathrm{u}_j \rangle_{\mathbb{L}_2(\lambda_{[0,1]})})_{j \in \mathbb{N}}$, is fixed and known in advance. □

§19.06 **Remark**. Under Assumptions §19.02 and §19.05 we impose in the sequel that $f \in \mathbb{F}_2 \subseteq \mathbb{L}_{\infty}(\lambda_{[0,1]})$ which in turn for all $j \in \mathbb{N}$ implies $Y \mathrm{u}_j(X) \in \mathcal{L}_2(\mathcal{U}_f)$ with

$$\begin{aligned} \mathcal{U}_f(Y^2 \mathrm{u}^2_j(X)) = \mathcal{U}_f(\xi^2 \mathrm{u}^2_j(X)) + \mathcal{U}_f(f^2(X) \mathrm{u}^2_j(X)) &= \mathcal{U}_f(\xi^2)\mathbb{P}^X(\mathrm{u}^2_j) + \mathbb{P}^X(f^2 \mathrm{u}^2_j) \\ &= \sigma^2_{\xi} + \lambda_{[0,1]}(f^2 \mathrm{u}^2_j) \leqslant \sigma^2_{\xi} + \|f\|^2_{\mathbb{L}_{\infty}(\lambda_{[0,1]})} \in \mathbb{R}^+. \end{aligned}$$

Alternatively, if $\mathrm{u}_j \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then for all $f \in \mathbb{L}_2(\lambda_{[0,1]})$ it follows that $Y\mathrm{u}_j(X) \in \mathcal{L}_2(\mathcal{U}_f)$ with $\mathcal{U}_f(Y^2\mathrm{u}_j^2(X)) \leqslant \sigma_\xi^2 + \|\mathrm{u}_j^2\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}\|f\|^2_{\mathbb{L}_2(\lambda_{[0,1]})} \in \mathbb{R}^+$. □

§19.07 **Notation.** Setting $\mathcal{Z} := [0,1] \times \mathbb{R}$, $\mathscr{Z} := \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ and $\psi_j(X,Y) := Y\mathrm{u}_j(X) \in \mathscr{Z}$ for each $j \in \mathbb{N}$ under the Assumptions §19.02 and §19.05 the stochastic process $\psi_\bullet = (\psi_j)_{j\in\mathbb{N}} \in \mathscr{Z} \otimes 2^{\mathbb{N}}$ satisfies $\psi_j \in \mathcal{L}_1(\mathcal{U}_f)$ for each $j \in \mathbb{N}$. Similar to an Empirical mean model §10.07 we define $\widehat{f}_\bullet := \widehat{\mathbb{P}}_n(\psi_\bullet) \in \mathscr{Z}^{\otimes n} \otimes 2^{\mathbb{N}}$ with $z^n = ((x_i,y_i))_{i\in[\![n]\!]} \mapsto \widehat{f}_j(z^n) = (\widehat{\mathbb{P}}_n(\psi_j))(z^n) = n^{-1}\sum_{i\in[\![n]\!]}\psi_j(x_i,y_i) = n^{-1}\sum_{i\in[\![n]\!]} y_i\mathrm{u}_j(x_i)$ for each $j \in \mathbb{N}$. For $f \in \mathbb{L}_2(\lambda_{[0,1]})$ by construction $f_\bullet = (f_j = \mathcal{U}_f(\psi_j))_{j\in\mathbb{N}} \in 2^{\mathbb{N}}$ is the ℓ_2-mean of $\widehat{f}_\bullet$. Consequently, $\varepsilon_\bullet := n^{1/2}(\widehat{\mathbb{P}}_n - \mathcal{U}_f)(\psi_\bullet) = (\varepsilon_j = n^{1/2}(\widehat{\mathbb{P}}_n(\psi_j) - f_j))_{j\in\mathbb{N}} \in \mathscr{Z}^{\otimes n} \otimes 2^{\mathbb{N}}$ is centred, i.e. $\varepsilon_j \in \mathbb{L}_1(\mathcal{U}_f^{\otimes n})$ with $\mathcal{U}_f^{\otimes n}(\varepsilon_j) = 0$. Evidently, $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ is a noisy version of $f_\bullet$ (see Definition §10.19). Moreover, if $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then $\widehat{f}_\bullet$ admits a covariance function $\mathrm{cov}^f_{\bullet\bullet} \in \mathbb{R}^{\mathbb{N}^2}$ given for $j, j_\circ \in \mathbb{N}$ by

$$n\,\mathbb{C}\mathrm{ov}(\widehat{f}_j, \widehat{f}_{j_\circ}) = \mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_{j_\circ}) = \mathcal{U}_f^{\otimes n}(\varepsilon_j\varepsilon_{j_\circ}) = \mathcal{U}_f(\psi_j\psi_{j_\circ}) - \mathcal{U}_f(\psi_j)\mathcal{U}_f(\psi_{j_\circ})$$
$$= \mathcal{U}_f(Y^2\mathrm{u}_j(X)\mathrm{u}_{j_\circ}(X)) - f_j f_{j_\circ} =: \mathrm{cov}^f_{j,j_\circ}.$$

Consequently, we have $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^f_{\bullet\bullet})}$ and $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet \sim \mathrm{P}_{(f_\bullet, n^{-1}\mathrm{cov}^f_{\bullet\bullet})}$ (see Definition §10.19). □

§19.08 **Noisy regression coefficients.** Under Assumptions §19.02 and §19.05 the stochastic process $\varepsilon_\bullet = n^{1/2}(\widehat{\mathbb{P}}_n - \mathcal{U}_f)(\psi_\bullet)$ satisfies Assumption §10.04, i.e. $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$, and $\varepsilon_\bullet$ has mean zero under $\mathcal{U}_f^{\otimes n}$. The stochastic process $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ with ℓ_2-mean $f_\bullet$ is called a *noisy version* of the regression coefficients $f_\bullet = \mathrm{U}f \in \ell_2$, or *noisy regression coefficients* for short. Moreover, if $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then $\varepsilon_\bullet$ admits under $\mathcal{U}_f^{\otimes n}$ a covariance function $\mathrm{cov}^f_{\bullet\bullet} \in \mathbb{R}^{\mathbb{N}^2}$ given for $j, j_\circ \in \mathbb{N}$ by $\mathrm{cov}^f_{j,j_\circ} = \mathcal{U}_f(Y^2\mathrm{u}_j(X)\mathrm{u}_{j_\circ}(X)) - f_j f_{j_\circ}$. We eventually write $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^f_{\bullet\bullet})}$ and $\widehat{f}_\bullet \sim \mathrm{P}_{(f_\bullet, n^{-1}\mathrm{cov}^f_{\bullet\bullet})}$. If in addition $\varepsilon_\bullet$ admits a covariance operator $\Gamma_f \in \mathbb{L}(\ell_2)$ then we write $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma_f)}$ and $\widehat{f}_\bullet \sim \mathrm{P}_{(f_\bullet, n^{-1}\Gamma_f)}$ for short. □

§19.09 **Remark.** The centred stochastic process $\varepsilon_\bullet := (\varepsilon_j)_{j\in\mathbb{N}}$ of error terms in Definition §19.08 is in general not a white noise process. □

§19.10 **Lemma.** *Under Assumptions §19.02 and §19.05 consider $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ as in Definition §19.08.*

(i) *If $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then under $\mathcal{U}_f^{\otimes n}$, $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}^f_{\bullet\bullet})}$ admits a covariance operator $\Gamma_f \in \mathbb{L}(\ell_2)$ given by*

$$a_\bullet \mapsto \Gamma_f a_\bullet = (\nu_{\mathbb{N}}(\mathrm{cov}^f_{j,\bullet}a_\bullet) = \sum_{j_\circ\in\mathbb{N}} \mathrm{cov}^f_{j,j_\circ}a_{j_\circ})_{j\in\mathbb{N}}$$

where $\|\Gamma_f\|_{\mathbb{L}(\ell_2)} \leqslant \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}$.

(ii) *If $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ and $\sigma_\xi^2 \in \mathbb{R}^+_{\setminus 0}$ then $\Gamma_f \in \mathbb{L}(\ell_2)$ is invertible with inverse $\Gamma_f^{-1} \in \mathbb{L}(\ell_2)$ where $\|\Gamma_f^{-1}\|_{\mathbb{L}(\ell_2)} \leqslant \sigma_\xi^{-2}$.*

Consequently, if $\mathbb{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ then for all $a_\bullet \in \ell_2$ we have

$$\mathbb{v}_f^{-1}\|a_\bullet\|^2_{\ell_2} \leqslant \|a_\bullet\|^2_{\Gamma_f} = \langle \Gamma_f a_\bullet, a_\bullet\rangle_{\ell_2} \leqslant \mathbb{v}_f\|a_\bullet\|^2_{\ell_2}.$$

§19.11 **Proof** of Lemma §19.10. is given in the lecture. □

§19.12 **Reminder.** Consider the orthonormal basis $(\mathbb{1}_\bullet^{\{j\}})_{j\in\mathbb{N}}$ in ℓ_2 (compare Remark §15.12). If $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ from Lemma §19.10 (i) for each $j \in \mathbb{N}$ we obtain

$$\mathcal{U}_f^{\otimes n}(\varepsilon_j^2) = \mathcal{U}_f^{\otimes n}(|\nu_{\mathbb{N}}(\mathbb{1}_\bullet^{\{j\}}\varepsilon_\bullet)|^2) = \langle \Gamma_f\mathbb{1}_\bullet^{\{j\}}, \mathbb{1}_\bullet^{\{j\}}\rangle_{\ell_2} \leqslant (\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})\|\mathbb{1}_\bullet^{\{j\}}\|^2_{\ell_2} = \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}$$

Keeping the last identities in mind if $\mathfrak{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ then for all $j \in \mathbb{N}$ we have $\mathfrak{v}_f^{-1} \leqslant \mathcal{U}_f^{\otimes n}(\varepsilon_j^2) \leqslant \mathfrak{v}_f$ due to Lemma §19.10. □

§20 Projection regression estimator

§20.01 **Notation** *(Reminder).* Consider the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ as in Notation §10.11. For $w_\bullet \in \mathbb{R}^{\mathbb{N}}$ define the multiplication map $\mathrm{M}_{w_\bullet} : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ with $a_\bullet \mapsto \mathrm{M}_{w_\bullet} a_\bullet := w_\bullet a_\bullet$. Note that each $w_\bullet \in \mathbb{R}^{\mathbb{N}}$ is $2^{\mathbb{N}}$-$\mathscr{B}$-measurable. We denote by $\mathbb{M}_{\mathbb{R}^{\mathbb{N}}}$ the set of all multiplication maps defined on $\mathbb{R}^{\mathbb{N}}$. If in addition $w_\bullet \in \ell_\infty = \mathbb{L}_\infty(\mathbb{N}, 2^{\mathbb{N}}, \nu_{\mathbb{N}})$ then we have also $\mathrm{M}_{w_\bullet} : \ell_2 \to \ell_2$. We set $\mathbb{L}^{\mathrm{M}}(\ell_2) = \{\mathrm{M}_{w_\bullet} \in \mathbb{M}_{\mathbb{R}^{\mathbb{N}}} : w_\bullet \in \ell_\infty\} \subseteq \mathbb{L}(\ell_2)$ noting that $\|\mathrm{M}_{w_\bullet}\|_{\mathbb{L}(\ell_2)} = \sup\{\|w_\bullet a_\bullet\|_{\ell_2} : \|a_\bullet\|_{\ell_2} \leqslant 1\} \leqslant \|w_\bullet\|_{\ell_\infty}$ for each $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\ell_2)$. □

§20.02 **Reminder**. If $w_\bullet \in \ell_\infty$ then $\mathrm{M}_{w_\bullet} \in \mathbb{L}^{\mathrm{M}}(\ell_2)$, and $\mathrm{M}_{w_\bullet}^\dagger : \ell_2 \supseteq \mathrm{dom}(\mathrm{M}_{w_\bullet}^\dagger) \to \ell_2$. Moreover, we have $\mathrm{dom}(\mathrm{M}_{w_\bullet}) = \ell_2$, $\mathrm{ran}(\mathrm{M}_{w_\bullet}) = \ell_2 w_\bullet$ and $\ker(\mathrm{M}_{w_\bullet}) = \ell_2 \mathbb{1}_\bullet^{\mathcal{N}_w}$ with $\mathcal{N}_w = \{j \in \mathbb{N} : w_j = 0\} \in 2^{\mathbb{N}}$ (see Property §11.03), and $\mathrm{dom}(\mathrm{M}_{w_\bullet}^\dagger) = \ell_2 w_\bullet \oplus \ell_2 \mathbb{1}_\bullet^{\mathcal{N}_w}$ (see Property §11.05). Consequently, if in addition $\nu_{\mathbb{N}}(\mathcal{N}_w) = 0$ or in equal $w_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, then $w_\bullet^\dagger = w_\bullet^{-1} \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$, hence $w_\bullet^{2\dagger} = w_\bullet^{-2} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$, and $\ell_2^w = \mathrm{dom}(\mathrm{M}_{w_\bullet^{-1}}) = \ell_2 w_\bullet = \mathbb{L}_2(w_\bullet^{-2}\nu_{\mathbb{N}}) =: \ell_2(w_\bullet^{-2})$. For each $m \in \mathbb{N}$ we write $\mathbb{1}_\bullet^m = (\mathbb{1}_j^m)_{j\in\mathbb{N}} := \mathbb{1}_\bullet^{[\![m]\!]}$ and $\mathbb{1}_\bullet^{m|\perp} := \mathbb{1}_\bullet - \mathbb{1}_\bullet^m$ with $[\![m]\!] := [-m, m] \cap \mathbb{N}$. Consequently, $\mathrm{M}_{\mathbb{1}_\bullet^m} \in \mathbb{L}^{\geqslant}(\ell_2)$ and $\mathrm{M}_{\mathbb{1}_\bullet^{m|\perp}} \in \mathbb{L}^{\geqslant}(\ell_2)$ is the *orthogonal projection* onto the linear subspace $\ell_2 \mathbb{1}_\bullet^m \subseteq \ell_2$ and its orthogonal complement $\ell_2 \mathbb{1}_\bullet^{m|\perp} = (\ell_2 \mathbb{1}_\bullet^m)^\perp \subseteq \ell_2$, respectively, that is $\ell_2 = \ell_2 \mathbb{1}_\bullet^m \oplus \ell_2 \mathbb{1}_\bullet^{m|\perp}$ (see Property §11.07). Finally, given $h_\bullet = \mathrm{U} h \in \ell_2$ for $h \in \mathbb{L}_2(\lambda_{[0,1]})$ we consider the orthogonal projections $h_\bullet^m = h_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m$ and $h^m := \mathrm{U}^\star h_\bullet^m \in \mathbb{L}_2(\lambda_{[0,1]})$ (Definition §11.08). □

§20.03 **Notation** *(Reminder).* Consider the stochastic processes $\varepsilon_\bullet = n^{1/2}(\widehat{\mathbb{P}}_n - \mathcal{U}_f)(\psi_\bullet) \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ given in Definition §19.08. The observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U} f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19). Under Assumptions §19.02 and §19.05 $\varepsilon_\bullet$ is centred and admits a covariance function $\mathrm{cov}_{\bullet,\bullet}^f \in \mathbb{R}^{\mathbb{N}^2}$ given in Definition §19.08, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(0,\mathrm{cov}_{\bullet,\bullet}^f)}$ and $\widehat{f}_\bullet \sim \mathrm{P}_{(f_\bullet, n^{-1}\mathrm{cov}_{\bullet,\bullet}^f)}$. If in addition $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then $\varepsilon_\bullet$ admits a covariance operator $\Gamma_f \in \mathbb{L}^{\geqslant}(\ell_2)$ given in Lemma §19.10, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(0,\Gamma_f)}$ and $\widehat{f}_\bullet \sim \mathrm{P}_{(f_\bullet, n^{-1}\Gamma_f)}$ □

§20.04 **Definition**. Given a noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U} f \in \ell_2$ for each $m \in \mathbb{N}$ we call $\widehat{f}_\bullet^m := \widehat{f}_\bullet \mathbb{1}_\bullet^m$ *orthogonal projection estimator (OPE)* of $f_\bullet$. □

§20.05 **Remark**. If $f = \mathrm{U}^\star f_\bullet$ (for example $(\mathrm{u}_j)_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathbb{L}_2(\lambda_{[0,1]})$), then we have

$$\|\mathrm{U}^\star \widehat{f}_\bullet^m - f\|^2_{\mathbb{L}_2(\lambda_{[0,1]})} = \|\widehat{f}_\bullet^m - f_\bullet\|^2_{\ell_2}.$$

In this situation all results for the OPE $\widehat{f}_\bullet^m$ of the regression coefficients immediately transfer onto the *orthogonal projection regression estimator* $\widehat{f}^m := \mathrm{U}^\star \widehat{f}_\bullet^m$ of the regression function f. □

§20|01 Global and maximal global $\mathfrak{v}$-risk

We measure first the accuracy of the OPE $\widehat{f}_\bullet^m = \widehat{f}_\bullet \mathbb{1}_\bullet^m$ of $f_\bullet^m = f_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m$ with $f_\bullet = \mathrm{U} f \in \ell_2$ by a global mean-$\mathfrak{v}$-error, i.e. $\mathfrak{v}$-risk.

§20.06 **Reminder**. If $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $f_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ then we have $f_\bullet^m = f_\bullet \mathbb{1}_\bullet^m \in \ell_2(\mathfrak{v}_\bullet^2)$ too and $\|f_\bullet^m - f_\bullet\|^2_{\mathfrak{v}} = o(1)$ as $m \to \infty$ (Property §11.09). Moreover, $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ given in Definition §19.08 satisfies $\mathfrak{v}_\bullet \varepsilon_\bullet \mathbb{1}_\bullet^m \in \ell_2$ (note that $\mathbb{1}_\bullet^m \in \ell_2$ and $\mathfrak{v}_\bullet \mathbb{1}_\bullet^m, \varepsilon_\bullet \mathbb{1}_\bullet^m \in \ell_\infty$) and thus also

$$n^{-1/2}\mathfrak{v}_\bullet \varepsilon_\bullet \mathbb{1}_\bullet^m + \mathfrak{v}_\bullet f_\bullet^m = \mathfrak{v}_\bullet \widehat{f}_\bullet^m \in \ell_2. \tag{20.01}$$

Finally, under Assumptions §19.02 and §19.05 and $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ due to Lemma §19.10 we have $\mathcal{U}_f^{\otimes n}(\varepsilon_\bullet^2) \in \ell_\infty$, more precisely, $\|\mathcal{U}_f^{\otimes n}(\varepsilon_\bullet^2)\|_{\ell_\infty} \leqslant \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}$ (see Reminder §19.12). □

§20|01|01 Global $\mathfrak{v}$-risk

§20.07 **Proposition** *(Upper bound).* *Let Assumptions §19.02 and §19.05, $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $f_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ be satisfied and for all $n, m \in \mathbb{N}$ set*

$$\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) := \|f_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\circ}(f_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \quad (20.02)$$

If $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then we have $\mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{m_n^\circ} - f_\bullet\|_{\mathfrak{v}}^2) \leqslant 1 \vee (\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})\, \mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet)$.

§20.08 **Proof** of Proposition §20.07. is given in the lecture. □

§20.09 **Oracle inequality.** *Under Assumptions §19.02 and §19.05 let $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $f_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$. If in addition $\mathbb{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ then $\mathbb{v}_f^{-1} \leqslant \mathbb{v}_j^f := \mathcal{U}_f^{\otimes n}(\varepsilon_j^2) \leqslant \mathbb{v}_f$ for all $j \in \mathbb{N}$ (see Reminder §19.12), and hence Property §12.15 implies*

$$\mathbb{v}_f^{-1}\,\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) \leqslant \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2) = n^{-1}\nu_{\mathbb{N}}(\mathbb{v}_\bullet^f \mathfrak{v}_\bullet^2 \mathbb{1}_\bullet^m) + \|f_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2$$
$$\leqslant \mathbb{v}_f \mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) \quad \text{for all } m, n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality (see Definition §12.14)

$$\mathbb{v}_f^{-1}\,\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2) \leqslant \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{m_n^\circ} - f_\bullet\|_{\mathfrak{v}}^2)$$
$$\leqslant \mathbb{v}_f\,\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{v}_f^2 \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2), \quad (20.03)$$

and, hence $\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet)$, m_n° and the statistic $\widehat{f}_\bullet^{m_n^\circ}$, respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant $\mathbb{v}_f^2$). We observe that $\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) = \mathrm{o}(1)$ as $n \to \infty$ (Remark §12.16), and thus, $\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet)$ is an oracle rate. However, note that the oracle dimension $m_n^\circ = m_n^\circ(f_\bullet, \mathfrak{v}_\bullet)$ depends on the unknown regression coefficients $f_\bullet$, and thus also the oracle optimal statistic $\widehat{f}_\bullet^{m_n^\circ}$. In other words $\widehat{f}_\bullet^{m_n^\circ}$ is not a feasible estimator. □

§20.10 **Illustration.** We illustrate the last results considering usual behaviour for the bias and variance term. We distinguish the following two cases

(p) $\mathfrak{v}_\bullet \in \ell_2$ or there is $m \in \mathbb{N}$ with $\|f_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2 = 0$,

(np) $\mathfrak{v}_\bullet \notin \ell_2$ and for all $m \in \mathbb{N}$ holds $\|f_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2 \in \mathbb{R}^+_{\setminus 0}$.

Interestingly, in case **(p)** the oracle bound is parametric, that is, $n\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) = \mathrm{O}(1)$, in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) = \infty$. In case **(np)** consider the following two specifications:

Table 01 [§20]

Order of the oracle rate $\mathrm{R}^{\circ}_n(f_\bullet, \mathfrak{v}_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$ $\mathfrak{v}_j^2 = j^{2v}$	$(a \in \mathbb{R}^+_{\setminus 0})$ f_j^2	(squarred bias) $\|f_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|^2_{\mathfrak{v}}$	(variance) $\|\mathbb{1}_\bullet^m\|^2_{\mathfrak{v}}$	m°_n	$\mathrm{R}^{\circ}_n(f_\bullet, \mathfrak{v}_\bullet)$
(o) $v \in (-1/2, a)$	j^{-2a-1}	$m^{-2(a-v)}$	m^{2v+1}	$n^{\frac{1}{2a+1}}$	$n^{-\frac{2(a-v)}{2a+1}}$
$v = -1/2$	j^{-2a-1}	m^{-2a-1}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a+1}}$	$\frac{\log n}{n}$
(s) $v + 1/2 \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+} e^{-m^{2a}}$	m^{2v+1}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{2v+1}{2a}}}{n}$
$v = -1/2$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 01 [§20] the order of the oracle rate $\mathrm{R}^{\circ}_n(f_\bullet, \mathfrak{v}_\bullet)$ is depict for $v \geqslant -1/2$ only. In case $v < -1/2$ the oracle rate $\mathrm{R}^{\circ}_n(f_\bullet, \mathfrak{v}_\bullet)$ is parametric. □

§20|01|02 Maximal global $\mathfrak{v}$-risk

§20.11 **Assumption**. Consider weights $\mathfrak{a}_\bullet, \mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\mathfrak{v})_\bullet := (\mathfrak{a}_j \mathfrak{v}_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet \mathfrak{v}_\bullet \in \ell_\infty$. We write $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete*, i.e an *orthonormal basis* in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}^2_\bullet = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2)** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2 \mathrm{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \tau^2_{\mathfrak{a},\mathrm{u}}$ for $\tau_{\mathfrak{a},\mathrm{u}} \in [1, \infty)$. □

§20.12 **Reminder**. Under Assumption §20.11 we have $\ell_2^{\mathfrak{a}_\bullet} = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}_\bullet^{-1}}) = \ell_2 \mathfrak{a}_\bullet \subseteq \ell_2$ and the three measures $\nu_{\mathbb{N}}$, $\mathfrak{a}_\bullet^{-2}\nu_{\mathbb{N}}$ and $\mathfrak{v}_\bullet^2 \nu_{\mathbb{N}}$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\ell_2^{\mathfrak{a}}$ endowed with $\|\cdot\|_{\mathfrak{a}^{-1}} = \|\mathrm{M}_{\mathfrak{a}_\bullet^{-1}} \cdot\|_{\ell_2}$ and given a constant $\mathrm{r} \in \mathbb{R}^+_{\setminus 0}$ the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}} := \{b_\bullet \in \ell_2^{\mathfrak{a}} : \|b_\bullet\|_{\mathfrak{a}^{-1}} \leqslant \mathrm{r}\} \subseteq \ell_2^{\mathfrak{a}}$. Since $(\mathfrak{a}\mathfrak{v})_\bullet \in \ell_\infty$, and hence $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ we have $\ell_2^{\mathfrak{a}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ (Property §11.15), and $\|b_\bullet \mathbb{1}_\bullet^{m\vert\perp}\|_{\mathfrak{v}} \leqslant \mathrm{r}\,(\mathfrak{a}\mathfrak{v})_{(m)}$ for all $b_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.17). □

§20.13 **Lemma**. *Under Assumption §20.11 set*

$$\mathbb{F}_2^{\mathfrak{a},\mathrm{r}} := \big\{h \in \mathbb{L}_2(\lambda_{[0,1]}) \colon h_\bullet = \mathrm{U}h \in \ell_2^{\mathfrak{a},\mathrm{r}}\big\}. \tag{20.04}$$

Then we have $\sup\big\{\|h\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \colon h \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\} \leqslant \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}$.

§20.14 **Proof** of Lemma §20.13. is given in the lecture. □

§20.15 **Proposition** *(Upper bound)*. *Let Assumptions §19.02 and §20.11 be satisfied. For $n \in \mathbb{N}$ considering $m^\star_n \in \mathbb{N}$ and $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \in \mathbb{R}^+$ as in* (12.06) *(Proposition §12.21) we have*

$$\sup\big\{\mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{m^\star_n} - f_\bullet\|^2_{\mathfrak{v}}) \colon f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\} \leqslant \mathrm{C}\,\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet).$$

with constant $\mathrm{C} = \sigma_\xi^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + \mathrm{r}^2$.

§20.16 **Proof** of Proposition §20.15. is given in the lecture. □

§20.17 **Illustration**. The *trigonometric basis* given for $x \in [0,1]$ by

$$\mathrm{u}_1 := \mathbb{1}_{[0,1]},\ \mathrm{u}_{2k}(x) := \sqrt{2}\cos(2\pi k t),\ \mathrm{u}_{2k+1}(x) := \sqrt{2}\sin(2\pi k t),\ k \in \mathbb{N},$$

is an orthonormal basis of $\mathbb{L}_2(\lambda_{[0,1]})$, hence it satisfies Assumption §20.11 **(os1)**. Keeping in mind that $\|\mathrm{u}_j^2\|_{\mathbb{L}_2(\lambda_{[0,1]})} \leqslant 2$ for all $j \in \mathbb{N}$ also the Assumption §20.11 **(os3)** is satisfied for all $\mathfrak{a}_\bullet \in \ell_2$

because $\tau_{\mathfrak{a},\mathfrak{u}}^2 \leqslant 2\|\mathfrak{a}_\bullet\|_{\ell_2}^2$ (see also Illustration §16.18). In Table 02 [§12] (Illustration §12.26) the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is depict for the two cases **(o)** and **(s)**. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^+$. □

§20.18 **Remark**. In Proposition §20.15 an upper bound is shown under Assumption §19.02 which amongst others imposes that $\mathcal{U}_f = \mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi$ with $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$. Recall that $\mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+} \subseteq \mathcal{W}(\mathscr{B})$ denotes the subset of all probability distributions over $(\mathbb{R}, \mathscr{B})$ with finite second moment and mean zero. For $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$ let us further introduce $\mathrm{P}_{\{0\}\times(0,\sigma^2]} \subseteq \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$ containing only probability distributions with second moment bounded by σ^2. In what follows we treat the distribution $\mathbb{P}^\xi$ of the error term as a nuisance parameter and consider the maximal risk over both $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ and $\mathrm{P}_{\{0\}\times(0,\sigma^2]}$ (see Definition §13.07). □

§20.19 **Corollary** *(Upper bound).* *Let Assumptions §19.02 and §20.11 be satisfied. For $n \in \mathbb{N}$ considering $m_n^\star \in \mathbb{N}$ and $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \in \mathbb{R}^+$ as in (12.06) (Proposition §12.21) we have*

$$\sup\big\{(\mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi)^{\otimes n}\big(\|\widehat{f}_\bullet^{m_n^\star} - f_\bullet\|_{\mathfrak{v}}^2\big) \colon f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times(0,\sigma^2]}\big\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet).$$

with constant $\mathrm{C} = \sigma^2 + \mathrm{r}^2\tau_{\mathfrak{a},\mathfrak{u}}^2 + \mathrm{r}^2$.

§20.20 **Proof** of Corollary §20.19. is given in the lecture. □

§20|02 Local and maximal local ϕ-risk

We measure secondly the accuracy of the OPE $\widehat{f}_\bullet^m = \widehat{f}_\bullet \mathbb{1}_\bullet^m$ of $f_\bullet^m = f_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m$ with $f_\bullet = \mathrm{U} f \in \ell_2$ by a local mean-ϕ-error, i.e. ϕ-risk.

§20.21 **Reminder**. If $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}}) := \big\{a_\bullet \in \ell_2 \colon \phi_\bullet a_\bullet \in \ell_1\big\}$, then we have $f_\bullet^m = f_\bullet \mathbb{1}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ too and $|\phi\nu_{\mathbb{N}}(f_\bullet - f_\bullet^m)| = \mathrm{o}(1)$ as $m \to \infty$ (Property §11.22). Moreover, $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ given in Definition §19.08 satisfies $\varepsilon_\bullet \mathbb{1}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ (note that $\phi_\bullet \mathbb{1}_\bullet^m, \varepsilon_\bullet \mathbb{1}_\bullet^m \in \ell_2$) and thus also

$$n^{-1/2}\varepsilon_\bullet \mathbb{1}_\bullet^m + f_\bullet^m = \widehat{f}_\bullet^m \in \mathrm{dom}(\phi\nu_{\mathbb{N}}). \tag{20.05}$$

Finally, under Assumptions §19.02 and §19.05 and $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ due to Lemma §19.10 (i) the process $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ admits a covariance operator $\Gamma_f \in \mathbb{L}(\ell_2)$, i.e. $\varepsilon_\bullet \sim \mathrm{P}_{(\mathfrak{0},\Gamma_f)}$, satisfying $\|\Gamma_f\|_{\mathbb{L}(\ell_2)} \leqslant \sigma_\xi^2 + \|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}^2$. □

§20|02|01 Local ϕ-risk

§20.22 **Proposition** *(Upper bound).* *Let Assumptions §19.02 and §19.05, $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ and $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ be satisfied and for all $n, m \in \mathbb{N}$ set*

$$\mathrm{R}_n^m(f_\bullet, \phi_\bullet) := |\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\textit{and} \quad \mathrm{R}_n^\circ(f_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\circ}(f_\bullet, \phi_\bullet) := \min\big\{\mathrm{R}_n^m(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \tag{20.06}$$

If $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ then we have $\mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{m_n^\circ} - f_\bullet)|^2) \leqslant 1 \vee (\sigma_\xi^2 + \|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}^2)\,\mathrm{R}_n^\circ(f_\bullet, \phi_\bullet)$.

§20.23 **Proof** of Proposition §20.22. is given in the lecture. □

§20.24 **Oracle inequality**. *Under Assumptions §19.02 and §19.05 let* $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ *and* $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$. *If in addition* $\mathbb{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ *then* $\max(\|\Gamma_f\|_{\mathbb{L}(\ell_2)}, \|\Gamma_f^{-1}\|_{\mathbb{L}(\ell_2)}) \leqslant \mathbb{v}_f$ *(see Lemma §19.10), and hence Property §12.36 implies*

$$\mathbb{v}_f^{-1}\mathrm{R}_n^m(f_\bullet,\phi_\bullet) \leqslant \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet)|^2) = n^{-1}\|\phi_\bullet \mathbb{1}_\bullet^m\|^2_{\Gamma_f} + |\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})|^2$$
$$\leqslant \mathbb{v}_f \mathrm{R}_n^m(f_\bullet,\phi_\bullet) \quad \textit{for all } m,n \in \mathbb{N}.$$

As a consequence we immediately obtain the following oracle inequality (see Definition §12.34)

$$\mathbb{v}_f^{-1}\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet)|^2) \leqslant \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{m_n^\circ} - f_\bullet)|^2)$$
$$\leqslant \mathbb{v}_f \mathrm{R}_n^\circ(f_\bullet,\phi_\bullet) \leqslant \mathbb{v}_f^2 \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet)|^2), \quad (20.07)$$

and hence, $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$, m_n° *and the statistic* $\widehat{f}_\bullet^{m_n^\circ}$, *respectively, is an oracle bound, an oracle dimension and oracle optimal (up to the constant* $\mathbb{v}_f^2$*). We observe that* $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet) = \mathrm{o}(1)$ *as* $n \to \infty$ *(Remark §12.37), and thus,* $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$ *is an oracle rate. However, note that the oracle dimension* $m_n^\circ = m_n^\circ(f_\bullet,\phi_\bullet)$ *depends on the unknown regression coefficients* $f_\bullet$*, and thus also the oracle optimal statistic* $\widehat{f}_\bullet^{m_n^\circ}$. *In other words* $\widehat{f}_\bullet^{m_n^\circ}$ *is not a feasible estimator.* □

§20.25 **Illustration**. We illustrate the last results considering usual behaviour for both the variance and the bias term. Similar to the two cases **(p)** and **(np)** in Illustration §20.10 we distinguish here the following two cases

(p) $\phi_\bullet \in \ell_2$ or there is $K \in \mathbb{N}$ with $\sup\{|\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m \in \mathbb{N} \cap [K,\infty)\} = 0$,

(np) $\phi_\bullet \notin \ell_2$ and for all $m \in \mathbb{N}$ holds $\sup\{|\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 : m \in \mathbb{N} \cap [K,\infty)\} \in \mathbb{R}^+_{\setminus 0}$.

In case **(p)** the oracle bound is again parametric, i.e. $n\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet) = \mathrm{O}(1)$, while in case **(np)** the oracle bound is nonparametric, i.e. $\lim_{n\to\infty} n\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet) = \infty$. In case **(np)** consider the following two specifications

Table 02 [§20]

Order of the oracle rate $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$ as $n \to \infty$

$(j \in \mathbb{N})$	$(a \in \mathbb{R}^+_{\setminus 0})$	(squarred bias)	(variance)		
$\phi_j = j^{v-1/2}$	f_j	$\lvert\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m\lvert\perp})\rvert^2$	$\lVert\mathbb{1}_\bullet^m\rVert^2_\phi$	m_n°	$\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$
(o) $v \in (0,a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{1}{2a}}$	$n^{-\frac{(a-v)}{a}}$
$v = 0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	$(\frac{n}{\log n})^{\frac{1}{2a}}$	$\frac{\log n}{n}$
(s) $v \in \mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+}e^{-2m^{2a}}$	m^{2v}	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$
$v = 0$	$e^{-j^{2a}}$	$m^{(1-2a)_+}e^{-m^{2a}}$	$\log m$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$

We note that in Table 02 [§20] the order of the oracle rate $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$ is depict for $v \geqslant 0$ only. For $v < 0$ the oracle rate $\mathrm{R}_n^\circ(f_\bullet,\phi_\bullet)$ is parametric. □

§20|02|02 Maximal local ϕ-risk

§20.26 **Assumption**. Consider $\phi_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\phi)_\bullet := (\mathfrak{a}_j\phi_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\phi_\bullet \in \ell_2$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete*, i.e an *orthonormal basis* in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}_\bullet^2 = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1],\mathscr{B}_{[0,1]})$ satisfies **(os2)** $\|\nu_{\mathbb{N}}(\mathfrak{a}_\bullet^2\mathrm{u}_\bullet^2)\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \tau_{\mathfrak{a},\mathrm{u}}^2$ for $\tau_{\mathfrak{a},\mathrm{u}} \in [1,\infty)$. □

§20.27 **Reminder**. Under Assumption §20.26 we have $\ell_2^{\mathfrak{a}} = \mathrm{dom}(\mathrm{M}_{\mathfrak{a}_\bullet^{-1}}) = \ell_2\mathfrak{a}_\bullet \subseteq \ell_2$ and the three measures $\nu_{\mathbb{N}}$, $\mathfrak{a}_\bullet^{-2}\nu_{\mathbb{N}}$ and $|\phi_\bullet|\nu_{\mathbb{N}}$ dominate mutually each other, i.e. they share the same null sets (see Property §11.05). We consider $\ell_2^{\mathfrak{a}_\bullet}$ endowed with $\|\cdot\|_{\mathfrak{a}^{-1}} = \|\mathrm{M}_{\mathfrak{a}_\bullet^{-1}}\cdot\|_{\ell_2}$ and given a constant $\mathrm{r} \in \mathbb{R}_{\setminus 0}^+$ the ellipsoid $\ell_2^{\mathfrak{a},\mathrm{r}} := \{a_\bullet \in \ell_2^{\mathfrak{a}_\bullet} : \|a_\bullet\|_{\mathfrak{a}^{-1}} \leqslant \mathrm{r}\} \subseteq \ell_2^{\mathfrak{a}}$. Since $(\mathfrak{a}\phi)_\bullet \in \ell_2$, and hence $\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \|(\mathfrak{a}\phi)_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_2} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$ ($\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence) we have $\ell_2^{\mathfrak{a}} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ (Property §11.27), and $|\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})| \leqslant \mathrm{r}\,\|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi$ for all $f_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ (Lemma §11.29). □

§20.28 **Remark**. Under Assumption §20.26 considering the set $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ of regression functions in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (20.04) we have $\|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}$ for all $f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ due to Lemma §20.13. □

§20.29 **Proposition** *(Upper bound)*. *Let Assumptions §19.02 and §20.26 be satisfied. For $n \in \mathbb{N}$ considering $m_n^\star \in \mathbb{N}$ and $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \in \mathbb{R}^+$ as in (12.13) (Proposition §12.42) we have*

$$\sup\big\{\mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{m} - f_\bullet)|^2) \colon f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet).$$

with constant $\mathrm{C} = \sigma_\xi^2 + \mathrm{r}^2\tau_{\mathfrak{a},\mathrm{u}}^2$.

§20.30 **Proof** of Proposition §20.29. is given in the lecture. □

§20.31 **Illustration**. Consider the *trigonometric basis* as in Illustration §20.17 which satisfies Assumption §20.26 for all $\mathfrak{a}_\bullet \in \ell_2$. In Table 04 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §12.47. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^+$. □

§20.32 **Corollary** *(Upper bound)*. *Let Assumptions §19.02 and §20.26 be satisfied. For $n \in \mathbb{N}$ considering $m_n^\star \in \mathbb{N}$ and $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \in \mathbb{R}^+$ as in (12.13) (Proposition §12.42) we have*

$$\sup\big\{(\mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi)^{\otimes n}\big(|\phi\nu_{\mathbb{N}}((\widehat{f}_\bullet^{m} - f_\bullet)\mathbb{1}_\bullet^m)|^2\big) \colon f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times(0,\sigma^2]}\big\} \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet).$$

with constant $\mathrm{C} = \sigma^2 + \mathrm{r}^2\tau_{\mathfrak{a},\mathrm{u}}^2$.

§20.33 **Proof** of Corollary §20.32. is given in the lecture. □

§21 Minimax optimal regression

§21|01 Maximal local ϕ-risk

§21.01 **Reminder** *(Maximal local ϕ-risk)*. Under Assumptions §19.02 and §20.26 the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U}f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §20.26 in Proposition §20.29 is shown an upper bound for a maximal local ϕ-risk of an OPE over the class $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}} \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ of regression functions defined in (20.04). More precisely, assuming $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$ with $\sigma_\xi^2 = \mathbb{P}^\xi(\mathrm{id}_{\mathbb{R}}^2) \in \mathbb{R}_{\setminus 0}^+$ and for $f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ setting $\mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi$ the performance of the OPE $\widehat{f}_\bullet^{m} = \widehat{f}_\bullet\mathbb{1}_\bullet^m \in \ell_2\mathbb{1}_\bullet^m \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m \in \mathbb{N}$ is measured by its maximal local ϕ-risk, that is

$$\mathcal{R}_n^\phi[\,\widehat{f}_\bullet^{m} \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] := \sup\big\{\mathcal{U}_f^{\otimes}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{m} - f_\bullet)|^2\big) \colon \mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi, f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

indicating explicitaly the dependence on the error distribution $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$. Let us recall (12.13) (Proposition §12.42) where for $n, m \in \mathbb{N}$ we have defined

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) := \|\mathfrak{a}_\bullet \mathbb{1}_\bullet^{m|\perp}\|_\phi^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (21.01)$$

By Proposition §20.29 under Assumptions §19.02 and §20.26 the maximal local ϕ-risk of an OPE $\widehat{f}_\bullet^{m_n^\star}$ with optimally choosen dimension $m_n^\star$ as in (21.01) satisfies

$$\mathcal{R}_n^\phi[\,\widehat{f}_\bullet^{m_n^\star}\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}},\{\mathbb{P}^\xi\}\,] \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet,\phi_\bullet)$$

with $\mathrm{C} = \sigma_\xi^2 + \mathrm{r}^2\tau_{\mathfrak{a},\mathrm{u}}^2$. □

§21.02 **Lemma** *(Lower bound based on two hypotheses).* *Given $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$ if there are $f^0, f^1 \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ with associated probability measures $\mathbb{P}_0 := \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^0(X)}$ and $\mathbb{P}_1 := \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^1(X)}$ such that $\mathrm{KL}(\mathbb{P}_0|\mathbb{P}_1) \leqslant 2n^{-1}$ then for all $n \geqslant 2$ we have*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^\phi[\,\widetilde{f}_\bullet\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}},\{\mathbb{P}^\xi\}\,] \geqslant \frac{1}{64}|\phi\nu_{\mathbb{N}}(f_\bullet^0 - f_\bullet^1)|^2.$$

where the infimum is taken over all possible estimators.

§21.03 **Proof** of Lemma §21.02. is given in the lecture. □

§21.04 **Remark.** If we consider furthermore candidate regression functions $f^0 := f^*$ and $f^1 = -f^*$ for some $f^* \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$, and hence by definition $f^0, f^1 \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$, then trivially $|\phi\nu(f_\bullet^0 - f_\bullet^1)|^2 = 4|\phi\nu_{\mathbb{N}}(f_\bullet^*)|^2$. If the associated probability measures $\mathcal{U}_{f^0} = \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^0(X)}$ and $\mathcal{U}_{f^1} = \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^1(X)}$ satisfy $\mathrm{KL}(\mathcal{U}_{f^0}|\mathcal{U}_{f^1}) \leqslant 2n^{-1}$ then due to Lemma §21.02 for all $n \geqslant 2$ we have

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^\phi[\,\widetilde{f}_\bullet\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}},\{\mathbb{P}^\xi\}\,] \geqslant \frac{1}{16}|\phi\nu_{\mathbb{N}}(f_\bullet^*)|^2. \tag{21.02}$$

We find a minimax-optimal lower bound by constructing a candidate $f^* = \mathrm{U}^\star f_\bullet^* \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ that has the largest possible $|\phi\nu_{\mathbb{N}}(f_\bullet^*)|^2$-value but $\mathcal{U}_{f^0}^{\otimes n}$ and $\mathcal{U}_{f^1}^{\otimes n}$ are still statistically indistinguishable in the sense that $\mathrm{KL}(\mathcal{U}_{f^0}|\mathcal{U}_{f^1}) \leqslant 2n^{-1}$. □

§21.05 **Assumption.** The distribution $\mathbb{P}^\xi \in \mathcal{W}(\mathscr{B})$ admits a Lebesgue-density $\mathbb{p}^\xi := \mathrm{d}\mathbb{P}^\xi/\mathrm{d}\lambda$ and $\xi + x \sim \mathbb{P}_x^\xi$ for all $x \in \mathbb{R}$. There exist constants $\mathrm{C}_\xi, x_\xi \in \mathbb{R}_{\setminus 0}^+$ such that

$$\forall\, x \in [-x_\xi, x_\xi]: \quad \mathrm{KL}(\mathbb{P}^\xi|\mathbb{P}_x^\xi) = \int \mathbb{p}^\xi(u)\log\Big(\frac{\mathbb{p}^\xi(u)}{\mathbb{p}^\xi(u-x)}\Big)\lambda(\mathrm{d}u) \leqslant \mathrm{C}_\xi\, x^2.$$

□

§21.06 **Lemma.** *Let $\mathbb{P}^\xi \in \mathcal{W}(\mathscr{B})$ satisfy Assumption §21.05 with constants $\mathrm{C}_\xi, x_\xi \in \mathbb{R}_{\setminus 0}^+$ and under Assumption §20.26 let $f_\bullet^\star \in \ell_2^{\mathfrak{a},\mathrm{r}}$ fulfill $\|f_\bullet^\star\|_{\mathfrak{a}^{-1}} \leqslant x_\xi/(2\tau_{\mathfrak{a},\mathrm{u}})$. Setting $f^0 := \mathrm{U}^\star f^\star$ and $f^1 := -\mathrm{U}^\star f^\star$ the distributions $\mathcal{U}_{f^\tau} := \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^\tau(X)}$, $\tau \in \{0,1\}$ satisfy $\mathrm{KL}(\mathcal{U}_{f^0}|\mathcal{U}_{f^1}) \leqslant 4\mathrm{C}_\xi\|f^\star\|^2_{\mathbb{L}_2(\lambda_{[0,1]})}$.*

§21.07 **Proof** of Lemma §21.06. is given in the lecture. □

§21.08 **Reminder.** Under Assumption §20.26 let in addition $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ (see Notation §13.23), then Assumption §13.24 is satisfied. If $\mathfrak{a}_2^2 > n^{-1}$ then exploiting the definition (21.01) of $m_n^\star$ we have $\mathfrak{a}_{m_n^\star}^2 > n^{-1} \geqslant \mathfrak{a}_{m_n^\star+1}^2$ (see Comment §13.25) which we use in the next proof. □

§21.09 **Proposition** *(Lower bound).* *Let $\mathbb{P}^\xi \in \mathcal{W}(\mathscr{B})$ satisfy Assumption §21.05 with constants $\mathrm{C}_\xi, x_\xi \in \mathbb{R}_{\setminus 0}^+$ and let Assumptions §19.02 and §20.26 be fulfilled. If $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ then for all $n \in \mathbb{N} \cap (1 \vee \mathfrak{a}_2^{-2}, \infty)$ we have*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^\phi[\,\widetilde{f}_\bullet\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}},\{\mathbb{P}^\xi\}\,] \geqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet,\phi_\bullet) \tag{21.03}$$

with constant $\mathrm{C} := 16^{-1}\big(\mathrm{r}^2 \wedge x_\xi^2/(4\tau_{\mathfrak{a},\mathrm{u}}^2) \wedge 1/(2\mathrm{C}_\xi)\big)$ and infimum taken over all estimators.

§21.10 **Proof** of Proposition §21.09. is given in the lecture. □

§21.11 **Comment**. If ξ is normally distributed with mean zero and variance $\sigma_\xi^2 \in \mathbb{R}_{\setminus 0}^+$, i.e. $\xi \sim \mathrm{N}_{(0,\sigma_\xi^2)}$, then for all $x \in \mathbb{R}$ we have

$$\mathrm{KL}(\mathrm{N}_{(0,\sigma_\xi^2)}|\mathrm{N}_{(x,\sigma_\xi^2)}) = \mathrm{N}_{(0,\sigma_\xi^2)}\big(\log \frac{\mathrm{dN}_{(0,\sigma_\xi^2)}}{\mathrm{dN}_{(x,\sigma_\xi^2)}}\big) = \frac{x^2}{2\sigma_\xi^2}$$

and thus Assumption §21.05 holds with $\mathrm{C}_\xi = 1/(2\sigma_\xi^2)$ and $x_\xi^2 = \infty$ (see Proof §13.15). Consequently, from Proposition §21.09 we obtain immediately,

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^{\phi}[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathrm{N}_{(0,\sigma_\xi^2)}\}\,] \geqslant \mathrm{C}\ \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \tag{21.04}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{r}^2 \wedge \sigma_\xi^2)$ and infimum taken over all estimators. □

§21.12 **Corollary** *(Lower bound).* *Let Assumptions §19.02 and §20.26 be fulfilled and let $\sigma^2 \in \mathbb{R}_{\setminus 0}^+$. If $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ then for all $n \in \mathbb{N} \cap (1 \vee \mathfrak{a}_2^{-2}, \infty)$ we have*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^{\phi}[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \mathrm{P}_{\{0\}\times(0,\sigma^2]}\,] \geqslant \mathrm{C}\ \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet) \tag{21.05}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{r}^2 \wedge \sigma^2)$ and infimum taken over all estimators.

§21.13 **Proof** of Corollary §21.12. is given in the lecture. □

§21.14 **Illustration**. Consider the *trigonometric basis* as in Illustration §20.17 which satisfies Assumption §20.26 for all $\mathfrak{a}_\bullet \in \ell_2$ (see Illustration §20.31). In Table 04 [§12] the order of the rate $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §12.47. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^+$. In both cases the additional assumption $\mathfrak{a}_\bullet^2 \in (\mathbb{R}_{\setminus 0}^+)_{\downarrow 0}^{\mathbb{N}}$ is satisfied. Consequently, due to Proposition §21.09 the Table 04 [§12] presents the order of the *minimax rate* $\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \phi_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{f}_\bullet^{m_n^\star} = \widehat{f}_\bullet \mathbb{1}_\bullet^{m_n^\star} \in \ell_2 \mathbb{1}_\bullet^{m_n^\star} \subseteq \mathrm{dom}(\phi_{\mathbb{N}})$ with optimally selected dimension $m_n^\star$ (Proposition §20.29). We shall stress, that the order of $m_n^\star$ given in the Table 04 [§12] depends on the parameter $a \in \mathbb{R}_{\setminus 0}^+$ characterising the (abstract) smoothness of the density of interest which is generally not known in advance. □

§21|02 Maximal global $\mathfrak{v}$-risk

§21.15 **Reminder** *(Maximal global $\mathfrak{v}$-risk).* Under Assumptions §19.02 and §20.11 the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U} f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §20.11 in Proposition §20.15 is shown an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE over the class $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}} \subseteq \mathbb{L}_2(\lambda_{[0,1]})$ of regression functions defined in (20.04). More precisely, assuming $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\setminus 0}^+}$ with $\sigma_\xi^2 = \mathbb{P}^\xi(\mathrm{id}_{\mathbb{R}}^2) \in \mathbb{R}_{\setminus 0}^+$ and for $f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ setting $\mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi$ the performance of the OPE $f_\bullet^m = \widehat{f}_\bullet \mathbb{1}_\bullet^m \in \ell_2 \mathbb{1}_\bullet^m \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}$-risk, that is

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{f}_\bullet^m \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] := \sup\big\{\mathcal{U}_f^{\otimes n}\big(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2\big) \colon \mathcal{U}_f := \mathcal{U}_{[0,1]} \odot \mathbb{P}_{f(X)}^\xi, f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

Let us recall (12.06) (Proposition §12.21) where for $n, m \in \mathbb{N}$ we have defined $(\mathfrak{a}\mathfrak{v})_{(m)}^2 = \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\ell_\infty}^2$ and

$$\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := [(\mathfrak{a}\mathfrak{v})_{(m)}^2 \vee n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2], \quad m_n^\star := \arg\min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}_n^m(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}. \tag{21.06}$$

By Proposition §20.15 under Assumptions §19.02 and §20.11 the maximal global $\mathfrak{v}$-risk of an OPE $\widehat{f}^{m_n^\star}_\bullet$ with optimally choosen dimension $m_n^\star$ as in (21.06) satisfies

$$\mathcal{R}_n^{\mathfrak{v}}[\,\widehat{f}^{m_n^\star}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] \leqslant \mathrm{C}\,\mathrm{R}_n^\star(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$$

with $\mathrm{C} = \sigma_\xi^2 + \mathrm{r}^2\tau_{\mathfrak{a},\mathrm{u}}^2 + \mathrm{r}^2$. Furthermore, as in Notation §13.29 for $m \in \mathbb{N}$ we set $\mathcal{T}_m := \{-1,1\}^m$ and for each $\tau := (\tau_j)_{j\in[\![m]\!]} \in \mathcal{T}_m$ and $j \in [\![m]\!]$ we introduce $\tau^{(j)} \in \mathcal{T}_m$ given by $\tau_j^{(j)} := -\tau_j$ and $\tau_l^{(j)} := \tau_l$ for $l \in [\![m]\!]\backslash\{j\}$. □

§21.16 **Lemma** *(Assouad's cube technique).* *Given $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}_{\backslash 0}^+}$ if for each $\tau \in \mathcal{T}_m$ there is $f^\tau \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ with associated probability measure $\mathcal{U}_{f^\tau} := \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^\tau(X)}$ such that for all $\tau \in \mathcal{T}_m$ and $j \in [\![m]\!]$ we have $\mathrm{KL}(\mathcal{U}_{f^\tau}|\mathcal{U}_{f^{\tau^{(j)}}}) \leqslant 2n^{-1}$ then for all $n \geqslant 2$*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] \geqslant 2^{-m} \sum_{\tau\in\mathcal{T}_m} \tfrac{1}{64} \sum_{j\in[\![m]\!]} \big(\mathfrak{v}_j^2 |f_j^\tau - f_j^{\tau^{(j)}}|^2\big)$$

where the infimum is taken over all possible estimators.

§21.17 **Proof** of Lemma §21.16. is given in the lecture. □

§21.18 **Remark**. Assume candidate regression functions $f^\tau := \mathrm{U}^\star f^\tau_\bullet$ with $f^\tau_\bullet := (\tau_j f_j^\star \mathbb{1}_j^m)_{j\in\mathbb{N}}$, $\tau \in \mathcal{T}_m$, for some $f^\star_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$, where evidently $f^\tau_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ too, then trivially $\sum_{j\in[\![m]\!]} \big(\mathfrak{v}_j^2 |f_j^\tau - f_j^{\tau^{(j)}}|^2\big) = 4\|f^\star_\bullet \mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}$. If for all $\tau \in \mathcal{T}_m$ and $j \in [\![m]\!]$ the associated probability measures $\mathcal{U}_{f^\tau} = \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^\tau(X)}$ and $\mathcal{U}_{f^{\tau^{(j)}}} := \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^{\tau^{(j)}}(X)}$ satisfy $\mathrm{KL}(\mathcal{U}_{f^\tau}|\mathcal{U}_{f^{\tau^{(j)}}}) \leqslant 2n^{-1}$ then due to Lemma §21.16 for all $n \geqslant 2$ we have

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}_n^{\mathfrak{v}}[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] \geqslant 2^{-m} \sum_{\tau\in\mathcal{T}_m} \tfrac{1}{16}\|f^\star_\bullet \mathbb{1}^m_\bullet\|^2_{\mathfrak{v}} = \tfrac{1}{16}\|f^\star_\bullet \mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}. \tag{21.07}$$

We find a minimax-optimal lower bound by choosing the parameter m and the function $f^\star_\bullet$ that have the largest possible $\|f^\star_\bullet \mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}$-value although that the associated $\mathcal{U}_{f^\tau}^{\otimes n}$, $\tau \in \mathcal{T}_m$ are still statistically indistinguishable in the sense that $\mathrm{KL}(\mathcal{U}_{f^\tau}|\mathcal{U}_{f^{\tau^{(j)}}}) \leqslant 2n^{-1}$ for all $j \in [\![m]\!]$ and $\tau \in \mathcal{T}_m$. □

§21.19 **Lemma**. *Let $\mathbb{P}^\xi \in \mathcal{W}(\mathscr{B})$ satisfy Assumption §21.05 with constants $\mathrm{C}_\xi, x_\xi \in \mathbb{R}_{\backslash 0}^+$ and under Assumption §20.11 let $f^\star_\bullet \in \ell_2^{\mathfrak{a},\mathrm{r}}$ fulfill $\|f^\star_\bullet\|_{\mathfrak{a}^{-1}} \leqslant x_\xi/(2\tau_{\mathfrak{a},\mathrm{u}})$. For each $\tau \in \mathcal{T}_m$ introduce $f^\tau_\bullet := (\tau_j f_j^\star \mathbb{1}_j^m)_{j\in\mathbb{N}} \in \ell_2^{\mathfrak{a},\mathrm{r}}$ and $f^\tau := \mathrm{U}^\star f^\tau_\bullet \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ with associated probability measure $\mathcal{U}_{f^\tau} = \mathcal{U}_{[0,1]} \odot \mathbb{P}^\xi_{f^\tau(X)}$. Then for each $j \in [\![m]\!]$ we have $\mathrm{KL}(\mathcal{U}_{f^\tau}|\mathcal{U}_{f^{\tau^{(j)}}}) \leqslant 4\mathrm{C}_\xi \|f^\star_\bullet \mathbb{1}^m_\bullet\|^2_{\ell_\infty}$.*

§21.20 **Proof** of Lemma §21.19. is given in the lecture. □

§21.21 **Reminder**. For $w_\bullet \in \ell_\infty$ we set $w^2_{(0)} := \|w^2_\bullet\|_{\ell_\infty}$ and $w^2_{(\bullet)} = (w^2_{(j)} := \|w^2_\bullet \mathbb{1}^{j|\perp}_\bullet\|_{\ell_\infty})_{j\in\mathbb{N}}$ (Notation §13.34) where by construction $w^2_{(j)} = \sup\big\{w_i^2 \colon i \in \mathbb{N} \cap [j+1,\infty)\big\}$, $j \in \mathbb{N}_0$ and $w^2_{(\bullet)} \in (\mathbb{R}_{\backslash 0}^+)^{\mathbb{N}}_{\searrow}$. Under Assumption §20.11 let in addition $(\mathfrak{a}\mathfrak{v})^2_{(\bullet)} \in (\mathbb{R}_{\backslash 0}^+)^{\mathbb{N}}_{\downarrow 0}$ and there exists $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})} \in (0,1]$ such that $\mathrm{C}_{(\mathfrak{a}\mathfrak{v})}\|(\mathfrak{a}\mathfrak{v})^{-2}_\bullet \mathbb{1}^m_\bullet\|_{\ell_\infty} \leqslant (\mathfrak{a}\mathfrak{v})^{-2}_{(m-1)}$ or in equal

$$(\mathfrak{a}\mathfrak{v})^2_{(m-1)} \geqslant \min\big\{(\mathfrak{a}\mathfrak{v})^2_j \colon j \in [\![m]\!]\big\} \geqslant \mathrm{C}_{(\mathfrak{a}\mathfrak{v})}(\mathfrak{a}\mathfrak{v})^2_{(m-1)}$$

for all $m \in \mathbb{N}$, then Assumption §13.35 is satisfied. For $m_n^\star$ and $\mathrm{R}_n^\star := \mathrm{R}_n^{m_n^\star}(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ as in (21.06) we distinguish case i) : $\mathrm{R}_n^\star = n^{-1}\|\mathbb{1}^{m_n^\star}_\bullet\|^2_{\mathfrak{v}} > (\mathfrak{a}\mathfrak{v})^2_{(m_n^\star)}$ and case ii) : $\mathrm{R}_n^\star = (\mathfrak{a}\mathfrak{v})^2_{(m_n^\star)} \geqslant n^{-1}\|\mathbb{1}^{m_n^\star}_\bullet\|^2_{\mathfrak{v}}$. Due to Comment §13.36 if $(\mathfrak{a}\mathfrak{v})^2_{(1)} > n^{-1}\mathfrak{v}_1^2$ then in case i) we obtain $(\mathfrak{a}\mathfrak{v})^2_{(m_n^\star-1)} \geqslant n^{-1}\|\mathbb{1}^{m_n^\star}_\bullet\|^2_{\mathfrak{v}}$, while in case ii) setting (the defining set is not empty since $(\mathfrak{a}\mathfrak{v})^2_{(\bullet)} \in (\mathbb{R}_{\backslash 0}^+)^{\mathbb{N}}_{\downarrow 0}$)

$$m_n^\diamond := \min\big\{m \in \mathbb{N} \cap [m_n^\star + 1, \infty) \colon n^{-1}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}} \geqslant (\mathfrak{a}\mathfrak{v})^2_{(m)}\big\} \tag{21.08}$$

we have $(\mathfrak{a}\mathfrak{v})^2_{(m_n^\star)} = (\mathfrak{a}\mathfrak{v})^2_{(m_n^\diamond-1)} \leqslant n^{-1}\|\mathbb{1}^{m_n^\diamond}_\bullet\|^2_{\mathfrak{v}}$. We use those estimates in the next proof. □

§21.22 **Proposition** *(Lower bound).* *Let* $\mathbb{P}^\xi \in \mathcal{W}(\mathscr{B})$ *satisfy Assumption §21.05 with constants* $\mathrm{C}_\xi, x_\xi \in \mathbb{R}^+_{\setminus 0}$ *and let Assumptions §19.02 and §20.11 be fulfilled. If* $(\mathfrak{av})^2_{(\bullet)} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$ *and there exists* $\mathrm{C}_{(\mathfrak{av})} \in (0,1]$ *such that* $\mathrm{C}_{(\mathfrak{av})} \|(\mathfrak{av})^{-2}_\bullet \mathbb{1}^m_\bullet\|_{\ell_\infty} \leqslant (\mathfrak{av})^{-2}_{(m-1)}$ *for all* $m \in \mathbb{N}$*, then for all* $n \in \mathbb{N} \cap (1 \vee \mathfrak{v}_1^2(\mathfrak{av})^{-2}_{(1)}, \infty)$ *we have*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}^{\mathfrak{v}}_n[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathbb{P}^\xi\}\,] \geqslant \mathrm{C}\ \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \tag{21.09}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{C}_{(\mathfrak{av})} x_\xi^2/(4\tau^2_{\mathfrak{a},\mathrm{u}}) \wedge \mathrm{C}_{(\mathfrak{av})}\mathrm{r}^2 \wedge 1/(2\mathrm{C}_\xi))$ *and infimum taken over all estimators.*

§21.23 **Proof** of Proposition §21.22. is given in the lecture. □

§21.24 **Comment**. If $\xi \sim \mathrm{N}_{(0,\sigma_\xi^2)}$ with $\sigma_\xi^2 \in \mathbb{R}^+_{\setminus 0}$ then Assumption §21.05 holds with $\mathrm{C}_\xi = 1/(2\sigma_\xi^2)$ and $x_\xi^2 = \infty$ (see Comment §21.11). Consequently, from Proposition §21.22 we obtain immediately

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}^{\mathfrak{v}}_n[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \{\mathrm{N}_{(0,\sigma_\xi^2)}\}\,] \geqslant \mathrm{C}\ \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) \tag{21.10}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{C}_{(\mathfrak{av})}\mathrm{r}^2 \wedge \sigma_\xi^2)$ and infimum taken over all estimators. □

§21.25 **Corollary** *(Lower bound).* *Let Assumptions §19.02 and §20.11 be fulfilled and let* $\sigma^2 \in \mathbb{R}^+_{\setminus 0}$*. If* $(\mathfrak{av})^2_{(\bullet)} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$ *and there exists* $\mathrm{C}_{(\mathfrak{av})} \in (0,1]$ *such that* $\mathrm{C}_{(\mathfrak{av})} \|(\mathfrak{av})^{-2}_\bullet \mathbb{1}^m_\bullet\|_{\ell_\infty} \leqslant (\mathfrak{av})^{-2}_{(m-1)}$ *for all* $m \in \mathbb{N}$*, then for all* $n \in \mathbb{N} \cap (1 \vee \mathfrak{v}_1^2(\mathfrak{av})^{-2}_{(1)}, \infty)$ *we have*

$$\inf_{\widetilde{f}_\bullet} \mathcal{R}^{\phi}_n[\,\widetilde{f}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}, \mathrm{P}_{\{0\}\times(0,\sigma^2]}\,] \geqslant \mathrm{C}\ \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \phi_\bullet) \tag{21.11}$$

with constant $\mathrm{C} := 16^{-1}(\mathrm{C}_{(\mathfrak{av})}\mathrm{r}^2 \wedge \sigma^2)$ *and infimum taken over all estimators.*

§21.26 **Proof** of Corollary §21.25. is given in the lecture. □

§21.27 **Illustration**. Consider the *trigonometric basis* as in Illustration §20.17 which satisfies Assumption §20.11 for all $\mathfrak{a}_\bullet \in \ell_2$ (see Illustration §20.17). In Table 02 [§12] the order of the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §12.26. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}^+_{\setminus 0}$. In both cases the additional assumptions, $(\mathfrak{av})^2_{(\bullet)} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}_{\downarrow 0}$ and there exists $\mathrm{C}_{(\mathfrak{av})} \in (0,1]$ such that $\mathrm{C}_{(\mathfrak{av})} \|(\mathfrak{av})^{-2}_\bullet \mathbb{1}^m_\bullet\|_{\ell_\infty} \leqslant (\mathfrak{av})^{-2}_{(m-1)}$ for all $m \in \mathbb{N}$, are satisfied. Consequently, due to Proposition §21.22 the Table 02 [§12] presents the order of the *minimax rate* $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ which is attaind by the *minimax-optimal* OPE $\widehat{f}^{m^\star_n}_\bullet = \widehat{f}_\bullet \mathbb{1}^{m^\star_n}_\bullet \in \ell_2 \mathbb{1}^{m^\star_n}_\bullet \subseteq \ell_2(\mathfrak{v}^2_\bullet)$ with optimally selected dimension $m^\star_n$ (Proposition §20.15). We shall stress, that the order of $m^\star_n$ given in the Table 02 [§12] depends on the parameter $a \in \mathbb{R}^+_{\setminus 0}$ characterising the (abstract) smoothness of the regression function of interest which is generally not known in advance. □

§22 Data-driven regression

§22|01 Data-driven global estimation by model selection

§22.01 **Reminder**. Talagrand's inequality stated in the form of Lemma §18.01 provides again our key argument in order to control the deviations of the reminder term. Let us briefly recall how we intend to apply Talagrand's inequality (see Remark §18.02 for a similar approach). Reconsider the stochastic process $\psi_\bullet = (\psi_j(X,Y) := Y\mathrm{u}_j(X))_{j\in\mathbb{N}} \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B}) \otimes 2^{\mathbb{N}}$ where $\psi_j \in \mathcal{L}_1(\mathcal{U}_f)$ for each $j \in \mathbb{N}$ and the OPE $\widehat{f}^m_\bullet = \widehat{f}_\bullet \mathbb{1}^m_\bullet \in \ell_2 \mathbb{1}^m_\bullet$ with dimension $m \in \mathbb{N}$ (Definition §20.04). $\widehat{f}_\bullet = \widehat{\mathbb{P}}_n\psi_\bullet = (\widehat{\mathbb{P}}_n\psi_j)_{j\in\mathbb{N}}$ are noisy versions (Definition §15.08) of the regression coefficients $f_\bullet = \mathrm{U}f = \mathcal{U}_f(\psi_\bullet) = (\mathcal{U}_f\psi_j = \mathcal{U}_f(Y\mathrm{u}_j(X)))_{j\in\mathbb{N}}$ (see Notation §19.07). For $m \in \mathbb{N}$ introduce

the unit ball $\mathbb{B}_m := \{a_\bullet \in \ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet : \|a_\bullet\|_\mathfrak{v} \leqslant 1\}$ contained in the linear subspace $\ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet$ spanned by $(\mathbb{1}^{\{j\}}_\bullet)_{j\in[\![m]\!]}$. Clearly, for each $a_\bullet \in \ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet$ we have $\mathrm{r}_{a_\bullet} := \sum_{j\in[\![m]\!]} \mathfrak{v}^2_j a_j \psi_j = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \psi_\bullet) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$, i.e. it is a $\mathscr{B}_{[0,1]} \otimes \mathscr{B}$-$\mathscr{B}$-measurable function, where $\widehat{\mathbb{P}}_n(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \widehat{\mathbb{P}}_n \psi_\bullet) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \widehat{f}_\bullet)$, $\mathcal{U}_f(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \mathcal{U}_f \psi_\bullet) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet f_\bullet)$ and hence $\bar{\mathrm{r}}_{a_\bullet} = \widehat{\mathbb{P}}_n(\mathrm{r}_{a_\bullet}) - \mathcal{U}_f(\mathrm{r}_{a_\bullet}) = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet(\widehat{f}_\bullet - f_\bullet)) = \langle \widehat{f}_\bullet - f_\bullet, a_\bullet\rangle_\mathfrak{v}$. Let $\mathcal{B}_m$ be a countable dense subset of the unit ball $\mathbb{B}_m$ (see Remark §18.02 for more details), then we obtain

$$\|\widehat{f}^m_\bullet - f^m_\bullet\|^2_\mathfrak{v} = \sup\{|\langle \widehat{f}_\bullet - f_\bullet, a_\bullet\rangle_\mathfrak{v}|^2 : a_\bullet \in \mathcal{B}_m\} = \sup\{|\bar{\mathrm{r}}_{a_\bullet}|^2 : a_\bullet \in \mathcal{B}_m\}.$$

The last identity provides the necessary argument to apply below Talagrand's inequality (§18.01) where we need to calculate the three constants h, H and v. We note that $\psi_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B}) \otimes 2^{\mathbb{N}}$ and thus $\mathrm{r}_{a_\bullet} = \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \psi_\bullet) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ is not bounded. Therefore, we decompose $\psi_\bullet = \psi^{\mathrm{b}}_\bullet + \psi^{\mathrm{u}}_\bullet$ into two parts $\psi^{\mathrm{b}}_\bullet, \psi^{\mathrm{u}}_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B}) \otimes 2^{\mathbb{N}}$, a bounded and a remaining unbounded one. To be more precise, for $a_\bullet \in \ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^m_\bullet$ setting $\mathrm{r}^{\mathrm{b}}_{a_\bullet} := \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \psi^{\mathrm{b}}_\bullet) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ and $\mathrm{r}^{\mathrm{u}}_{a_\bullet} := \nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet \psi^{\mathrm{u}}_\bullet) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ let $\sup\{|\mathrm{r}^{\mathrm{b}}_{a_\bullet}(x,y)| : a_\bullet \in \mathcal{B}_d, x \in [0,1], y \in \mathbb{R}\} \in \mathbb{R}$ be satisfied. Introducing further $\bar{\mathrm{r}}^{\mathrm{b}}_{a_\bullet} := \widehat{\mathbb{P}}_n(\mathrm{r}^{\mathrm{b}}_{a_\bullet}) - \mathcal{U}_f(\mathrm{r}^{\mathrm{b}}_{a_\bullet})$ and $\bar{\mathrm{r}}^{\mathrm{u}}_{a_\bullet} := \widehat{\mathbb{P}}_n(\mathrm{r}^{\mathrm{u}}_{a_\bullet}) - \mathcal{U}_f(\mathrm{r}^{\mathrm{u}}_{a_\bullet})$ we evidently have

$$\begin{aligned}\|\widehat{f}^m_\bullet - f^m_\bullet\|^2_\mathfrak{v} &= \sup\{|\bar{\mathrm{r}}^{\mathrm{b}}_{a_\bullet} + \bar{\mathrm{r}}^{\mathrm{u}}_{a_\bullet}|^2 : a_\bullet \in \mathcal{B}_m\}\\ &\leqslant 2\sup\{|\bar{\mathrm{r}}^{\mathrm{b}}_{a_\bullet}|^2 : a_\bullet \in \mathcal{B}_m\} + 2\sup\{|\bar{\mathrm{r}}^{\mathrm{u}}_{a_\bullet}|^2 : a_\bullet \in \mathcal{B}_m\}\\ &= 2\|(\widehat{\mathbb{P}}_n\psi^{\mathrm{b}}_\bullet - \mathcal{U}_f\psi^{\mathrm{b}}_\bullet)\mathbb{1}^m_\bullet\|^2_\mathfrak{v} + 2\|(\widehat{\mathbb{P}}_n\psi^{\mathrm{u}}_\bullet - \mathcal{U}_f\psi^{\mathrm{u}}_\bullet)\mathbb{1}^m_\bullet\|^2_\mathfrak{v} \qquad (22.01)\end{aligned}$$

Considering the first term on the right hand side provided that

$$\begin{aligned}&\sup\{\|\psi^{\mathrm{b}}_\bullet(x,y)\mathbb{1}^m_\bullet\|_\mathfrak{v} : x \in [0,1], y \in \mathbb{R}\} = \sup\{|\mathrm{r}^{\mathrm{b}}_{a_\bullet}(x,y)| : a_\bullet \in \mathcal{B}_d, x \in [0,1], y \in \mathbb{R}\} \leqslant \mathrm{h}_m,\\ &\mathcal{U}_f^{\otimes n}\big(\|(\widehat{\mathbb{P}}_n\psi^{\mathrm{b}}_\bullet - \mathcal{U}_f\psi^{\mathrm{b}}_\bullet)\mathbb{1}^m_\bullet\|^2_\mathfrak{v}\big) = \mathcal{U}_f^{\otimes n}\big(\sup\{|\bar{\mathrm{r}}^{\mathrm{b}}_{a_\bullet}|^2 : a_\bullet \in \mathcal{B}_d\}\big) \leqslant \mathrm{H}^2_m,\\ &\sup\{\mathcal{U}_f\big(|\nu_{\mathbb{N}}(\mathfrak{v}^2_\bullet a_\bullet(\psi^{\mathrm{b}}_\bullet - \mathcal{U}_f\psi^{\mathrm{b}}_\bullet))|^2\big) : a_\bullet \in \mathcal{B}_d\} = \sup\{n\mathcal{U}_f^{\otimes n}\big(|\bar{\mathrm{r}}^{\mathrm{b}}_{a_\bullet}|^2\big) : a_\bullet \in \mathcal{B}_d\} \leqslant \mathrm{v}_m \qquad (22.02)\end{aligned}$$

we eventually apply Talagrand's inequality (§18.01) and we obtain

$$\mathcal{U}_f^{\otimes n}\big(\big(\|(\widehat{\mathbb{P}}_n\psi^{\mathrm{b}}_\bullet - \mathcal{U}_f\psi^{\mathrm{b}}_\bullet)\mathbb{1}^m_\bullet\|^2_\mathfrak{v} - 6\mathrm{H}^2_m\big)_+\big) \leqslant \mathrm{C}_{\mathrm{tal}}\left\{\frac{\mathrm{v}_m}{n}\exp\left(\frac{-n\mathrm{H}^2_m}{6\mathrm{v}_m}\right) + \frac{\mathrm{h}^2_m}{n^2}\exp\left(\frac{-n\mathrm{H}_m}{100\mathrm{h}_m}\right)\right\} \qquad (22.03)$$

for some universal numerical constant $\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$. □

§22|01|01 Global $\mathfrak{v}$-risk

§22.02 **Assumption**. The weights $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ satisfy

$$\forall x \in \mathbb{R}^+_{\setminus 0}: \quad \sum_{m\in\mathbb{N}}\{x\|\mathfrak{v}^2_\bullet\mathbb{1}^m_\bullet\|_{\ell_\infty}\exp\big(-\|\mathfrak{v}_\bullet\mathbb{1}^m_\bullet\|^2_{\ell_2}/(x\|\mathfrak{v}^2_\bullet\mathbb{1}^m_\bullet\|_{\ell_\infty})\big)\} =: \mathrm{C}_\mathfrak{v}(x) \in \mathbb{R}^+. \qquad (22.04)$$

The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete*, i.e an *orthonormal basis* in $\mathbb{L}_2(\lambda_{[0,1]})$ and satisfies as process $\mathrm{u}^2_\bullet = (\mathrm{u}^2_j)_{j\in\mathbb{N}}$ on $([0,1],\mathscr{B}_{[0,1]})$ for $\tau_{\mathfrak{v},\mathrm{u}} \in [1,\infty)$ and for all $m \in \mathbb{N}$ **(os3)** $\sup\{\|\mathrm{u}_\bullet(x)\mathbb{1}^m_\bullet\|^2_\mathfrak{v} : x \in [0,1]\} \leqslant \tau^2_{\mathfrak{v},\mathrm{u}}\|\mathbb{1}^m_\bullet\|^2_\mathfrak{v} \in \mathbb{R}^+$. □

§22.03 **Remark**. Under Assumption §22.02 (18.05) we have $\|\mathfrak{v}_\bullet\mathbb{1}^m_\bullet\|^{-2}_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$ (Comment §14.22), see also Illustration §14.23 for an example when (18.05) is not satisfied. □

§22.04 **Reminder** *(Global oracle $\mathfrak{v}$-risk)*. Given Assumptions §19.02 and §22.02 we consider an OPE as in Definition §20.04. Here the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression

coefficients $f_\bullet = \mathrm{U} f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §22.02, (and hence Assumption §19.05 and $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$) and $f_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ in §20.09 an *oracle inequality* for the global $\mathfrak{v}$-risk of the OPE's is shown. More presicely, as in (20.02) (Proposition §20.07) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) := \|f_\bullet \mathbb{1}_\bullet^{m|\perp}\|_{\mathfrak{v}}^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad m_n^\circ := \arg\min\{\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}_n^{m_n^\circ}(f_\bullet, \mathfrak{v}_\bullet) = \min\{\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\}. \quad (22.05)$$

and assuming $\mathbb{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ due to Property §20.09 the (infeasible) OPE $\widehat{f}_\bullet^{m_n^\circ} = \widehat{f}_\bullet \mathbb{1}_\bullet^{m_n^\circ} \in \ell_2(\mathfrak{v}_\bullet^2)\mathbb{1}_\bullet^{m_n^\circ} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ with oracle dimension m_n° as in (22.05) satisfies

$$\mathbb{v}_f^{-1}\,\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2) \leqslant \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{m_n^\circ} - f_\bullet\|_{\mathfrak{v}}^2)$$
$$\leqslant \mathbb{v}_f\,\mathrm{R}_n^\circ(f_\bullet, \mathfrak{v}_\bullet) \leqslant \mathbb{v}_f^2 \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^m - f_\bullet\|_{\mathfrak{v}}^2),$$

and hence it is *oracle optimal* (with constant $\mathbb{v}_f^2$). □

Partially known penalty sequence

§22.05 **Notation**. Consider a sequence of penalties $\mathfrak{pen}_\bullet^{f,\mathfrak{v}} = (\mathfrak{pen}_m^{f,\mathfrak{v}})_{m\in\mathbb{N}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given by

$$\mathfrak{pen}_m^{f,\mathfrak{v}} := 48\mathrm{v}_f^2\tau_{\mathfrak{v},\mathrm{u}}^2 n^{-1}\|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2, \quad \text{for each } m \in \mathbb{N} \text{ with } \mathrm{v}_f^2 := 1 + \mathcal{U}_f(Y^2) \quad (22.06)$$

which is obviously only *partially known* in advance, and the in advance known upper bound (where the defining set is not empty)

$$\mathrm{M}^{\mathfrak{v}} := \max\{m \in \mathbb{N}\colon \|\mathbb{1}_\bullet^m\|_{\mathfrak{v}}^2 \leqslant n\mathfrak{v}_1^2,\ m \leqslant n^{-2/3}\exp(\tfrac{n^{1/6}}{100})\}. \quad (22.07)$$

Considering the partially data-driven OSE $\widehat{f}_\bullet^{\widehat{m}} = \widehat{f}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter

$$\widehat{m} := \arg\min\{-\|\widehat{f}_\bullet^m\|_{\mathfrak{v}} + \mathfrak{pen}_m^{f,\mathfrak{v}}\colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\} \quad (22.08)$$

we derive below an upper bound for its global $\mathfrak{v}_\bullet$-risk, $\mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{\widehat{m}} - f_\bullet\|_{\mathfrak{v}}^2)$. □

§22.06 **Lemma**. *Under Assumptions §19.02 and §22.02, $Y \in \mathcal{L}_5(\mathcal{U}_f)$ and $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ for $\mathfrak{pen}_\bullet^{f,\mathfrak{v}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ as in (22.06) and $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ as in (22.07) we have*

$$\mathcal{U}_f^{\otimes n}\big(\max\{(\|\widehat{f}_\bullet^m - f_\bullet^m\|_{\mathfrak{v}}^2 - \mathfrak{pen}_m^{f,\mathfrak{v}}/4)_+ \colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\}\big)$$
$$\leqslant 2\mathrm{C}_{\mathrm{tal}}\tau_{\mathfrak{v},\mathrm{u}}^2(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f}) + \mathfrak{v}_1^2)(1 + \mathcal{U}_f(Y^2) + \mathcal{U}_f(|Y|^5))\, n^{-1} \quad (22.09)$$

for some universal numerical constant $\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ and $x_{\xi,f} = 6(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})/(\mathrm{v}_f^2\tau_{\mathfrak{v},\mathrm{u}}^2) \in \mathbb{R}^+$.

§22.07 **Proof** of Lemma §22.06. is given in the lecture. □

§22.08 **Proposition** *(Upper bound)*. *Under Assumptions §19.02 and §22.02, $Y \in \mathcal{L}_5(\mathcal{U}_f)$ and $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ for $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ as in (22.07) and $\mathfrak{pen}_\bullet^{f,\mathfrak{v}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ as in (22.06) the partially data-driven OPE $\widehat{f}_\bullet^{\widehat{m}} = \widehat{f}_\bullet \mathbb{1}_\bullet^{\widehat{m}} \in \ell_2(\mathfrak{v}_\bullet^2)\mathbb{1}_\bullet^{\widehat{m}} \subseteq \ell_2(\mathfrak{v}_\bullet^2)$ of $f_\bullet \in \ell_2(\mathfrak{v}_\bullet^2)$ with data-driven dimension $\widehat{m} \in [\![\mathrm{M}^{\mathfrak{v}}]\!]$ as in (22.08) satisfies*

$$\mathcal{U}_f^{\otimes n}(\|\widehat{f}_\bullet^{\widehat{m}} - f_\bullet\|_{\mathfrak{v}}^2) \leqslant 192(1 + \mathcal{U}_f(Y^2))\tau_{\mathfrak{v},\mathrm{u}}^2 \min\{\mathrm{R}_n^m(f_\bullet, \mathfrak{v}_\bullet)\colon m \in [\![\mathrm{M}^{\mathfrak{v}}]\!]\}$$
$$+ \mathrm{C}\tau_{\mathfrak{v},\mathrm{u}}^2(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f}) + \mathfrak{v}_1^2)(1 + \mathcal{U}_f(Y^2) + \mathcal{U}_f(|Y|^5))\, n^{-1} \quad (22.10)$$

for some universal numerical constant $\mathrm{C} = 16\mathrm{C}_{\mathrm{tal}} \in [1,\infty)$ and $x_{\xi,f} = 6(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})\mathrm{v}_f^{-2}\tau_{\mathfrak{v},\mathrm{u}}^{-2} \in \mathbb{R}^+$.

§22.09 **Proof** of Proposition §22.08. is given in the lecture. □

§22.10 **Comment**. The oracle bound $\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet}) = \mathrm{R}^{m^{\circ}_{n}}_{n}(f_{\bullet},\mathfrak{v}_{\bullet}) = \min\{\mathrm{R}^{m}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})\colon m\in\mathbb{N}\}$ (for details see Reminder §22.04) satisfies $n\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet}) \geqslant \|\mathbb{1}^{m^{\circ}_{n}}_{\bullet}\|^{2}_{\mathfrak{v}} \geqslant \mathfrak{v}_{1}^{2}$. Consequently, the last upper bound in (22.10) and the oracle bound $\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})$ coincide up to a constant $(192(1+\mathcal{U}_{f}(Y^{2}))\tau^{2}_{\mathfrak{v},\mathrm{u}} + \mathrm{C}\tau^{2}_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f})\mathfrak{v}_{1}^{-2}+1)(1+\mathcal{U}_{f}(Y^{2})+\mathcal{U}_{f}(|Y|^{5})))$ provided the oracle dimension fulfils $m^{\circ}_{n}\in[\![\mathrm{M}^{\mathfrak{v}}]\!]$. Therefore, we wish the upper bound $\mathrm{M}^{\mathfrak{v}}$ to be as large as possible. The next assertion shows that $\mathrm{M}^{\mathfrak{v}}$ as in (22.07) is a suitable choice for the upper bound. □

§22.11 **Corollary**. *Under the assumptions of Proposition §22.08 for each $n\in\mathbb{N}$ such that $\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})\leqslant\mathfrak{v}_{1}^{2}$ and $m^{\circ}_{n}\leqslant n^{-2/3}\exp(\frac{n^{1/6}}{100})$ we have*

$$\mathcal{U}_{f}^{\otimes n}\big(\|\widehat{f}^{\widehat{m}}_{\bullet}-f_{\bullet}\|^{2}_{\mathfrak{v}}\big)\leqslant\mathrm{K}\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})$$

and, hence up to the constant $\mathrm{K}=32(\mathrm{C}_{\mathrm{tal}}+12)\tau^{2}_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f})\mathfrak{v}_{1}^{-2}+1)(1+\mathcal{U}_{f}(|Y|^{5}))$ the infeasible partially data-driven estimator $\widehat{f}^{\widehat{m}}_{\bullet}$ is oracle optimal.

§22.12 **Proof** of Corollary §22.11. is given in the lecture. □

§22.13 **Illustration**. Consider the *trigonometric basis* as in Illustration §20.17 which satisfies Assumption §22.02 **(os1)**, **(os3)** for all $\mathfrak{a}_{\bullet}\in\ell_{2}$. In Table 01 [§12] (Illustration §12.19) the order of the rate $\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})$ is depict for the two specifications **(o)** and **(s)**. We note that we have $\mathfrak{a}_{\bullet}\in\ell_{2}$ in case **(o)** for $a>1/2$ while in case **(s)** for $a\in\mathbb{R}^{+}_{\backslash 0}$. The sequence $\mathfrak{v}_{\bullet}$ satisfies Assumption §22.02, i.e. (22.04), for $v\geqslant -1/2$. Moreover, the optimal dimension m°_{n} given in Table 01 [§12] satisfies $m^{\circ}_{n}\leqslant n^{-2/3}\exp(\frac{n^{1/6}}{100})$, and thus (under the above restrictions) the partially data-driven (hence not feasible) density estimator attains the oracle rate $\mathrm{R}^{\circ}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})$ up to the constant given in Corollary §22.11. □

Estimated penalty sequence

§22.14 **Notation**. The penalty sequence $\mathfrak{pen}^{f,\mathfrak{v}}_{\bullet}\in(\mathbb{R}^{+}_{\backslash 0})^{\mathbb{N}}$ given in (22.06) still depends amongst others on characteristics of the unknown regression function f. More precisely, for $m\in\mathbb{N}$ the term $\mathfrak{pen}^{f,\mathfrak{v}}_{m}$ involves the quantity $\mathrm{v}_{f}^{2}=1+\mathcal{U}_{f}(Y^{2})$ which we eventually estimate without bias by $\widehat{\mathrm{v}}^{2}:=1+\widehat{\mathbb{P}}_{n}(Y^{2})$ (keeping in mind that we identify Y and the coordinate map Π_{R}). Therewith, let us introduce a fully data-driven sequence of penalties $\widehat{\mathfrak{pen}}^{\mathfrak{v}}_{\bullet}=(\widehat{\mathfrak{pen}}^{\mathfrak{v}}_{m})_{m\in\mathbb{N}}\in(\mathbb{R}^{+}_{\backslash 0})^{\mathbb{N}}$ given by

$$\widehat{\mathfrak{pen}}^{\mathfrak{v}}_{m}:=2\times 48\widehat{\mathrm{v}}^{2}\tau^{2}_{\mathfrak{v},\mathrm{u}}n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^{2}_{\mathfrak{v}}\quad\text{for each } m\in\mathbb{N}\ \text{ with }\ \widehat{\mathrm{v}}^{2}:=1+\widehat{\mathbb{P}}_{n}\big(Y^{2}\big)\tag{22.11}$$

and the upper bound $\mathrm{M}^{\mathfrak{v}}\in\mathbb{N}$ given in (22.07) which are both *fully known* in advance. Considering the data-driven OSE $\widehat{f}^{\widehat{m}}_{\bullet}=\widehat{f}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet}$ with dimension parameter selected by

$$\widehat{m}:=\arg\min\big\{-\|\widehat{f}^{m}_{\bullet}\|_{\mathfrak{v}}+\widehat{\mathfrak{pen}}^{\mathfrak{v}}_{m}\colon m\in[\![\mathrm{M}^{\mathfrak{v}}]\!]\big\}\tag{22.12}$$

we derive below an upper bound for its global $\mathfrak{v}$-risk, $\mathcal{U}_{f}^{\otimes n}\big(\|\widehat{f}^{\widehat{m}}_{\bullet}-f_{\bullet}\|^{2}_{\mathfrak{v}}\big)$. □

§22.15 **Proposition** *(Upper bound)*. *Under Assumptions §19.02 and §22.02, $Y\in\mathcal{L}_{5}(\mathcal{U}_{f})$ and $f\in\mathbb{L}_{\infty}(\lambda_{[0,1]})$ for $\mathrm{M}^{\mathfrak{v}}\in\mathbb{N}$ as in (22.07) and $\widehat{\mathfrak{pen}}^{\mathfrak{v}}_{m}\in(\mathbb{R}^{+}_{\backslash 0})^{\mathbb{N}}$ as in (22.11) the fully data-driven OPE $\widehat{f}^{\widehat{m}}_{\bullet}=\widehat{f}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet}\in\ell_{2}(\mathfrak{v}_{\bullet}^{2})\mathbb{1}^{\widehat{m}}_{\bullet}\subseteq\ell_{2}(\mathfrak{v}_{\bullet}^{2})$ of $f_{\bullet}\in\ell_{2}(\mathfrak{v}_{\bullet}^{2})$ with data-driven dimension $\widehat{m}\in[\![\mathrm{M}^{\mathfrak{v}}]\!]$ as in (22.08) satisfies*

$$\begin{aligned}\mathcal{U}_{f}^{\otimes n}\big(\|\widehat{f}^{\widehat{m}}_{\bullet}-f_{\bullet}\|^{2}_{\mathfrak{v}}\big)\leqslant 288\tau^{2}_{\mathfrak{v},\mathrm{u}}(1+\mathcal{U}_{f}(Y^{2}))\min\big\{\mathrm{R}^{m}_{n}(f_{\bullet},\mathfrak{v}_{\bullet})\colon m\in[\![\mathrm{M}^{\mathfrak{v}}]\!]\big\}\\ +\mathrm{C}\tau^{2}_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f})+\mathfrak{v}_{1}^{2})(1+\mathcal{U}_{f}(|Y|^{5}))\,n^{-1}\end{aligned}\tag{22.13}$$

for $x_{\xi,f} = 6(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})\mathrm{v}_f^{-2}\tau_{\mathfrak{v},\mathrm{u}}^{-2} \in \mathbb{R}^+$ *and some universal numerical constant* $\mathrm{C} = 3(16\mathrm{C}_{\mathrm{tal}} + 384) \in [1,\infty)$.

§22.16 **Proof** of Proposition §22.15. is given in the lecture. □

§22.17 **Comment**. We shall stress that the last upper bound (22.13) in Proposition §22.15 (for the fully data-driven procedure) and the upper bound (22.10) in Proposition §22.08 (for the partially data-driven procedure) differ only in the constants. Thus, Comment §22.10 still applies here and the proof of the next results follows line by line their counterparts above. □

§22.18 **Corollary**. *Under the assumptions of Proposition §22.15 for each* $n \in \mathbb{N}$ *such that* $\mathrm{R}^\circ_n(f_\bullet, \mathfrak{v}_\bullet) \leqslant \mathfrak{v}_1^2$ *and* $m^\circ_n \leqslant n^{-2/3}\exp(\frac{n^{1/6}}{100})$ *we have*

$$\mathcal{U}_f^{\otimes n}\big(\|\widehat{f}^{\widehat{m}}_\bullet - f_\bullet\|^2_{\mathfrak{v}}\big) \leqslant \mathrm{K}\mathrm{R}^\circ_n(f_\bullet, \mathfrak{v}_\bullet)$$

and, hence up to the constant $\mathrm{K} = 5(16\mathrm{C}_{\mathrm{tal}} + 384)\tau^2_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_{\xi,f})\mathfrak{v}_1^{-2} + 1)(1 + \mathcal{U}_f(|Y|^5))$ *the feasible fully data-driven estimator* $\widehat{f}^{\widehat{m}}_\bullet$ *is oracle optimal.*

§22.19 **Proof** of Corollary §22.18. is given in the lecture. □

§22|01|02 Maximal global $\mathfrak{v}$-risk

§22.20 **Assumption**. Consider weights $\mathfrak{a}_\bullet, \mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\mathfrak{v})_\bullet := (\mathfrak{a}_j\mathfrak{v}_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\mathfrak{v}_\bullet \in \ell_\infty$. We write $(\mathfrak{a}\mathfrak{v})_{(m)} := \|(\mathfrak{a}\mathfrak{v})_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\ell_\infty} \in \mathbb{R}^+$ for each $m \in \mathbb{N}$. The weights $\mathfrak{v}_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$ satisfy (18.05). The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete*, i.e an *orthonormal basis* in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}^2_\bullet = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ satisfies **(os2)** $\|\nu_{\mathbb{N}}(\mathfrak{a}^2_\bullet \mathrm{u}^2_\bullet)\|_{\mathbb{L}_2(\lambda_{[0,1]})} \leqslant \tau^2_{\mathfrak{a},\mathrm{u}}$ and for all $m \in \mathbb{N}$, **(os3)** $\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}_\bullet} : x \in [0,1]\big\} \leqslant \tau^2_{\mathfrak{v},\mathrm{u}}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}_\bullet} \in \mathbb{R}^+$. for $\tau_{\mathfrak{a},\mathrm{u}}, \tau_{\mathfrak{v},\mathrm{u}} \in [1,\infty)$. □

§22.21 **Reminder** *(Maximal global $\mathfrak{v}$-risk).* Given Assumptions §19.02 and §22.20 we consider an OPE as in Definition §20.04. Here the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U}f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §22.20 in Proposition §16.16 an upper bound for a maximal global $\mathfrak{v}$-risk of an OPE is shown over the set $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ given in (20.04) (Lemma §20.13). More precisely, the performance of the OPE $\widehat{f}^m_\bullet = \widehat{f}_\bullet\mathbb{1}^m_\bullet \in \ell_2(\mathfrak{v}^2)\mathbb{1}^m_\bullet \subseteq \ell_2(\mathfrak{v}^2_\bullet)$ with dimension $m \in \mathbb{N}$ is measured by its maximal global $\mathfrak{v}$-risk over the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}^{\mathfrak{v}}_n[\,\widehat{f}^m_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathcal{U}_f^{\otimes n}(\|\widehat{f}^m_\bullet - f_\bullet\|^2_{\mathfrak{v}}) : f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

As in (12.06) (Proposition §12.21) for $n, m \in \mathbb{N}$ setting $(\mathfrak{a}\mathfrak{v})^2_{(m)} := \|(\mathfrak{a}\mathfrak{v})^2_\bullet \mathbb{1}^{m|\perp}_\bullet\|_{\ell_\infty}$ and

$$\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := (\mathfrak{a}\mathfrak{v})^2_{(m)} \vee n^{-1}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}}, \quad m^\star_n := \arg\min\big\{\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) := \mathrm{R}^{m^\star_n}_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) = \min\big\{\mathrm{R}^m_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet) : m \in \mathbb{N}\big\} \quad (22.14)$$

by Proposition §20.15 under Assumptions §19.02 and §22.02 the maximal global $\mathfrak{v}$-risk of an OPE $\widehat{f}^{\widehat{m}^\star_n}_\bullet$ with optimally choosen dimension $m^\star_n$ as in (22.14) satisfies

$$\mathcal{R}^{\mathfrak{v}}_n[\,\widehat{f}^{\widehat{m}^\star_n}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$$

with $\mathrm{C} = \sigma_\xi^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + \mathrm{r}^2$. Moreover, due to Proposition §21.22 $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ provides (up to a constant) also a lower bound of the maximal global $\mathfrak{v}$-risk over the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}^\star_n(\mathfrak{a}_\bullet, \mathfrak{v}_\bullet)$ is a minimax bound and $\widehat{f}^{\widehat{m}^\star_n}_\bullet$ is minimax optimal. However, the optimal dimension $m^\star_n$ depends on $\mathfrak{a}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ characterising the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$. □

§22.22 **Proposition** *(Upper bound).* *Under Assumptions §19.02 and §22.20 and* $\mathbb{P}^\xi \in \mathrm{P}_{\{0\}\times\mathbb{R}^+_{\setminus 0}}$ *with* $\sigma_\xi^2 := \mathbb{P}^\xi(\mathrm{id}_{\mathbb{R}}^2) \in \mathbb{R}^+_{\setminus 0}$ *and* $\kappa_\xi^5 := \mathbb{P}^\xi(|\mathrm{id}_{\mathbb{R}}|^5) \in \mathbb{R}^+_{\setminus 0}$ *for* $\mathrm{M}^{\mathfrak{v}} \in \mathbb{N}$ *as in* (22.07) *and* $\widehat{\mathfrak{p}}\mathrm{en}^{\mathfrak{v}}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ *as in* (22.11) *the fully data-driven OPE* $\widehat{f}^{\widehat{m}}_\bullet = \widehat{f}_\bullet \mathbb{1}^{\widehat{m}}_\bullet \in \ell_{2(\mathfrak{v}^2_\bullet)}\mathbb{1}^{\widehat{m}}_\bullet \subseteq \ell_2(\mathfrak{v}^2_\bullet)$ *of* $f_\bullet \in \ell_2(\mathfrak{v}^2_\bullet)$ *with fully data-driven dimension* $\widehat{m} \in [\![\mathrm{M}^{\mathfrak{v}}]\!]$ *as in* (22.12) *satisfies*

$$\mathcal{R}^{\mathfrak{v}}_n[\,\widehat{f}^{m^\star_n}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant (3\mathrm{r}^2 + 288\tau^2_{\mathfrak{v},\mathrm{u}}(1+\sigma_\xi^2+\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}}))\min\big\{\mathrm{R}^m_n(f_\bullet,\mathfrak{v}_\bullet)\colon m\in[\![\mathrm{M}^{\mathfrak{v}}]\!]\big\} + \mathrm{C}\tau^2_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_\xi)+\mathfrak{v}_1^2)(1+\kappa_\xi^5+\mathrm{r}^5\tau^5_{\mathfrak{a},\mathrm{u}})n^{-1} \quad (22.15)$$

for $x_\xi := 6(\sigma_\xi^2+\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}})\tau^{-2}_{\mathfrak{v},\mathrm{u}} \in \mathbb{R}^+$ *and some universal numerical constant* $\mathrm{C} = 96(16\mathrm{C}_{\mathrm{tal}}+384) \in [1,\infty)$.

§22.23 **Proof** of Proposition §22.22. is given in the lecture. □

§22.24 **Comment**. The minimax bound $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) = \mathrm{R}^{m^\star_n}_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) = \min\big\{\mathrm{R}^m_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)\colon m\in\mathbb{N}\big\}$ (for details see Reminder §18.12) satisfies $n\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) \geqslant \|\mathbb{1}^{m^\star_n}_\bullet\|^2_{\mathfrak{v}} \geqslant \mathfrak{v}_1^2$. Consequently, the last upper bound in (22.15) and the minimax bound $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ coincide up to a constant $3\mathrm{r}^2+288\tau^2_{\mathfrak{v},\mathrm{u}}(1+\sigma_\xi^2+\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}})+\mathrm{C}\tau^2_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_\xi)\mathfrak{v}_1^{-2}+1)(1+\kappa_\xi^5+\mathrm{r}^5\tau^5_{\mathfrak{a},\mathrm{u}})$ provided the minimax dimension fulfils $m^\star_n \in [\![\mathrm{M}_n]\!]$. Therefore, we wish the upper bound $\mathrm{M}^{\mathfrak{v}}$ to be as large as possible. The next assertion shows that $\mathrm{M}^{\mathfrak{v}}$ as in (22.07) is a suitable choice for the upper bound. □

§22.25 **Corollary**. *Under the assumptions of Proposition §22.22 for each* $n\in\mathbb{N}$ *such that* $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) \leqslant \mathfrak{v}_1^2$ *and* $m^\star_n \leqslant n^{-2/3}\exp(\frac{n^{1/6}}{100})$ *we have*

$$\begin{aligned}\mathcal{R}^{\mathfrak{v}}_n[\,\widehat{f}^{\widehat{m}}_\bullet \,|\, \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] &\leqslant (3\mathrm{r}^2 + 288\tau^2_{\mathfrak{v},\mathrm{u}}(1+\sigma_\xi^2+\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}}))\min\big\{\mathrm{R}^m_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)\colon m\in[\![\mathrm{M}^{\mathfrak{v}}]\!]\big\}\\ &\quad + 96(16\mathrm{C}_{\mathrm{tal}}+384)\tau^2_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_\xi)+\mathfrak{v}_1^2)(1+\kappa_\xi^5+\mathrm{r}^5\tau^5_{\mathfrak{a},\mathrm{u}})n^{-1}\\ &\leqslant \mathrm{K}\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet) \quad (22.16)\end{aligned}$$

and, hence up to the constant $\mathrm{K} := 3\mathrm{r}^2+\mathrm{C}\tau^2_{\mathfrak{v},\mathrm{u}}(\mathrm{C}_{\mathfrak{v}}(x_\xi)\mathfrak{v}_1^{-2}+1)(1+\kappa_\xi^5+\mathrm{r}^5\tau^5_{\mathfrak{a},\mathrm{u}})$ *with universal numerical constant* $\mathrm{C} = 99(16\mathrm{C}_{\mathrm{tal}}+384) \in [1,\infty)$ *the feasible data-driven estimator* $\widehat{f}^{\widehat{m}}_\bullet$ *is minimax optimal.*

§22.26 **Proof** of Corollary §22.25. is given in the lecture. □

§22.27 **Illustration**. Consider the *trigonometric basis* as in Illustration §20.17 which satisfies Assumption §20.11 for all $\mathfrak{a}_\bullet \in \ell_2$ (see Illustration §20.17). In Table 02 [§12] the order of the rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ is depict for the two cases **(o)** and **(s)** introduced in Illustration §12.26. We note that we have $\mathfrak{a}_\bullet \in \ell_2$ in case **(o)** for $a > 1/2$ while in case **(s)** for $a \in \mathbb{R}^+_{\setminus 0}$. The sequence $\mathfrak{v}_\bullet$ satisfies Assumption §22.20, i.e. (22.04), for $v \geqslant -1/2$. Moreover, the optimal dimension $m^\star_n$ given in Table 02 [§12] satisfies $m^\star_n \leqslant n^{-2/3}\exp(\frac{n^{1/6}}{100})$, and thus (under the above restrictions) the adaptive density estimator attains the minimax optimal rate $\mathrm{R}^\star_n(\mathfrak{a}_\bullet,\mathfrak{v}_\bullet)$ up to the constant given in Corollary §22.25. □

§22|02 Data-driven local estimation by Goldenshluger and Lepskij's method

§22.28 **Reminder**. The Bernstein inequality stated in the form of Lemma §18.19 provides again our key argument in order to control the deviations of the reminder term. Let us briefly recall how we intend to apply the Bernstein inequality (see Remark §18.21 for a similar approach). Reconsider the stochastic process $\psi_\bullet = (\psi_j(X,Y) := Y\mathrm{u}_j(X))_{j\in\mathbb{N}} \in (\mathscr{B}_{[0,1]}\otimes\mathscr{B})\otimes 2^{\mathbb{N}}$ where $\psi_j \in \mathcal{L}_1(\mathcal{U}_f)$ for each $j\in\mathbb{N}$ and the OPE $\widehat{f}^m_\bullet = \widehat{f}_\bullet\mathbb{1}^m_\bullet \in \ell_2\mathbb{1}^m_\bullet$ with dimension $m\in\mathbb{N}$ (Definition §20.04). $\widehat{f}_\bullet = \widehat{\mathbb{P}}_n\psi_\bullet = (\widehat{\mathbb{P}}_n\psi_j)_{j\in\mathbb{N}}$ are noisy versions (Definition §15.08) of the regression coefficients $f_\bullet = \mathrm{U}f = \mathcal{U}_f(\psi_\bullet) = (\mathcal{U}_f\psi_j = \mathcal{U}_f(Y\mathrm{u}_j(X)))_{j\in\mathbb{N}}$ (see Notation §19.07). Clearly, $\mathrm{r}_m :=$

$\phi\nu_{\mathbb{N}}(\psi_\bullet \mathbb{1}_\bullet^m)$ is a $\mathscr{B}_{[0,1]} \otimes \mathscr{B}$-$\mathscr{B}$-measurable function, where $\widehat{\mathbb{P}}_n(\mathrm{r}_m) = \phi\nu_{\mathbb{N}}(\widehat{\mathbb{P}}_n(\psi_\bullet)\mathbb{1}_\bullet^m) = \phi\nu_{\mathbb{N}}(\widehat{f}_\bullet \mathbb{1}_\bullet^m)$ and $\mathcal{U}_f(\mathrm{r}_m) = \phi\nu_{\mathbb{N}}(\mathcal{U}_f(\psi_\bullet)\mathbb{1}_\bullet^m) = \phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^m)$, and thus $\bar{\mathrm{r}}_m = \widehat{\mathbb{P}}_n(\mathrm{r}_m) - \mathcal{U}_f(\mathrm{r}_m) = \phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet^m)$. We note that $\psi_\bullet$ and thus $\mathrm{r}_m = \phi\nu_{\mathbb{N}}(\psi_\bullet \mathbb{1}_\bullet^m)$ is not bounded. Therefore, we decompose $\psi_\bullet = \psi_\bullet^{\mathrm{b}} + \psi_\bullet^{\mathrm{u}}$ into two parts $\psi_\bullet^{\mathrm{b}}, \psi_\bullet^{\mathrm{u}} \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B}) \otimes 2^{\mathbb{N}}$, a bounded and a remaining unbounded one. To be more precise, setting $\mathrm{r}_m^{\mathrm{b}} := \phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}} \mathbb{1}_\bullet^m) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ and $\mathrm{r}_m^{\mathrm{u}} := \phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{u}} \mathbb{1}_\bullet^m) \in \mathscr{B}_{[0,1]} \otimes \mathscr{B}$ let $\sup\{|\mathrm{r}_m^{\mathrm{b}}(x,y)| \colon x \in [0,1], y \in \mathbb{R}\} \in \mathbb{R}$ be satisfied. Introducing further $\bar{\mathrm{r}}_m^{\mathrm{b}} := \widehat{\mathbb{P}}_n(\mathrm{r}_m^{\mathrm{b}}) - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{b}})$ and $\bar{\mathrm{r}}_m^{\mathrm{u}} := \widehat{\mathbb{P}}_n(\mathrm{r}_m^{\mathrm{u}}) - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{u}})$ we evidently have $\bar{\mathrm{r}}_m = \bar{\mathrm{r}}_m^{\mathrm{b}} + \bar{\mathrm{r}}_m^{\mathrm{u}}$ and hence

$$\begin{aligned} |\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet^m)| = |\bar{\mathrm{r}}_m^{\mathrm{b}} + \bar{\mathrm{r}}_m^{\mathrm{u}}|^2 &\leqslant 2|\bar{\mathrm{r}}_m^{\mathrm{b}}|^2 + 2|\bar{\mathrm{r}}_m^{\mathrm{u}}|^2 \\ &= 2|\widehat{\mathbb{P}}_n(\mathrm{r}_m^{\mathrm{b}}) - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{b}})|^2 + 2|\widehat{\mathbb{P}}_n(\mathrm{r}_m^{\mathrm{u}}) - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{u}})|^2. \end{aligned} \tag{22.17}$$

Considering the first term on the right hand side provided that

$$\begin{aligned} &\mathcal{U}_f(|\mathrm{r}_m^{\mathrm{b}} - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{b}})|^2) = \mathcal{U}_f(|\phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}} \mathbb{1}_\bullet^m) - \mathcal{U}_f(\phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}} \mathbb{1}_\bullet^m))|^2) \leqslant \mathrm{v}_{f,m}^2 \in \mathbb{R}^+, \\ &\sup\{|\mathrm{r}_m^{\mathrm{b}}(x,y)| \colon x \in [0,1], y \in \mathbb{R}\} \leqslant \mathrm{b}_m \in \mathbb{R}^+, \text{ and hence } |\mathrm{r}_m^{\mathrm{b}} - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{b}})| \leqslant 2\mathrm{b}_m, \end{aligned} \tag{22.18}$$

due to the Bernstein inequality (Lemma §18.19 (18.16)) we have

$$\mathcal{U}_f^{\otimes n}\big(\big(\big|n^{1/2}\bar{\mathrm{r}}_m^{\mathrm{b}}\big|^2 - (4\mathrm{v}_{f,m}^2 + 32\mathrm{b}_m^2(\log \mathrm{K})n^{-1})\log \mathrm{K}\big)_+\big) \leqslant 8\mathrm{K}^{-1}\{\mathrm{v}_{f,m}^2 + 16\mathrm{b}_m^2 n^{-1}\}. \tag{22.19}$$

for any $\mathrm{K} \in [1,\infty)$. □

§22|02|01 Local ϕ-risk

§22.29 **Assumption**. Let $\phi_\bullet \in (\mathbb{R}_{\backslash 0})^{\mathbb{N}}$ and the orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete* and satisfies as process $\mathrm{u}_\bullet = (\mathrm{u}_j)_{j\in\mathbb{N}}$ on $([0,1], \mathscr{B}_{[0,1]})$ for $\tau_{\mathrm{u}} \in [1,\infty)$ and for all $m \in \mathbb{N}$ satisfies **(os3)** $\sup\{\|\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\ell_2}^2 \colon x \in [0,1]\} \leqslant \tau_{\mathrm{u}}^2 m \in \mathbb{R}^+$. □

§22.30 **Remark**. Keeping Reminder §22.28 in mind we define $\psi_\bullet^{\mathrm{b}}, \psi_\bullet^{\mathrm{u}} \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B}) \otimes 2^{\mathbb{N}}$ as

$$\begin{aligned} \psi_\bullet^{\mathrm{b}}(x,y) = (\psi_j^{\mathrm{b}}(x,y) := y\mathbb{1}_{[0,n^{1/6}]}(|y|)\mathrm{u}_j(x))_{j\in\mathbb{N}} \text{ and} \\ \psi_\bullet^{\mathrm{u}}(x,y) = (\psi_j^{\mathrm{u}}(x,y) := y\mathbb{1}_{(n^{1/6},\infty)}(|y|)\mathrm{u}_j(x))_{j\in\mathbb{N}}, \quad x \in [0,1], y \in \mathbb{R} \end{aligned}$$

where evidently $\psi_\bullet^{\mathrm{b}} + \psi_\bullet^{\mathrm{u}} = \psi_\bullet$ and

$$|\mathrm{r}_m^{\mathrm{b}}| = |\phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}} \mathbb{1}_\bullet^m)| = |Y\mathbb{1}_{[0,n^{1/6}]}(|Y|)\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet(X)\mathbb{1}_\bullet^m)| \leqslant |Y\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet(X)\mathbb{1}_\bullet^m)| = |\phi\nu_{\mathbb{N}}(\psi_\bullet \mathbb{1}_\bullet^m)| = |\mathrm{r}_m|.$$

We use in the sequel that under Assumption §22.29 **(os3)** for each $m \in \mathbb{N}$

$$\begin{aligned} \sup\{|\mathrm{r}_m^{\mathrm{b}}(x,y)|^2 \colon x \in [0,1], y \in \mathbb{R}\} &= \sup\{|y\mathbb{1}_{[0,n^{1/6}]}(|y|)\phi\nu_{\mathbb{N}}(\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m)|^2 \colon x \in [0,1], y \in \mathbb{R}\} \\ &\leqslant n^{1/3}\|\mathbb{1}_\bullet^m\|_\phi^2 \sup\{\|\mathrm{u}_\bullet(x)\mathbb{1}_\bullet^m\|_{\ell_2}^2 \colon x \in [0,1]\} \leqslant n^{1/3}\tau_{\mathrm{u}}^2 m\|\mathbb{1}_\bullet^m\|_\phi^2 =: \mathrm{b}_m^2 \end{aligned} \tag{22.20}$$

by applying the Cauchy Schwarz inequality and moreover (see Proof §15.11)

$$\begin{aligned} \mathcal{U}_f(|\mathrm{r}_m^{\mathrm{b}} - \mathcal{U}_f(\mathrm{r}_m^{\mathrm{b}})|^2) \leqslant \mathcal{U}_f(|\mathrm{r}_m^{\mathrm{b}}|^2) = \mathcal{U}_f(|\phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}} \mathbb{1}_\bullet^m)|^2) =: \mathrm{v}_{f,m}^2 \\ \leqslant \mathcal{U}_f(|\phi\nu_{\mathbb{N}}(\psi_\bullet \mathbb{1}_\bullet^m)|^2) \leqslant (\sigma_\xi^2 + \|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}^2)\|\mathbb{1}_\bullet^m\|_\phi^2 \in \mathbb{R}^+. \end{aligned} \tag{22.21}$$

exploiting (??) in Proof §19.11. Combining (18.19), (18.20) and (18.18) (Remark §18.21) we obtain

$$\begin{aligned} &\mathcal{U}_f^{\otimes n}\big(\big(\big|n^{1/2}\bar{\mathrm{r}}_m^{\mathrm{b}}\big|^2 - (4\mathrm{v}_{f,m}^2 + 32\mathrm{b}_m^2(\log \mathrm{K})n^{-1})\log \mathrm{K}\big)_+\big) \\ &\qquad \leqslant 8\mathrm{K}^{-1}\{\sigma_\xi^2 + \|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})}^2 + 16\tau_{\mathrm{u}}^2 m n^{-2/3}\}\|\mathbb{1}_\bullet^m\|_\phi^2 \end{aligned} \tag{22.22}$$

for any $m \in \mathbb{N}$ and $\mathrm{K} \in [1,\infty)$. □

§22.31 **Reminder** *(Local oracle ϕ-risk).* Given Assumptions §19.02 and §22.29 we consider an OPE as in Definition §20.04. Here the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U}f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §22.29 Assumptions §15.02 and §18.22, (and hence Assumption §19.05 and $\phi_\bullet \in (\mathbb{R}_{\setminus 0})^{\mathbb{N}}$) and $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ in §20.24 an *oracle inequality* for the local ϕ-risk of the OPE's is shown. More presicely, as in (20.06) (Proposition §20.22) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}_n^m(f_\bullet, \phi_\bullet) := |\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{m|\perp})|^2 + n^{-1}\|\mathbb{1}_\bullet^m\|_\phi^2, \quad m_n^\circ := \arg\min\big\{\mathrm{R}_n^m(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}_n^\circ(f_\bullet, \phi_\bullet) := \mathrm{R}_n^{m_n^\circ}(f_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}_n^m(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (22.23)$$

and assuming $\mathbb{v}_f := \max(\sigma_\xi^{-2}, \sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}) \in \mathbb{R}^+_{\setminus 0}$ due to Property §20.09 the (infeasible) OPE $\widehat{f}_\bullet^{m_n^\circ} = \widehat{f}_\bullet \mathbb{1}_\bullet^{m_n^\circ} \in \ell_2 \mathbb{1}_\bullet^{m_n^\circ} \subseteq \mathrm{dom}(\phi\nu)$ with oracle dimension m_n° as in (22.23) satisfies

$$\mathbb{v}_f^{-1}\mathrm{R}_n^\circ(f_\bullet, \phi_\bullet) \leqslant \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet)|^2) \leqslant \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{m_n^\circ} - f_\bullet)|^2)$$
$$\leqslant \mathbb{v}_f \mathrm{R}_n^\circ(f_\bullet, \phi_\bullet) \leqslant \mathbb{v}_f^2 \inf_{m\in\mathbb{N}} \mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet)|^2),$$

and hence it is *oracle optimal* (with constant $\mathbb{v}_f^2$). □

Partially known penalty sequence

§22.32 **Notation.** Consider first a sequence of penalties $\mathfrak{pen}_\bullet^{f,\phi} = (\mathfrak{pen}_m^{f,\phi})_{m\in\mathbb{N}} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given by

$$\mathfrak{pen}_m^{f,\phi} := 24n^{-1}\big(\mathrm{v}_{f,m}^2 + 8\mathrm{b}_m^2(\log \mathrm{K}_m)n^{-1}\big)(\log \mathrm{K}_m) \quad \text{with} \quad \mathrm{v}_{f,m}^2 := \mathcal{U}_f\big(|\phi\nu_{\mathbb{N}}(\psi_\bullet^{\mathrm{b}}\mathbb{1}_\bullet^m)|^2\big),$$
$$\mathrm{b}_m^2 := \tau_{\mathrm{u}}^2 n^{1/3} m \|\mathbb{1}_\bullet^m\|_\phi^2, \quad \text{and} \quad \mathrm{K}_m := (1 \vee \|\mathbb{1}_\bullet^m\|_\phi^2)\, m^3 \in [1, \infty) \quad \text{for each } m \in \mathbb{N}, \quad (22.24)$$

which is obviously only *partially known* in advance, and fully known upper bound

$$\mathrm{M}^\phi := \max\big\{m \in \mathbb{N}\colon m\|\mathbb{1}_\bullet^m\|_\phi^2 \leqslant n^2\phi_1^2\big\} \in \mathbb{N} \quad (22.25)$$

where the defining set is not empty and finite (i.e. $\mathrm{M}^\phi \leqslant n^2$). Considering the data-driven OSE $\widehat{f}_\bullet^{\widehat{m}} = \widehat{f}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter selected by Goldenshluger and Lepskij's method

$$\widehat{m} := \arg\min\big\{\mathfrak{contr}_m^{f,\phi} + \mathfrak{pen}_m^{f,\phi}\colon m \in [\![\mathrm{M}^\phi]\!]\big\} \quad \text{and}$$
$$\mathfrak{contr}_m^{f,\phi} := \max\Big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^j - \widehat{f}_\bullet^m)|^2 - \mathfrak{pen}_j^{f,\phi} - \mathfrak{pen}_m^{f,\phi}\big)_+ \colon j \in]\!]m, \mathrm{M}^\phi]\!]\Big\}, \ m \in [\![\mathrm{M}^\phi]\!]. \quad (22.26)$$

Moreover, studying a ϕ-error the bias term introduced in (14.31) becomes

$$\mathfrak{bias}_m(f_\bullet, \phi_\bullet) = \sup\big\{|\phi\nu_{\mathbb{N}}(f_\bullet^j - f_\bullet^m)| = |\phi\nu_{\mathbb{N}}(f_\bullet \mathbb{1}_\bullet^{]\!]m,j]\!]})|\colon j \in [\![m, \infty]\!]\big\} \quad \forall m \in \mathbb{N}.$$

If $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ and hence $\nu_{\mathbb{N}}(|\phi_\bullet f_\bullet|) \in \mathbb{R}$ then $\mathfrak{bias}_m(f_\bullet, \phi_\bullet) \leqslant \nu_{\mathbb{N}}(|\phi_\bullet f_\bullet|\mathbb{1}_\bullet^{m|\perp}) = \mathrm{o}(1)$ as $m \to \infty$ by dominated convergence. Considering the data-driven OSE $\widehat{f}_\bullet^{\widehat{m}} = \widehat{f}_\bullet \mathbb{1}_\bullet^{\widehat{m}}$ with dimension parameter $\widehat{m}$ selected as in (22.26) with penalty sequence $\mathfrak{pen}_\bullet^{f,\phi}$ given in (22.24) and upper bound $\mathrm{M}^\phi \in \mathbb{N}$ as in (22.25) we derive below an upper bound for its local ϕ-risk, $\mathcal{U}_f^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^{\widehat{m}} - f_\bullet)|^2\big)$. □

§22.33 **Lemma.** *Under Assumptions §19.02 and §22.29, $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ and $Y \in \mathcal{L}_{14}(\mathcal{U}_f)$ for $\mathfrak{pen}_\bullet^{f,\phi} \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ as in (22.24) and $\mathrm{M}^\phi \in \mathbb{N}$ as in (22.25) we have*

$$\mathcal{U}_f^{\otimes n}\Big(\max\Big\{\big(|\phi\nu_{\mathbb{N}}(\widehat{f}_\bullet^m - f_\bullet^m)|^2 - \mathfrak{pen}_m^{f,\phi}/3\big)_+ \colon m \in [\![\mathrm{M}^\phi]\!]\Big\}\Big)$$
$$\leqslant \big\{28\sigma_\xi^2 + 28\|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 448\tau_{\mathrm{u}}^2 n^{-2/3} + 2\phi_1^2\tau_{\mathrm{u}}^2\mathcal{U}_f\big(Y^{14}\big)\big\}\, n^{-1} \quad (22.27)$$

§22.34 **Proof** of Lemma §22.33. is given in the lecture. □

§22.35 **Proposition** *(Upper bound). Under Assumptions §19.02 and §22.29,* $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *and* $Y \in \mathcal{L}_{14}(\mathcal{U}_f)$ *for* $\mathfrak{pen}^{f,\phi}_\bullet \in (\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ *as in* (22.24) *and* $\mathrm{M}^\phi \in \mathbb{N}$ *as in* (22.25) *the OPE* $\widehat{f}^{\widehat{m}}_\bullet = \widehat{f}_\bullet \mathbb{1}^{\widehat{m}}_\bullet \in \ell_2 \mathbb{1}^{\widehat{m}}_\bullet \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with partially data-driven dimension* $\widehat{m} \in [\![\mathrm{M}^\phi]\!]$ *as in* (22.26) *satisfies for all* $n \in \mathbb{N}$

$$\begin{aligned}\mathcal{U}_f^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}^{\widehat{m}}_\bullet - f_\bullet)|^2\big) &\leqslant 128(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 8\tau_{\mathrm{u}}^2)\\ &\quad \times \min\big\{\mathfrak{bias}^2_m(\mathfrak{p}_\bullet, \phi_\bullet) + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi (\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m) m n^{-2/3}) \colon m \in [\![\mathrm{M}]\!]\big\}\\ &\quad + 56\big(14\sigma_\xi^2 + 14\|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 224\tau_{\mathrm{u}}^2 n^{-2/3} + \phi_1^2\tau_{\mathrm{u}}^2 \mathcal{U}_f(Y^{14})\big) n^{-1}. \quad (22.28)\end{aligned}$$

§22.36 **Proof** of Proposition §22.35. is given in the lecture. □

§22.37 **Comment**. Let us compare the dominating part of the upper bound given in (22.28), that is

$$\min\big\{\mathfrak{bias}^2_m(f_\bullet, \phi_\bullet) + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi (\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m) m n^{-2/3}) \colon m \in [\![\mathrm{M}^\phi]\!]\big\} \quad (22.29)$$

with the oracle bound $\mathrm{R}^\circ_n(f_\bullet, \phi_\bullet) = \min\big\{|\phi\nu_{\mathbb{N}}(f^m_\bullet - f_\bullet)|^2 + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi : m \in \mathbb{N}\big\}$ (for details see Reminder §22.31). In (22.29) we face eventually a deterioration by three sources. First, we generally have $\mathfrak{bias}_m(f_\bullet, \phi_\bullet) \geqslant |\phi\nu_{\mathbb{N}}(f^m_\bullet - f_\bullet)|$, but note that for $f_\bullet\phi_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$ equality holds, that is

$$\mathfrak{bias}_m(f_\bullet, \phi_\bullet) = \sup\big\{\nu_{\mathbb{N}}(\phi_\bullet f_\bullet \mathbb{1}^{[\![m,j]\!]}_\bullet) \colon j \in [\![m, \infty]\!]\big\} = \nu_{\mathbb{N}}(\phi_\bullet f_\bullet \mathbb{1}^{m|\perp}_\bullet) = |\phi\nu_{\mathbb{N}}(f^m_\bullet - f_\bullet)|$$

for all $m \in \mathbb{N}$. Secondly, the variance term features an additional factor $(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m) m n^{-2/3})$, and finally the upper bound M^ϕ might impose an additional deterioration. The next assertion shows that M^ϕ is a suitable choice for the upper bound. Moreover, we estimate the bias term by $\mathfrak{bias}_m(f_\bullet, \phi_\bullet) \leqslant \nu(|\phi_\bullet f_\bullet| \mathbb{1}^{m|\perp}_\bullet)$ where equality holds whenever $f_\bullet\phi_\bullet \in (\mathbb{R}^+)^{\mathbb{N}}$. □

§22.38 **Corollary**. *For* $n, m \in \mathbb{N}$ *we set*

$$\begin{aligned}\mathrm{R}^m_n(f_\bullet, \phi_\bullet) &:= \big(\nu_{\mathbb{N}}(|\phi_\bullet f_\bullet| \mathbb{1}^{m|\perp}_\bullet)\big)^2\\ &\quad + (1 + \big(\log\|\mathbb{1}^m_\bullet\|^2_\phi\big)_+ + \log m)\big(1 + ((\big(\log\|\mathbb{1}^m_\bullet\|^2_\phi\big)_+ + \log m) m n^{-2/3}\big)\, n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi,\\ m^\diamond &:= \arg\min\big\{\mathrm{R}^m_n(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\} \quad \textit{and}\\ \mathrm{R}^\diamond_n(f_\bullet, \phi_\bullet) &:= \mathrm{R}^{m^\diamond}_n(f_\bullet, \phi_\bullet) = \min\big\{\mathrm{R}^m_n(f_\bullet, \phi_\bullet) : m \in \mathbb{N}\big\}. \quad (22.30)\end{aligned}$$

Under the assumptions of Proposition §22.35 for each $n \in \mathbb{N}$ *such that* $m^\diamond \in [\![\mathrm{M}^\phi]\!]$ *we have*

$$\begin{aligned}\mathcal{U}_f^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}^{\widehat{m}}_\bullet - f_\bullet)|^2\big) &\leqslant 1152(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 8\tau_{\mathrm{u}}^2)\mathrm{R}^\diamond_n(f_\bullet, \phi_\bullet)\\ &\quad + 56\big(14\sigma_\xi^2 + 14\|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 224\tau_{\mathrm{u}}^2 n^{-2/3} + \phi_1^2\tau_{\mathrm{u}}^2 \mathcal{U}_f(Y^{14})\big) n^{-1}\\ &\leqslant 9216\big((\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})(1 + \phi_1^{-2}) + \tau_{\mathrm{u}}^2(1 + 2\phi_1^{-2} n^{-2/3} + \mathcal{U}_f(Y^{14}))\big)\, \mathrm{R}^\diamond_n(f_\bullet, \phi_\bullet). \quad (22.31)\end{aligned}$$

§22.39 **Proof** of Proof §22.39. is given in the lecture. □

§22.40 **Comment**. The data-driven bound $\mathrm{R}^\diamond_n(f_\bullet, \phi_\bullet)$ compared to the oracle bound $\mathrm{R}^\circ_n(f_\bullet, \phi_\bullet)$ features a deterioration of the variance term at least by a logarithmic factor. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§22.41 **Illustration**. We illustrate the last results considering the two specifications **(o)** and **(s)** given in Table 03 [§12] (Illustration §12.40). We restrict ourselves to the case $\phi_\bullet \notin \ell_2$ only.

Table 01 [§22]

Order of the oracle rate $\mathrm{R}^\circ_n(f_\bullet,\phi_\bullet)$ and the data-driven rate $\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)$ as $n\to\infty$

$(j\in\mathbb{N})$ $\phi_j=j^{v-1/2}$	$(a\in\mathbb{R}^+_{\setminus 0})$ f_j	(squarred bias) $(\nu_{\mathbb{N}}(\lvert\phi_\bullet f_\bullet\rvert\mathbb{1}_\bullet^{m\lvert\perp}))^2$	(variance) $\lVert\mathbb{1}_\bullet^m\rVert_\phi^2$	M^ϕ	$m^\diamond$	$\mathrm{R}^\circ_n(f_\bullet,\phi_\bullet)$	$\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)$
(o) $v\in(0,a)$	$j^{-a-1/2}$	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{2}{2v+1}}$		$n^{-\frac{(a-v)}{a}}$	
$a\in(3/4,\infty)$					$\big(\frac{n}{\log n}\big)^{\frac{1}{2a}}$		$\big(\frac{\log n}{n}\big)^{\frac{(a-v)}{a}}$
$a\in(0,3/4]$					$\big(\frac{n^{5/6}}{\log n}\big)^{\frac{1}{a+1/2}}$		$\big(\frac{\log n}{n^{5/6}}\big)^{\frac{2(a-v)}{a+1/2}}$
$v=0$	$j^{-a-1/2}$	m^{-2a}	$\log m$	$\frac{n^2}{\log n}$		$\frac{\log n}{n}$	
$a\in(3/4,\infty)$					$\big(\frac{n}{(\log n)^2}\big)^{\frac{1}{2a}}$		$\frac{(\log n)^2}{n}$
$a\in(0,3/4]$					$\big(\frac{n^{5/3}}{(\log n)^3}\big)^{\frac{1}{2a+1}}$		$\big(\frac{(\log n)^3}{n^{5/3}}\big)^{\frac{a}{a+1/2}}$
(s) $v\in\mathbb{R}^+_{\setminus 0}$	$e^{-j^{2a}}$	$m^{(1-2(a-v))_+}e^{-2m^{2a}}$	m^{2v}	$n^{\frac{2}{2v+1}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}}(\log\log n)}{n}$
$v=0$	$e^{-j^{2a}}$	$m^{(1-2a)_+}e^{-2m^{2a}}$	$\log m$	$\frac{n^2}{\log n}$	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 01 [§22] the order of the oracle rate $\mathrm{R}^\circ_n(f_\bullet,\phi_\bullet)$ and the data-driven rate $\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)$ is depict for $v\geqslant 0$ only. In case $v<0$ we have $\phi_\bullet\in\ell_2$. Moreover, in case **(s)** for $a\in\mathbb{R}^+_{\setminus 0}$ and **(o)** for $a\in(3/4,\infty)$ the rate $\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)$ features only an additional logarithmic factor compared with the oracle rate $\mathrm{R}^\circ_n(f_\bullet,\phi_\bullet)$. □

Estimated penalty sequence

§22.42 **Notation**. The penalty sequence $\mathfrak{pen}^{f,\mathfrak{v}}_\bullet\in(\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given in (22.24) still depends on characteristics of the unknown regression function f. More precisely, for $m\in\mathbb{N}$ the term $\mathfrak{pen}^{f,\mathfrak{v}}_m$ involves the quantity $\mathrm{v}^2_{f,m}=\mathcal{U}_f\big(\lvert\phi\nu_{\mathbb{N}}(\psi^{\mathrm{b}}_\bullet\mathbb{1}^m_\bullet)\rvert^2\big)$ which we eventually estimate without bias by $\widehat{\mathrm{v}}^2_m:=\widehat{\mathbb{P}}_n\big(\lvert\phi\nu_{\mathbb{N}}(\psi^{\mathrm{b}}_\bullet\mathbb{1}^m_\bullet)\rvert^2\big)$. Based on this estimator let us introduce a fully data-driven sequence of penalties $\widehat{\mathfrak{pen}}^\phi_\bullet=(\widehat{\mathfrak{pen}}^\phi_m)_{m\in\mathbb{N}}\in(\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ given by

$$\widehat{\mathfrak{pen}}^\phi_m:=24n^{-1}\big(2\widehat{\mathrm{v}}^2_m+3\times 8\mathrm{b}^2_m(\log\mathrm{K}_m)n^{-1}\big)(\log\mathrm{K}_m)\quad\text{with}\quad\widehat{\mathrm{v}}^2_m:=\widehat{\mathbb{P}}_n\big(\lvert\phi\nu_{\mathbb{N}}(\psi^{\mathrm{b}}_\bullet\mathbb{1}^m_\bullet)\rvert^2\big),$$
$$\mathrm{b}^2_m:=\tau_{\mathrm{u}}^2n^{1/3}m\lVert\mathbb{1}^m_\bullet\rVert^2_\phi,\quad\text{and}\quad\mathrm{K}_m:=(1\vee\lVert\mathbb{1}^m_\bullet\rVert^2_\phi)\,m^3\in[1,\infty)\quad\text{for each } m\in\mathbb{N},\tag{22.32}$$

which is now *fully known* in advance, and fully known upper bound $\mathrm{M}^\phi\in\mathbb{N}$ defined in (22.25). Considering the data-driven OSE $\widehat{f}^{\widehat{m}}_\bullet=\widehat{f}_\bullet\mathbb{1}^{\widehat{m}}_\bullet$ with dimension parameter selected by Goldenshluger and Lepskij's method

$$\widehat{m}:=\arg\min\big\{\widehat{\mathfrak{c}}\mathrm{ontr}^\phi_m+\widehat{\mathfrak{pen}}^\phi_m\colon m\in[\![\mathrm{M}^\phi]\!]\big\}\quad\text{and}$$
$$\widehat{\mathfrak{c}}\mathrm{ontr}^\phi_m:=\max\Big\{\big(\lvert\phi\nu_{\mathbb{N}}(\widehat{\mathbb{p}}^j_\bullet-\widehat{\mathbb{p}}^m_\bullet)\rvert^2-\widehat{\mathfrak{pen}}^\phi_j-\widehat{\mathfrak{pen}}^\phi_m\big)_+\colon j\in[\![m,\mathrm{M}^\phi]\!]\Big\},\ m\in[\![\mathrm{M}^\phi]\!]\tag{22.33}$$

we derive below an upper bound for its local ϕ-risk, $\mathcal{U}^{\otimes n}_f\big(\lvert\phi\nu_{\mathbb{N}}(\widehat{f}^{\widehat{m}}_\bullet-f_\bullet)\rvert^2\big)$. □

§22.43 **Lemma**. *Under Assumptions §19.02 and §22.29 and $f\in\mathbb{L}_\infty(\lambda_{[0,1]})$ for $\mathfrak{pen}^{f,\phi}_\bullet,\widehat{\mathfrak{pen}}^\phi_\bullet\in(\mathbb{R}^+_{\setminus 0})^{\mathbb{N}}$ as in (22.24) and (22.32), respectively, and for any $\mathrm{M}\in\mathbb{N}$ we have*

$$\mathcal{U}^{\otimes n}_f\Big(\max\Big\{\big(\mathfrak{pen}^{f,\phi}_j-\widehat{\mathfrak{pen}}^\phi_j\big)_+\colon j\in[\![\mathrm{M}]\!]\Big\}\Big)\leqslant 80\{\sigma_\xi^2+\lVert f\rVert^2_{\mathbb{L}_\infty(\lambda_{[0,1]})}+6\tau_{\mathrm{u}}^2n^{-2/3}\}n^{-1}.\tag{22.34}$$

§22.44 **Proof** of Lemma §22.43. is given in the lecture. □

§22.45 **Proposition** *(Upper bound).* *Under Assumptions §19.02 and §22.29,* $Y \in \mathcal{L}_{14}(\mathcal{U}_f)$ *and* $f \in \mathbb{L}_\infty(\lambda_{[0,1]})$ *for* $\widehat{\mathfrak{pen}}^\phi_\bullet \in (\mathbb{R}^+_{\backslash 0})^{\mathbb{N}}$ *as in* (22.32) *and for* $\mathrm{M}^\phi \in \mathbb{N}$ *as in* (22.25) *the OPE* $\widehat{f}^{\widehat{m}}_\bullet = \widehat{f}_\bullet \mathbb{1}^{\widehat{m}}_\bullet \in \ell_2 \mathbb{1}^{\widehat{m}}_\bullet \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *of* $f_\bullet \in \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with fully data-driven dimension* $\widehat{m} \in [\![\mathrm{M}^\phi]\!]$ *as in* (22.33) *satisfies for all* $n \in \mathbb{N}$

$$\begin{aligned}\mathcal{U}_f^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}^{\widehat{m}}_\bullet - f_\bullet)|^2\big) &\leqslant 224(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 12\tau_{\mathrm{u}}^2)\\ &\quad\times \min\big\{\mathfrak{bias}^2_m(f_\bullet,\phi) + n^{-1}\|\mathbb{1}^m_\bullet\|^2_\phi(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m)mn^{-2/3}) \colon m \in [\![\mathrm{M}^\phi]\!]\big\}\\ &\quad + 72\big(40\sigma_\xi^2 + 40\|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 240\tau_{\mathrm{u}}^2 n^{-2/3} + \phi_1^2\tau_{\mathrm{u}}^2\mathcal{U}_f(Y^{14})\big)n^{-1}. \quad (22.35)\end{aligned}$$

§22.46 **Proof** of Proposition §22.45. is given in the lecture. □

§22.47 **Comment**. We shall stress that the last upper bound (22.35) in Proposition §22.45 (for the fully data-driven procedure) and the upper bound (22.28) in Proposition §22.35 (for the partially data-driven procedure) differ only in the numerical constants. Thus, thus the proof of the next result follows line by line their counterparts above. □

§22.48 **Corollary**. *Under the assumptions of Proposition §22.45 for each* $n \in \mathbb{N}$ *such that* $m^\diamond \in [\![\mathrm{M}^\phi]\!]$ *we have*

$$\begin{aligned}\mathcal{U}_f^{\otimes n}\big(|\phi\nu_{\mathbb{N}}(\widehat{f}^{\widehat{m}}_\bullet - f_\bullet)|^2\big) &\leqslant 2016(\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 12\tau_{\mathrm{u}}^2)\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)\\ &\quad + 72\big(40\sigma_\xi^2 + 40\|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})} + 240\tau_{\mathrm{u}}^2 n^{-2/3} + \phi_1^2\tau_{\mathrm{u}}^2\mathcal{U}_f(Y^{14})\big)n^{-1}\\ &\leqslant 2016\big((\sigma_\xi^2 + \|f\|^2_{\mathbb{L}_\infty(\lambda_{[0,1]})})(1+\phi_1^{-2}) + 12\tau_{\mathrm{u}}^2(1 + \phi_1^{-2}n^{-2/3} + \mathcal{U}_f(Y^{14}))\big)\,\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet). \quad (22.36)\end{aligned}$$

§22.49 **Proof** of Proof §22.49. is given in the lecture. □

§22.50 **Comment**. The fullay data-driven bound $\mathrm{R}^\diamond_n(f_\bullet,\phi_\bullet)$ equals up to the numerical constants the bound in the partially known case. Therefore, the Comment §22.40 and the Illustration §22.41 apply here equally. □

§22|02|02 Maximal local ϕ-risk

§22.51 **Assumption**. Consider $\phi_\bullet, \mathfrak{a}_\bullet \in (\mathbb{R}_{\backslash 0})^{\mathbb{N}}$ with $\mathfrak{a}_\bullet \in \ell_\infty$ and $(\mathfrak{a}\phi)_\bullet := (\mathfrak{a}_j\phi_j)_{j\in\mathbb{N}} = \mathfrak{a}_\bullet\phi_\bullet \in \ell_2$, and hence $\|\mathfrak{a}_\bullet\mathbb{1}^{m|\perp}_\bullet\|_\phi = \|(\mathfrak{a}\phi)_\bullet\mathbb{1}^{m|\perp}_\bullet\|_{\ell_2} = \mathrm{o}(1)$ as $m \to \infty$. The orthonormal system $(\mathrm{u}_j)_{j\in\mathbb{N}}$ in $\mathbb{L}_2(\lambda_{[0,1]})$ is **(os1)** *complete*, i.e an *orthonormal basis* in $\mathbb{L}_2(\lambda_{[0,1]})$ and as process $\mathrm{u}^2_\bullet = (\mathrm{u}_j^2)_{j\in\mathbb{N}}$ on $([0,1],\mathscr{B}_{[0,1]})$ satisfies **(os2)** $\|\nu_{\mathbb{N}}(\mathfrak{a}^2_\bullet\mathrm{u}^2_\bullet)\|_{\mathbb{L}_2(\lambda_{[0,1]})} \leqslant \tau^2_{\mathfrak{a},\mathrm{u}}$ and for all $m \in \mathbb{N}$, **(os3)** $\sup\big\{\|\mathrm{u}_\bullet(x)\mathbb{1}^m_\bullet\|^2_{\ell_2} \colon x \in [0,1]\big\} \leqslant \tau^2_{\mathrm{u}}\|\mathbb{1}^m_\bullet\|^2_{\mathfrak{v}_\bullet} \in \mathbb{R}^+$. for $\tau_{\mathfrak{a},\mathrm{u}}, \tau_{\mathrm{u}} \in [1,\infty)$. □

§22.52 **Remark**. Under Assumption §22.51 considering the set $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ of regression functions in $\mathbb{L}_2(\lambda_{[0,1]})$ defined in (20.04) we have $\|f\|_{\mathbb{L}_\infty(\lambda_{[0,1]})} \leqslant \mathrm{r}\tau_{\mathfrak{a},\mathrm{u}}$ for all $f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ due to Lemma §20.13. Consequently, given in addition Assumption §19.02 all assumptions of Proposition §22.45 are satisfied. □

§22.53 **Reminder** *(Maximal local ϕ-risk).* Given Assumptions §19.02 and §22.51 we consider an OPE as in Definition §20.04. Here the observable noisy version $\widehat{f}_\bullet = f_\bullet + n^{-1/2}\varepsilon_\bullet$ of the regression coefficients $f_\bullet = \mathrm{U}f \in \ell_2$ take the form of a *statistical direct problem* (see Definition §10.19) where the stochastic processes $\varepsilon_\bullet \in (\mathscr{B}_{[0,1]} \otimes \mathscr{B})^{\otimes n} \otimes 2^{\mathbb{N}}$ is given in Definition §19.08. Under Assumptions §19.02 and §22.51 (and hence Assumption §20.26) in Proposition §20.29 an upper bound for a maximal local ϕ-risk of an OPE is shown over the set $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ given in (20.04) (Lemma §20.13)

More precisely, the performance of the OPE $\widehat{f}^{m}_{\bullet} = \widehat{f}_{\bullet}\mathbb{1}^{m}_{\bullet} \in \ell_2\mathbb{1}^{m}_{\bullet} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ with dimension $m \in \mathbb{N}$ is measured by its maximal global ϕ-risk over the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$, that is

$$\mathcal{R}^{\phi}_{n}[\,\widehat{f}^{m}_{\bullet}\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] := \sup\big\{\mathcal{U}_f^{\otimes n}(|\phi\nu_{\mathbb{N}}(\widehat{f}^{m}_{\bullet} - f_{\bullet})|^2)\colon f \in \mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\big\}.$$

As in (12.13) (Proposition §12.42) for all $n, m \in \mathbb{N}$ setting

$$\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) := \|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi} + n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}, \quad m^{\star}_{n} := \arg\min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) : m \in \mathbb{N}\big\}$$
$$\text{and} \quad \mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\phi) := \mathrm{R}^{m^{\star}_{n}}_{n}(\mathfrak{a}_{\bullet},\phi) = \min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) : m \in \mathbb{N}\big\}. \quad (22.37)$$

by Proposition §20.29 under Assumptions §19.02 and §22.51 the maximal local ϕ-risk of an OPE $\widehat{f}^{m^{\star}_{n}}_{\bullet}$ with optimally choosen dimension $m^{\star}_{n}$ as in (22.37) satisfies

$$\mathcal{R}^{\phi}_{n}[\,\widehat{f}^{m^{\star}_{n}}_{\bullet}\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant \mathrm{C}\,\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\phi)$$

with $\mathrm{C} = \sigma_{\xi}^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}}$. Moreover, due to Proposition §21.09 $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\phi)$ provides (up to a constant) also a lower bound of the maximal global ϕ-risk over the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$ for any estimator. Consequently, (up to a constant) $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\phi)$ is a minimax bound and $\widehat{f}^{m^{\star}_{n}}_{\bullet}$ is minimax optimal. However, the optimal dimension $m^{\star}_{n}$ depends on $\mathfrak{a}_{\bullet} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ characterising the ellipsoid $\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}$. □

§22.54 **Proposition** *(Upper bound).* *Under Assumptions §19.02 and §22.51 and* $\mathbb{P}^{\xi} \in \mathrm{P}_{\{0\}\times\mathbb{R}^{+}_{\setminus 0}}$ *with* $\sigma^2_{\xi} := \mathbb{P}^{\xi}(\mathrm{id}^2_{\mathbb{R}}) \in \mathbb{R}^{+}_{\setminus 0}$ *and* $\kappa^{14}_{\xi} := \mathbb{P}^{\xi}(\mathrm{id}^{14}_{\mathbb{R}}) \in \mathbb{R}^{+}_{\setminus 0}$ *for* $\mathrm{M}^{\phi} \in \mathbb{N}$ *as in* (22.25) *and and* $\widehat{\mathfrak{p}}\mathrm{en}^{\phi}_{\bullet} \in (\mathbb{R}^{+}_{\setminus 0})^{\mathbb{N}}$ *as in* (22.32) *the OPE* $\widehat{f}^{\widehat{m}}_{\bullet} = \widehat{f}_{\bullet}\mathbb{1}^{\widehat{m}}_{\bullet} \in \ell_2\mathbb{1}^{\widehat{m}}_{\bullet} \subseteq \mathrm{dom}(\phi\nu_{\mathbb{N}})$ *with fully data-driven dimension* $\widehat{m} \in [\![\mathrm{M}^{\phi}]\!]$ *as in* (22.33) *satisfies for all* $n \in \mathbb{N}$

$$\mathcal{R}^{\phi}_{n}[\,\widehat{f}^{\widehat{m}}_{\bullet}\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant 224(\sigma_{\xi}^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + 12\tau^2_{\mathrm{u}})$$
$$\times \min\big\{\|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi} + n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}(\log \mathrm{K}_m)(1 \vee (\log \mathrm{K}_m)mn^{-2/3})\colon m \in [\![\mathrm{M}^{\phi}]\!]\big\}$$
$$+ 576\big(5\sigma_{\xi}^2 + 5\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + 30\tau^2_{\mathrm{u}}n^{-2/3} + 2^{11}\phi_1^2\tau^2_{\mathrm{u}}(\kappa^{14}_{\xi} + \mathrm{r}^{14}\tau^{14}_{\mathfrak{a},\mathrm{u}})\big)n^{-1}. \quad (22.38)$$

§22.55 **Proof** of Proposition §22.54. is given in the lecture. □

§22.56 **Corollary**. *Under the assumptions of Proposition §22.54 for* $n, m \in \mathbb{N}$ *we set*

$$\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) := \|\mathfrak{a}_{\bullet}\mathbb{1}^{m|\perp}_{\bullet}\|^2_{\phi}$$
$$+ \big(1 + \big(\log\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}\big)_{+} + \log m\big)\big(1 + ((\big(\log\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi}\big)_{+} + \log m)mn^{-2/3}\big)\, n^{-1}\|\mathbb{1}^{m}_{\bullet}\|^2_{\phi},$$
$$m^{\diamond} := \arg\min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) : m \in \mathbb{N}\big\} \quad \textit{and}$$
$$\mathrm{R}^{\diamond}_{n}(\mathfrak{a}_{\bullet},\phi) := \mathrm{R}^{m^{\diamond}}_{n}(\mathfrak{a}_{\bullet},\phi) = \min\big\{\mathrm{R}^{m}_{n}(\mathfrak{a}_{\bullet},\phi) : m \in \mathbb{N}\big\}. \quad (22.39)$$

For each $n \in \mathbb{N}$ *such that* $m^{\diamond} \in [\![\mathrm{M}^{\phi}]\!]$ *we have*

$$\mathcal{R}^{\phi}_{n}[\,\widehat{f}^{\widehat{m}}_{\bullet}\,|\,\mathbb{F}_2^{\mathfrak{a},\mathrm{r}}\,] \leqslant 2016(\sigma_{\xi}^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + 12\tau^2_{\mathrm{u}})\mathrm{R}^{\diamond}_{n}(\mathfrak{a}_{\bullet},\phi)$$
$$+ 576\big(5\sigma_{\xi}^2 + 5\mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + 30\tau^2_{\mathrm{u}}n^{-2/3} + 2^{11}\phi_1^2\tau^2_{\mathrm{u}}(\kappa^{14}_{\xi} + \mathrm{r}^{14}\tau^{14}_{\mathfrak{a},\mathrm{u}})\big)n^{-1}$$
$$\leqslant 576\big((4 + 5\phi_1^{-2})(\sigma_{\xi}^2 + \mathrm{r}^2\tau^2_{\mathfrak{a},\mathrm{u}} + 12\tau^2_{\mathrm{u}}) + 2^{11}\tau^2_{\mathrm{u}}(\kappa^{14}_{\xi} + \mathrm{r}^{14}\tau^{14}_{\mathfrak{a},\mathrm{u}})\big)\mathrm{R}^{\diamond}_{n}(\mathfrak{a}_{\bullet},\phi). \quad (22.40)$$

§22.57 **Proof** of Corollary §22.56. is given in the lecture. □

§22.58 **Comment**. The data-driven bound $\mathrm{R}^{\diamond}_{n}(\mathfrak{a}_{\bullet},\phi)$ compared to the minimax bound $\mathrm{R}^{\star}_{n}(\mathfrak{a}_{\bullet},\phi)$ features a deterioration of the variance term at least by a logarithmic factor. The appearance of the logarithmic factor within the bound is a known fact in the context of local estimation (cf. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of nonparametric Gaussian regression and hence it is widely considered as an acceptable price for adaptation. □

§22.59 **Illustration**. We illustrate the last results considering the two specifications **(o)** and **(o)** given in Table 04 [§12] (Illustration §12.47). We restrict ourselves again to the case $\phi_{\bullet} \notin \ell_2$ only.

Table 02 [§22]

Order of the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ and the data-driven rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ as $n \to \infty$

$(j \in \mathbb{N})$ $\phi_j = j^{v-1/2}$	$(a \in \mathbb{R}_{\setminus 0}^{+})$ $\mathfrak{a}_j^2$	(squarred bias) $\lVert \mathfrak{a}_{\bullet} \mathbb{1}_{\bullet}^{m\vert\perp} \rVert_{\phi}^2$	(variance) $\lVert \mathbb{1}_{\bullet}^{m} \rVert_{\phi}^2$	M^{ϕ}	$m^{\diamond}$	$\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$	$\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$
(o) $v \in (0, a)$	j^{-a}	$m^{-2(a-v)}$	m^{2v}	$n^{\frac{2}{2v+1}}$		$n^{-\frac{(a-v)}{a}}$	
$a \in (3/4, \infty)$					$\big(\frac{n}{\log n}\big)^{\frac{1}{2a}}$		$\big(\frac{\log n}{n}\big)^{\frac{(a-v)}{a}}$
$a \in (0, 3/4]$					$\big(\frac{n^{5/6}}{\log n}\big)^{\frac{1}{a+1/2}}$		$\big(\frac{\log n}{n^{5/6}}\big)^{\frac{2(a-v)}{a+1/2}}$
$v = 0$	j^{-a}	m^{-2a}	$\log m$	$\frac{n^2}{\log n}$		$\frac{\log n}{n}$	
$a \in (3/4, \infty)$					$\big(\frac{n}{(\log n)^2}\big)^{\frac{1}{2a}}$		$\frac{(\log n)^2}{n}$
$a \in (0, 3/4]$					$\big(\frac{n^{5/3}}{(\log n)^3}\big)^{\frac{1}{2a+1}}$		$\big(\frac{(\log n)^3}{n^{5/3}}\big)^{\frac{a}{a+1/2}}$
(s) $v \in \mathbb{R}_{\setminus 0}^{+}$	$e^{-j^{2a}}$	$m^{2(v-a)_+} e^{-m^{2a}}$	m^{2v}	$n^{\frac{1}{2v}}$	$(\log n)^{\frac{1}{2a}}$	$\frac{(\log n)^{\frac{v}{a}}}{n}$	$\frac{(\log n)^{\frac{v}{a}} (\log\log n)}{n}$
$v = 0$	$e^{-j^{2a}}$	$e^{-m^{2a}}$	$\log m$	e^{n}	$(\log n)^{\frac{1}{2a}}$	$\frac{\log\log n}{n}$	$\frac{(\log\log n)^2}{n}$

We note that in Table 02 [§22] the order of the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ and the data-driven rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ is depict for $v \geqslant 0$ only. In case $v < 0$ we have $\phi_{\bullet} \in \ell_2$. Moreover, in case **(s)** for $a \in \mathbb{R}_{\setminus 0}^{+}$ and **(o)** for $a \in (3/4, \infty)$ the rate $\mathrm{R}_n^{\diamond}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$ features only an additional logarithmic factor compared with the minimax rate $\mathrm{R}_n^{\star}(\mathfrak{a}_{\bullet}, \phi_{\bullet})$. □

Appendix A

Probability theory

Elements of the PROBABILITY THEORY *are recalled along the lines of the text book* **Klenke** *[2008] where a detailed exposition with many examples can be found.*

§19 Fundamentals

§19.01 **Notation.** For $x, y \in \mathbb{R}$ we agree on the following notations $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leqslant x\}$ (integer part), $x \vee y = \max(x, y)$ (maximum), $x \wedge y = \min(x, y)$ (minimum), $x^+ = \max(x, 0)$ (positive part), $x^- = \max(-x, 0)$ (negative part) and $|x| = x^- + x^+$ (modulus).

(i) We set $\mathbb{R}^+ := [0, \infty)$, $\mathbb{R}^+_{\setminus 0} := (0, \infty)$, $\mathbb{R}_{\setminus 0} := \mathbb{R} \setminus \{0\}$, $\overline{\mathbb{R}} := [-\infty, \infty]$, $\overline{\mathbb{R}}^+ := [0, \infty]$.

(ii) For $a, b \in \mathbb{R}$ with $a < b$ we write $[\![a, b]\!] := [a, b] \cap \mathbb{Z}$, $[\![a, b[\![:= [a, b) \cap \mathbb{Z}$ and $]\!]a, b]\!] := (a, b] \cap \mathbb{Z}$. Moreover, let $[\![n]\!] := [\![1, n]\!]$ and $[\![n[\![:= [\![1, n[\![$ for $n \in \mathbb{N}$.

(iii) For $a^n = (a_i)_{i \in [\![n]\!]}, b^n = (b_i)_{i \in [\![n]\!]} \in \overline{\mathbb{R}}^n$ we write $a^n < b^n$, if $a_i < b_i$ for all $i \in [\![n]\!]$. For $a^n < b^n$, define the open *rectangle* as the Cartesian product $(a^n, b^n) := \bigtimes_{i=1}^n (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$. Analogously, we define $[a^n, b^n]$, $(a^n, b^n]$ and $[a^n, b^n)$.

(iv) We call $\overline{\mathscr{B}} := \mathscr{B}_{\overline{\mathbb{R}}}$ the Borel-σ-field over the compactified real line $\overline{\mathbb{R}}$, where the sets $\{-\infty\}$, $\{\infty\}$ and $\mathbb{R}$ are in $\overline{\mathbb{R}}$ closed and open, respectively, and hence Borel-measurable. In particular, the trace $\mathscr{B} := \mathscr{B}_{\mathbb{R}} = \overline{\mathscr{B}} \cap \mathbb{R}$ of $\overline{\mathscr{B}}$ over $\mathbb{R}$ is the Borel-σ-field over $\mathbb{R}$. Furthermore, we write $\overline{\mathscr{B}}^+ := \overline{\mathscr{B}} \cap \overline{\mathbb{R}}^+$, $\mathscr{B}^+ := \mathscr{B} \cap \mathbb{R}^+$ and $\mathscr{B}^+_{\setminus 0} := \mathscr{B} \cap \mathbb{R}^+_{\setminus 0}$.

(v) Given a measurable space $(\Omega, \mathscr{A})$ a Borel-measurable function $g : \Omega \to \mathbb{R}$ and $f : \Omega \to \overline{\mathbb{R}}$ is called *real* and *numerical*, respectively, and we write $g \in \mathscr{A}$ and $f \in \overline{\mathscr{A}}$ for short. g respectively f is called positive if $g(\Omega) \in \mathbb{R}^+$ respectively $f(\Omega) \in \overline{\mathbb{R}}^+$, then we write $g \in \mathscr{A}^+$ and $f \in \overline{\mathscr{A}}^+$ - We call a Borel-measurable function $f^k = (f_i)_{i \in [\![k]\!]} : \Omega \to \overline{\mathbb{R}}^k$, that is $f_i \in \overline{\mathscr{A}}$ for each $i \in [\![k]\!]$, and $g^k = (g_i)_{i \in [\![k]\!]} : \Omega \to \mathbb{R}^k$, *numerical* and *real*, respectively and we write $f^k \in \overline{\mathscr{A}}^k$ and $g^k \in \mathscr{A}^k$ for short. □

§19.02 **Property.**

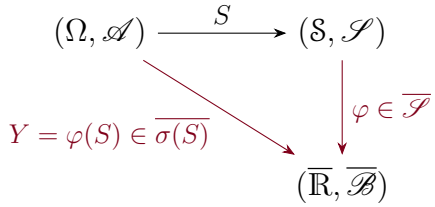
(i) *For $X, Y \in \overline{\mathscr{A}}$ and $a \in \mathbb{R}$ holds: $aX \in \overline{\mathscr{A}}$ (with convention $0 \times \infty = 0$); $X \vee Y := \max(X, Y), X \wedge Y := \min(X, Y) \in \overline{\mathscr{A}}$ and particularly $X^+ := X \vee 0$, $X^- := (-X)^+ \in \overline{\mathscr{A}}^+$, $|X| \in \overline{\mathscr{A}}^+$, $\{X < Y\}, \{X \leqslant Y\}, \{X = Y\} \in \mathscr{A}$, and $\lfloor X \rfloor \in \overline{\mathscr{A}}^+$.*

(ii) *For $X^n = (X_i)_{i \in [\![n]\!]} \in \mathscr{A}^n$, i.e., $X_i \in \mathscr{A}$, $i \in [\![n]\!]$, and Borel-measurable $h : \mathbb{R}^n \to \mathbb{R}^m$ holds $h(X^n) \in \mathscr{A}^m$, and in particular $X_1 + X_2$, $X_1 - X_2$, $X_1 X_2 \in \mathscr{A}$, and $X_1 / X_2 \in \overline{\mathscr{A}}$.*

(iii) *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence in $\overline{\mathscr{A}}$. Then $\sup_{n \in \mathbb{N}} X_n \in \overline{\mathscr{A}}$, $\inf_{n \in \mathbb{N}} X_n \in \overline{\mathscr{A}}$, $X_\star = \liminf_{n \to \infty} X_n \in \overline{\mathscr{A}}$ and $X^\star = \limsup_{n \to \infty} X_n \in \overline{\mathscr{A}}$. If $X := \lim_{n \to \infty} X_n$ exists, then $X \in \overline{\mathscr{A}}$.*

(iv) *Let $S : (\Omega, \mathscr{A}) \to (\mathbb{S}, \mathscr{S})$ be measurable, $\sigma(S) := S^{-1}(\mathscr{S}) \subseteq \mathscr{A}$ the sub-σ-field generated by S and $Y : \Omega \to \overline{\mathbb{R}}$. Then the following conditions are equivalent: (a) Y is $\sigma(S)$-measurable, symbolically $Y \in \overline{\sigma(S)}$; (b) There exists a measurable $\varphi : (\mathbb{S}, \mathscr{S}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, in short $\varphi \in \overline{\mathscr{S}}$, with $Y = \varphi(S)$. If Y is real, bounded or positive, then φ has each of those*

properties too.



The function φ is uniquely determined by Y on $S(\Omega)$, and for all $s \notin S(\Omega)$ it can be arbitrarily be extended.

(v) *For every $X \in \overline{\mathscr{A}}^+$ the sequence of simple random variables $(X_n)_{n\in\mathbb{N}}$ in $\overline{\mathscr{A}}^+$ given by $X_n := (2^{-n}\lfloor 2^n X\rfloor) \wedge n$ satisfies* (a) $X_n \uparrow X$; (b) $X_n \leqslant X \wedge n$; (c) *For each $c \in \mathbb{R}^+$ holds $\lim_{n\to\infty} X_n = X$ uniformly on $\{X \leqslant c\}$.* □

§19.03 **Notation.** For a measure μ on $(\Omega, \mathscr{A})$ we denote the integral of $f \in \overline{\mathscr{A}}$ with respect to μ by $\mu f := \int f d\mu$, if it exists. For $s \in \mathbb{R}^+_{\setminus 0}$ define $\|f\|_{\mathcal{L}_s(\mu)} := (\mu|f|^s)^{1/s}$, and $\|f\|_{\mathcal{L}_\infty(\mu)} := \inf\{c \in \mathbb{R}^+ : \mu(|f| > c) = 0\}$. For $s \in \overline{\mathbb{R}}^+_{\setminus 0} := (0, \infty]$ a function $f \in \overline{\mathscr{A}}$ is called $\mathcal{L}_s(\mu)$-integrable, if $\|f\|_{\mathcal{L}_s(\mu)} < \infty$. We denote the set of all $\mathcal{L}_s(\mu)$-integrable functions by $\mathcal{L}_s(\mu) := \mathcal{L}_s(\mathscr{A}, \mu) := \{f \in \overline{\mathscr{A}} : \|f\|_{\mathcal{L}_s(\mu)} < \infty\}$. Note that $\|\cdot\|_{\mathcal{L}_s(\mu)}$ is a seminorm on $\mathcal{L}_s(\mu)$ for each $s \in [1, \infty]$. Given a metric space $(\mathcal{X}, d)$ equipped with its Borel-σ-field $\mathscr{B}_\mathcal{X}$ we denote by $\mathcal{C}_b := \mathcal{C}_b(\mathcal{X})$ the set of all bounded and continuous functions mapping $\mathcal{X}$ into $\mathbb{R}$. For any finite measure μ on $(\mathcal{X}, \mathscr{B}_\mathcal{X})$ we have $\|h\|_{\mathcal{L}_\infty(\mu)} < \infty$ for all $h \in \mathcal{C}_b$ and thus $\mathcal{C}_b \subseteq \mathcal{L}_\infty(\mathscr{B}_\mathcal{X}, \mu)$ in equal. We denote by λ the Lebesgue measure on $(\mathbb{R}, \mathscr{B})$ and write shortly $\mathcal{L}_s := \mathcal{L}_s(\mathscr{B}) := \mathcal{L}_s(\mathscr{B}, \lambda)$. □

§19.04 **Notation.** We understand a vector $a^k = (a_i)_{i\in[\![k]\!]}$ as a column vector, i.e., $a^k = (a_1 \cdots a_k)^t \in \overline{\mathbb{R}}^k$ and hence we identify $\overline{\mathbb{R}}^k$ and $\overline{\mathbb{R}}^{(k,1)}$. We denote by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ the Euclidean norm and inner product on $\mathbb{R}^k$, respectively, i.e, $\|a^k\| = (\sum_{i\in[\![k]\!]} |a_i|^2)^{1/2}$ and $\langle a^k, b^k\rangle = \sum_{i\in[\![k]\!]} a_i b_i = (b^k)^t a^k$ for all $a^k, b^k \in \overline{\mathbb{R}}^k$. For $s \in \mathbb{R}^+_{\setminus 0}$ we define $\|a^k\|_s := (\sum_{i\in[\![k]\!]} |a_i|^s)^{1/s}$ and $\|a^k\|_\infty := \max_{i\in[\![k]\!]} |a_i|$. Note that $f^k \in \overline{\mathscr{A}}^k$ and $g^k \in \mathscr{A}^k$ imply $\|f^k\|_s \in \overline{\mathscr{A}}$ and $\|g^k\|_s \in \mathscr{A}$ for any $s \in \overline{\mathbb{R}}^+_{\setminus 0}$. We call $f^k = (f_i)_{i\in[\![k]\!]}$ $\mathcal{L}^k_s(\mu)$-integrable if $\|f^k\|_s \in \mathcal{L}_s(\mu)$ or equivalently $f_i \in \mathcal{L}_s(\mu)$ for each $i \in [\![k]\!]$. We define $\|f^k\|_{\mathcal{L}^k_s(\mu)} := \|\|f^k\|_p\|_{\mathcal{L}_s(\mu)}$ and $\mathcal{L}^k_s(\mu) := \mathcal{L}^k_s(\mathscr{A}, \mu) := \{f^k \in \overline{\mathscr{A}}^k : \|f^k\|_{\mathcal{L}^k_s(\mu)} < \infty\}$ with a slight abuse of notation. □

§19.05 **Notation.** Let X be a random variable, i.e. a measurable function, defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{X}, \mathscr{X})$. The probability measure on $(\mathcal{X}, \mathscr{X})$ induced by X is denoted by $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$ and we write $X \sim \mathbb{P}^X$ for short. For $f \in \overline{\mathscr{X}}$ the expectation of f with respect to $\mathbb{P}^X$ or equivalently of $f(X)$ with respect to $\mathbb{P}$ (if it exists) is denoted by $\mathbb{E}^f_X := \mathbb{P}^X f = \mathbb{P} f(X) =: \mathbb{E} f(X)$ for short. For example, when applied to the empirical measure $\widehat{\mathbb{P}}_n$ given by $\widehat{\mathbb{P}}_n(x^n) := \frac{1}{n}\sum_{i\in[\![n]\!]} \delta_{x_i}$ for $x^n = (x_i)_{i\in[\![n]\!]} \in \mathcal{X}^n$ this yields $\widehat{\mathbb{P}}_n f \in \overline{\mathscr{X}}$ with $x^n \mapsto (\widehat{\mathbb{P}}_n f)(x^n) := \frac{1}{n}\sum_{i\in[\![n]\!]} f(x_i)$. In other words, for each $x^n \in \mathcal{X}^n$, $(\widehat{\mathbb{P}}_n f)(x^n)$ is an abbreviation for the average $\frac{1}{n}\sum_{i\in[\![n]\!]} f(x_i)$. We denote by $\mathcal{W}(\mathscr{X})$ the set of all probability measures on $(\mathcal{X}, \mathscr{X})$ and for $\mathbb{R}^n$ equipped with its Borel-σ-field $\mathscr{B}^n := \mathscr{B}_{\mathbb{R}^n}$ by $\mathcal{W}_s(\mathscr{B}^n) \subseteq \mathcal{W}(\mathscr{B}^n)$ the subset of all probability measures on $(\mathbb{R}^n, \mathscr{B}^n)$ with finite $s \in \mathbb{R}^+$ absolute mean, that is, for all $\mathbb{P} \in \mathcal{W}_s(\mathscr{B}^n)$ the identity mapping $\mathrm{id}_n : \mathbb{R}^n \to \mathbb{R}^n$ belongs to $\mathcal{L}^n_s(\mathbb{P})$. Furthermore, for $Y \sim \mathbb{P}$ we write $\mathbb{E}(Y) = \mathbb{P}(Y) := \mathbb{P}(\mathrm{id}_n) = \big(\mathbb{P}(\Pi_i)\big)_{i\in[\![n]\!]}$ using for $i \in [\![n]\!]$ the coordinate map $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ with $x^n = (x_i)_{i\in[\![n]\!]} \mapsto \Pi_i(x^n) := x_i$. □

§19.06 **Property**. *Let $X \in \mathcal{L}_2^k(\mathbb{P})$, i.e. $\|X\|^2_{\mathcal{L}_2^k(\mathbb{P})} = \mathbb{P}(\|X\|^2) < \infty$. For each $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{(n,k)}$ we have $Y := AX + b \in \mathcal{L}_2^n(\mathbb{P})$. If we further denote by $\mu := \mathbb{P}X \in \mathbb{R}^k$ and $\Sigma := \mathbb{C}\mathrm{ov}(X) = \mathbb{P}(X-\mu)(X-\mu)^t = \mathbb{P}(XX^t) - \mu\mu^t \in \mathbb{R}^{(k,k)}$ expectation vector and covariance matrix of X, respectively, then $\mathbb{P}(Y) = A\mu + b \in \mathbb{R}^n$ and $\mathbb{C}\mathrm{ov}(Y) = A\Sigma A^t \in \mathbb{R}^{(n,n)}$.* □

§19.07 **Definition**. A $\mathcal{L}_2^k(\mathbb{P})$-random vector X with $\mu := \mathbb{P}(X)$ and $\Sigma := \mathbb{C}\mathrm{ov}(X)$ is *multivariate normally distributed*, $X \sim \mathrm{N}_{(\mu,\Sigma)}$ for short, if for each $c \in \mathbb{R}^k$ the real random variable $\langle X, c\rangle$ is normally distributed with mean $\langle \mu, c\rangle$ and variance $\langle \Sigma c, c\rangle$, i.e., $\langle X, c\rangle \sim \mathrm{N}_{(\langle \mu, c\rangle, \langle \Sigma c, c\rangle)}$. If Id_k denotes the k-dimensional identity matrix, then $X \sim \mathrm{N}_{(0,\mathrm{Id}_k)}$ is called a *standard normal random vector*. □

§19.08 **Property**. *A random vector $X = (X_i)_{i\in[\![k]\!]}$ is standard normal, i.e., $X \sim \mathrm{N}_{(0,\mathrm{Id}_k)}$ if and only if its components $\{X_i : i \in [\![k]\!] \in \mathbb{K}\}$ are independent and identically $\mathrm{N}_{(0,1)}$-distributed.* □

§19.09 **Remark**. In other words, a multivariate $\mathrm{N}_{(0,\mathrm{Id}_k)}$-distribution equals the product of its marginal $\mathrm{N}_{(0,1)}$-distributions, or $\mathrm{N}_{(0,\mathrm{Id}_k)} = \mathrm{N}_{(0,1)}{}^{\otimes k} := \bigotimes_{i\in[\![k]\!]} \mathrm{N}_{(0,1)}$ for short. □

§20 Convergence of random variables

Here and subsequently, a metric space is equipped with its Borel-σ-field.

§20.01 **Definition**. Let X and X_n, $n \in \mathbb{N}$, be random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a metric space $(\mathcal{X}, d)$. The sequence $(X_n)_{n\in\mathbb{N}}$ *converges to* X:

(a) *almost surely ($\mathbb{P}$-a.s.)*, if $\mathbb{P}(\lim_{n\to\infty} d(X_n, X) = 0) = 1$. We write $X_n \xrightarrow{n\to\infty} X$ $\mathbb{P}$-a.s., or briefly, $X_n \xrightarrow{\mathbb{P}\text{-a.s.}} X$.

(b) *almost completely ($\mathbb{P}$-a.c.)*, if $\sum_{n\in\mathbb{N}} \mathbb{P}(d(X_n, X) > \varepsilon) < \infty$ for all $\varepsilon \in \mathbb{R}^+_{\setminus 0}$. We write $X_n \xrightarrow{n\to\infty} X$ $\mathbb{P}$-a.c., or briefly, $X_n \xrightarrow{\mathbb{P}\text{-a.c.}} X$.

(c) *in probability*, if $\lim_{n\to\infty} \mathbb{P}\big(d(X_n, X) \geqslant \varepsilon\big) = 0$ for all $\varepsilon \in \mathbb{R}^+_{\setminus 0}$. We write $X_n \xrightarrow{n\to\infty} X$ in $\mathbb{P}$, or briefly, $X_n \xrightarrow{\mathbb{P}} X$.

(d) *in distribution*, if $\lim_{n\to\infty} \mathbb{P}^{X_n} f = \mathbb{P}^X f$ for any $f \in \mathcal{C}_b(\mathcal{X})$. We write $X_n \xrightarrow{n\to\infty} X$ in distribution, or briefly, $X_n \xrightarrow{d} X$ and with a slight abuse of notation also $X_n \xrightarrow{d} \mathbb{P}^X$.

(e) *in $\mathcal{L}_s(\mathbb{P})$ or s-th mean*, if $\lim_{n\to\infty} \mathbb{P}(d(X_n, X)^s) = 0$. We write $X_n \xrightarrow{n\to\infty} X$ in $\mathcal{L}_s(\mathbb{P})$, or briefly, $X_n \xrightarrow{\mathcal{L}_s(\mathbb{P})} X$. □

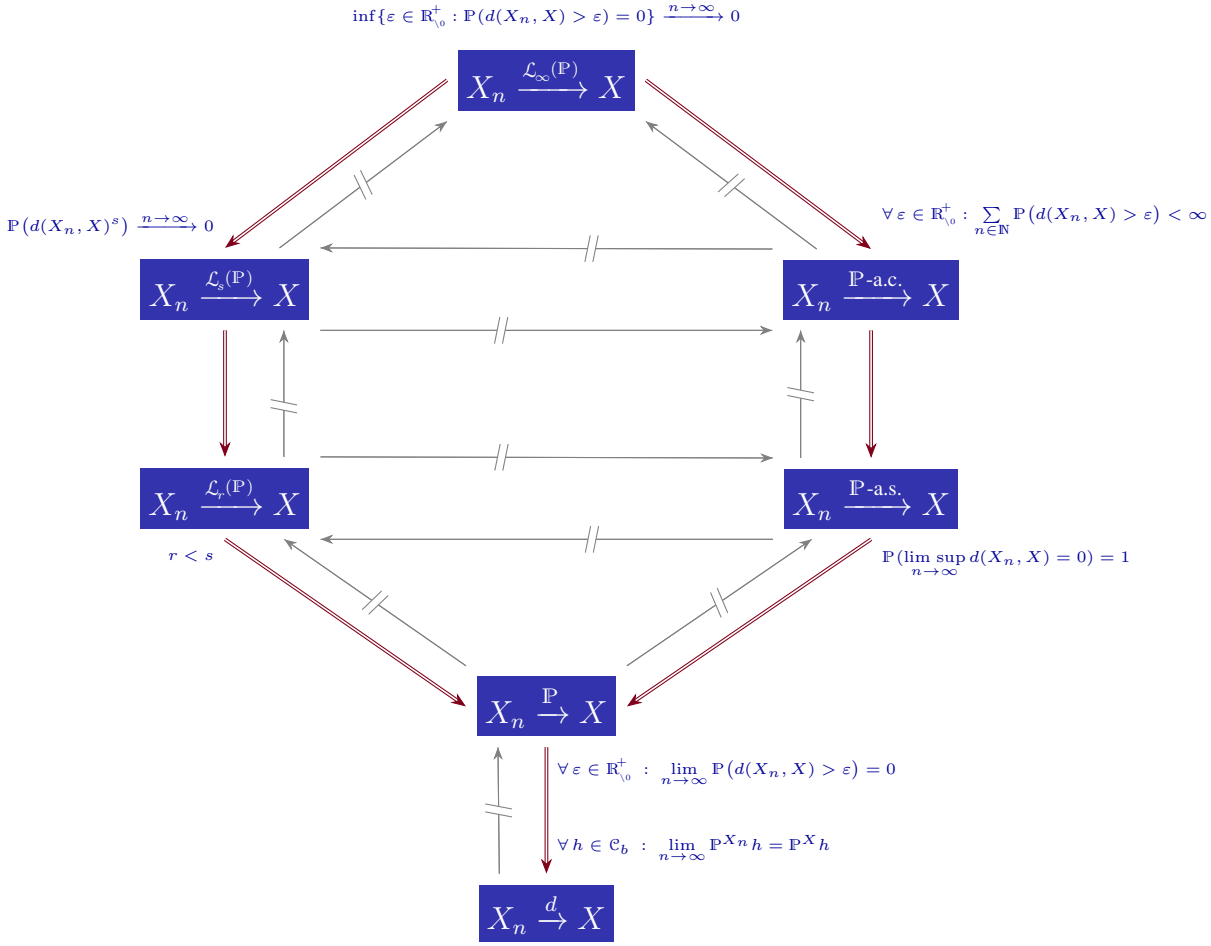
§20.02 **Remark**. Let X and X_n, $n \in \mathbb{N}$, be random vectors in $\mathbb{R}^k$, i.e., $(\mathbb{R}^k, \mathscr{B}^k)$-valued random variables, and $\|\cdot\|_s$ as in Notation §19.04. Convergence of $(X_n)_{n\in\mathbb{N}}$ to X in s-th mean, that is, $\mathbb{P}\|X_n - X\|_s^s = \|X_n - X\|^s_{\mathcal{L}_s^k(\mathbb{P})} \xrightarrow{n\to\infty} 0$, equals the component-wise convergence of $(X_n^i)_{n\in\mathbb{N}}$ to X^i in $\mathcal{L}_s(\mathbb{P})$, i.e., $\mathbb{P}|X_n^i - X^i|^s = \|X_n^i - X^i\|^s_{\mathcal{L}_s(\mathbb{P})} \xrightarrow{n\to\infty} 0$ for each $i \in [\![k]\!]$. □

§20.03 **Property**. *Let X and X_n, $n \in \mathbb{N}$, be random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a metric space $(\mathcal{X}, d)$.*

(i) *The following statements are equivalent:* (a) $X_n \xrightarrow{\mathbb{P}\text{-a.s.}} X$; (b) $\sup_{m\geqslant n} d(X_m, X_n) \xrightarrow{\mathbb{P}} 0$; (c) $\forall \varepsilon, \delta \in \mathbb{R}^+_{\setminus 0} : \exists N \in \mathbb{N} : \forall n \geqslant N : \mathbb{P}\big(\bigcap_{j\geqslant n} \{d(X_j, X) \leqslant \varepsilon\}\big) \geqslant 1 - \delta$ *and* (d) $\sup_{m\geqslant n} d(X_m, X) \xrightarrow{\mathbb{P}} 0$.

(ii) *(Continuous mapping theorem) Let $g : \mathcal{X} \to \mathbb{R}$ be continuous and let $(X_n)_{n\in\mathbb{N}}$ converge to X $\mathbb{P}$-a.s. (respectively, in probability or in distribution). Then $(g(X_n))_{n\in\mathbb{N}}$ converges to $g(X)$ $\mathbb{P}$-a.s. (respectively, in probability or in distribution).*

(iii) *Counter examples show, that the converse (in gray) of the following direct implications (in red) do not hold.* □



§20.04 **Definition**. A family of $\{X_{n,j} : j \in [\![k_n]\!], n \in \mathbb{N} \in \mathbb{K}\}$ of real $\mathcal{L}_2$-random variables is called a standardised array, if for every $n \in \mathbb{N}$ the family $\{X_{n,j} : j \in [\![k_n]\!] \in \mathbb{K}\}$ is independent, centred and normed, i.e., $\mathbb{E}(X_{n,j}) = 0$, $j \in [\![k_n]\!]$ and $\sum_{j\in[\![k_n]\!]} \mathfrak{var}_m(X_{n,j}) = 1$. A standardised array $\{X_{n,j} : j \in [\![k_n]\!], n \in \mathbb{N} \in \mathbb{K}\}$ is said to satisfy

(a) the *Lindeberg condition*, if $\lim_{n\to\infty} \sum_{j\in[\![k_n]\!]} \mathbb{E}\big(X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}|\geqslant\delta\}}\big) = 0$ for every $\delta \in \mathbb{R}^+_{\setminus 0}$;

(b) the *Lyapunov condition*, if there is $\delta \in \mathbb{R}^+_{\setminus 0}$ such that $\lim_{n\to\infty} \sum_{j\in[\![k_n]\!]} \mathbb{E}|X_{n,j}|^{2+\delta} = 0$. □

§20.05 **Property**. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent real random variables.*

(i) *(Law of Large Numbers) Let X_n, $n \in \mathbb{N}$, be identically distributed. Then $X_1 \in \mathcal{L}_1(\mathbb{P})$ if and only if $\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[\![n]\!]} X_i = \mathbb{P}(X_1)$ $\mathbb{P}$-a.s. (and then also in $\mathcal{L}_1(\mathbb{P})$).*

(ii) *(Lévy's equivalence theorem) For partial sums $(S_n := \sum_{i\in[\![n]\!]} X_i)_{n\in\mathbb{N}}$ $\mathbb{P}$-a.s. convergence is equivalent to convergence in probability. Otherwise, they diverge with probability one. (Kolmogorov's three-series theorem) $(S_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-a.s. if and only if there is $\varepsilon \in \mathbb{R}^+_{\setminus 0}$ such that each of the following three conditions holds:* **(a)** *$\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) < \infty$;* **(b)** *$\sum_{n\in\mathbb{N}} \mathbb{E}(X_n \mathbb{1}_{\{|X_n|\leqslant\varepsilon\}})$ converges; and* **(c)** *$\sum_{n\in\mathbb{N}} \mathfrak{var}_m(X_n \mathbb{1}_{\{|X_n|\leqslant\varepsilon\}}) < \infty$.*

Let $\{X_{n,j} : j \in [\![k_n]\!], n \in \mathbb{N} \in \mathbb{K}\}$ be a standardised array.

(iii) *The Lyapunov condition implies the Lindeberg condition.*

(iv) *(Central Limit Theorem of Lindeberg (1922)) If the Lindeberg condition hold, then (for the row sum)* $S_n^* = \sum_{j\in[\![k_n]\!]} X_{nj} \xrightarrow{d} \mathrm{N}_{(0,1)}$. □

§20.06 **Remark** *(Law of Large Numbers).* Let X_n^k, $n \in \mathbb{N}$, be i.i.d. random vector in $\mathbb{R}^k$. Then $\|X_1^k\|_{\mathcal{L}_1^k(\mathbb{P})} = \mathbb{P}\|X_1^k\|_1 < \infty$ if and only if $\frac{1}{n}\sum_{i\in[\![n]\!]} X_i^k \xrightarrow{\mathbb{P}\text{-a.s.}} \mathbb{E}(X_1^k)$ (then also in $\mathcal{L}_1^k(\mathbb{P})$). □

§20.07 **Property** *(Portemanteau). Let X and X_n, $n \in \mathbb{N}$, be random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a metric space $(\mathcal{X}, d)$. The following statements are equivalent:*

(i) $X_n \xrightarrow{d} X$;

(ii) $\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geqslant \mathbb{P}(X \in U)$ *for all open* $U \subseteq \mathcal{X}$;

(iii) $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leqslant \mathbb{P}(X \in F)$ *for all closed* $F \subseteq \mathcal{X}$;

(iv) $\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ *for all measurable* B *with* $\mathbb{P}(X \in \partial B) = 0$ *where* $\overline{B}$, B *and* $\partial B = \overline{B}\backslash B$ *is the closure, interior and the boundary of* B, *respectively.* □

§20.08 **Property** *(Helly-Bray). Let X and X_n, $n \in \mathbb{N}$, be random vectors in $\mathbb{R}^k$ with cumulative distribution function (c.d.f.) for each $x \in \mathbb{R}^k$ given by $\mathbb{F}(x) := \mathbb{P}(X \leqslant x)$ and $\mathbb{F}_n(x) := \mathbb{P}(X_n \leqslant x)$. Then the following statements are equivalent:* (i) $X_n \xrightarrow{d} X$ *and* (ii) $\lim_{n\to\infty} \mathbb{F}_n(x) = \mathbb{F}(x)$ *for all points of continuity* x *of* $\mathbb{F}$. □

§20.09 **Property** *(Continuous mapping theorem). Let $(\mathcal{X}_1, d_1)$ and $(\mathcal{X}_2, d_2)$ be metric spaces and let $\varphi : \mathcal{X}_1 \to \mathcal{X}_2$ be measurable. Denote by U_φ the set of points of discontinuity of φ. If X and X_n, $n \in \mathbb{N}$, are $\mathcal{X}_1$-valued random variables with $\mathbb{P}(X \in U_\varphi) = 0$ and $X_n \xrightarrow{d} X$, then $\varphi(X_n) \xrightarrow{d} \varphi(X)$.* □

§20.10 **Property** *(Slutzky's lemma). Let X and X_n, Y_n, $n \in \mathbb{N}$, be random variables taking values in a common metric space $(\mathcal{X}, d)$ and satisfying $X_n \xrightarrow{d} X$ and $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$. Then $Y_n \xrightarrow{d} X$.* □

§20.11 **Example**. Let X and X_n, $n \in \mathbb{N}$, be a random vector in $\mathbb{R}^k$ satisfying $X_n \xrightarrow{d} X$.

(a) If Y_n, $n \in \mathbb{N}$, are random vector in $\mathbb{R}^k$ and $c \in \mathbb{R}^k$ such that $Y_n \xrightarrow{d} c$, then $X_n + Y_n \xrightarrow{d} X + c$.

(b) If Σ_n, $n \in \mathbb{N}$ are random matrices in $\mathbb{R}^{(k,k)}$ and Σ is a matrix in $\mathbb{R}^{(k,k)}$ such that $\Sigma_n \xrightarrow{d} \Sigma$, then $\Sigma_n X_n \xrightarrow{d} \Sigma X$. If in addition Σ is strictly positive definite, and thus invertible, then $\Sigma_n^{-1} X_n \xrightarrow{d} \Sigma^{-1} X$ and $\Sigma_n^{-1/2} X_n \xrightarrow{d} \Sigma^{-1/2} X$. □

§20.12 **Property** *(Cramér-Wold device). Let X_n, $n \in \mathbb{N}$, be random vectors in $\mathbb{R}^k$. Then, the following are equivalent:* (a) *There is a random vector X with $X_n \xrightarrow{d} X$.* (b) *For any $v \in \mathbb{R}^k$, there is a real X^v with $\langle v, X_n\rangle \xrightarrow{d} X^v$. If* (a) *and* (b) *hold, then X^v and $\langle v, X\rangle$ are identically distributed (i.d.), $X^v \overset{d}{=} \langle v, X\rangle$ for short, for all $v \in \mathbb{R}^k$.* □

§20.13 **Property** *(Lindeberg-Feller CLT). For each $n \in \mathbb{N}$ let $\{Y_{n,j} : j \in [\![k_n]\!] \in \mathbb{K}\}$ be independent and centred $\mathcal{L}_2^p$-random vectors such that* (i) $\sum_{j\in[\![k_n]\!]} \mathbb{E}\|Y_{n,j}\|^2 \mathbb{1}_{\{\|Y_{n,j}\|>\varepsilon\}} \xrightarrow{n\to\infty} 0$ *for any* $\varepsilon \mathbb{R}_{\backslash 0}^+$ *and* (ii) $\sum_{j\in[\![k_n]\!]} \mathbb{E}(Y_{n,j} Y_{n,j}^t) \xrightarrow{n\to\infty} \Sigma$. *Then* $\sum_{j\in[\![k_n]\!]} Y_{n,j} \xrightarrow{d} \mathrm{N}_{(0,\Sigma)}$. □

§20.14 **Example**. Let X and X_n, $n \in \mathbb{N}$, be i.i.d. $\mathcal{L}_2^k(\mathbb{P})$-random vectors with $\mu = \mathbb{P}(X)$ and strictly positive definite $\Sigma = \mathbb{C}\mathrm{ov}(X)$.

(a) (CLT) $\frac{1}{\sqrt{n}}\sum_{i\in[\![n]\!]}(X_i-\mu)\xrightarrow{d}\mathrm{N}_{(0,\Sigma)}$,

(b) (LLN) $\overline{X}_n:=\frac{1}{n}\sum_{i\in[\![n]\!]}X_i\xrightarrow{\mathbb{P}}\mu$,

(c) (LLN) $\frac{1}{n}\sum_{i\in[\![n]\!]}X_iX_i^t\xrightarrow{\mathbb{P}}\mathbb{E}(XX^t)$,

(d) $\widehat{\Sigma}_n:=\frac{1}{n}\sum_{i\in[\![n]\!]}(X_i-\overline{X}_n)(X_i-\overline{X}_n)^t=\frac{1}{n}\sum_{i\in[\![n]\!]}X_iX_i^t-\overline{X}_n\overline{X}_n^t\xrightarrow{\mathbb{P}}\mathbb{E}(XX^t)-\mu\mu^t=\mathbb{C}\mathrm{ov}(X)=\Sigma$ (using (b) and (c) and continuous mapping theorem §20.03)

(e) $\sqrt{n}\Sigma_n^{-1/2}(\overline{X}-\mu)\xrightarrow{d}\mathrm{N}_{(0,\mathrm{Id}_s)}$ (using (a), (d) and Slutzky's lemma §20.10 as in the Example §20.11 (b)) □

§20.15 **Remark.** A map $\phi:\mathbb{R}^k\to\mathbb{R}^m$, that is defined at least in a neighbourhood of θ_o, is called differentiable at θ_o, if there exists a linear map (matrix) $\dot{\phi}_{\theta_o}:\mathbb{R}^k\to\mathbb{R}^m$ such that

$$\lim_{\theta\to\theta_o}\frac{\|\phi(\theta)-\phi(\theta_o)-\dot{\phi}_{\theta_o}(\theta-\theta_o)\|}{\|\theta-\theta_o\|}=0.$$

The linear map $x\mapsto\dot{\phi}_{\theta_o}(x)$ is called *(total) derivative* as opposed to partial derivatives. A sufficient condition for ϕ to be (totally) differentiable is that all partial derivatives $\partial\phi_j(\theta)/\partial\theta_l$ exist for θ in a neighbourhood of θ_o and are continuous at θ_o. □

§20.16 **Property** *(Delta method).* *Let $\phi:\mathbb{R}^k\supseteq\mathcal{D}_\phi\to\mathbb{R}^m$ be a map defined on a subset $\mathcal{D}_\phi$ of $\mathbb{R}^k$ and differentiable at θ_o. Let T and T_n, $n\in\mathbb{N}$ be random variables taking their values in the domain $\mathcal{D}_\phi$ of ϕ. If $r_n(T_n-\theta_o)\xrightarrow{d}T$ for numbers $r_n\to\infty$, then $r_n(\phi(T_n)-\phi(\theta_o))\xrightarrow{d}\dot{\phi}_{\theta_o}(T)$. Moreover, the difference between $r_n(\phi(T_n)-\phi(\theta_o))$ and $\dot{\phi}_{\theta_o}(r_n(T_n-\theta_o))$ converges to zero in probability.* □

§20.17 **Remark.** Commonly, $\sqrt{n}(T_n-\theta_o)\xrightarrow{d}\mathrm{N}_{(\mu,\Sigma)}$. Then applying the delta method it follows that $\sqrt{n}(\phi(T_n)-\phi(\theta_o))\xrightarrow{d}\mathrm{N}_{(\dot{\phi}_{\theta_o}\mu,\dot{\phi}_{\theta_o}\Sigma\dot{\phi}_{\theta_o}^t)}$. □

§20.18 **Property** *(Markov's inequality).* *If X is a $\mathcal{L}_s^k(\mathbb{P})$-random vector for some $s\geqslant 1$, then $\mathbb{P}(\|X\|_s>c)\leqslant c^{-s}\mathbb{P}(\|X\|_s^s)=c^{-s}\|X\|_{\mathcal{L}_s^k(\mathbb{P})}^s$.* □

§20.19 **Property** *(Monotone convergence).* *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of monotonically increasing real $\mathcal{L}_1(\mathbb{P})$-random variables converging $\mathbb{P}$-a.s. to a numerical random variable X, for short $X_n\uparrow X$ $\mathbb{P}$-a.s.. Then $\mathbb{P}X=\lim_{n\to\infty}\mathbb{P}X_n$.* □

§20.20 **Property** *(Dominated convergence).* *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real $\mathcal{L}_1(\mathbb{P})$-random variables converging $\mathbb{P}$-a.s. to a numerical random variable X, i.e., $X_n\xrightarrow{\mathbb{P}\text{-a.s.}}X$. If there is a real $\mathcal{L}_1(\mathbb{P})$ random variable Y with $\sup_{n\in\mathbb{N}}|X_n|\leqslant Y$ $\mathbb{P}$-a.s. (and thus $\sup_{n\in\mathbb{N}}|X_n|\in\mathcal{L}_1(\mathbb{P})$), then $X\in\mathcal{L}_1(\mathbb{P})$ and $X_n\xrightarrow{\mathcal{L}_1(\mathbb{P})}X$.* □

§20.21 **Definition.** A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with values in a metric space $(\mathcal{X},d)$ is called *(uniformly) tight* (straff) or *bounded in probability*, if, for any $\varepsilon\in\mathbb{R}^+_{\setminus 0}$, there exists a compact set $K_\varepsilon\subseteq\mathcal{X}$ such that $\mathbb{P}(X_n\in K_\varepsilon)\geqslant 1-\varepsilon$ for all $n\in\mathbb{N}$. □

§20.22 **Remark.** If $(\mathcal{X},d)$ is Polish, i.e., separable and complete, then every $\mathcal{X}$-valued random variable is bounded in probability and thus so is every finite family. □

§20.23 **Example.** A sequence $(X_n)_{n\in\mathbb{N}}$ of random vectors in $\mathbb{R}^k$ is bounded in probability, if for any $\varepsilon>0$, there exists a constant K_ε such that $\mathbb{P}(\|X_n\|>K_\varepsilon)\leqslant\varepsilon$ for all $n\in\mathbb{N}$. □

§20.24 **Property** *(Prohorov's theorem).* *Let X and X_n, $n \in \mathbb{N}$, be random variables with values in a Polish space.*

(i) *If $X_n \xrightarrow{d} X$, then $(X_n)_{n\in\mathbb{N}}$ is bounded in probability.*

(ii) *If $(X_n)_{n\in\mathbb{N}}$ is bounded in probability, then there exists a sub-sequence $(X_{n_k})_{k\in\mathbb{N}}$ which converges in distribution.* □

§20.25 **Landau notation.** Let X_n, $n \in \mathbb{N}$, be random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a metric space $(\mathcal{X}, d)$ and let x_n, $n \in \mathbb{N}$, belong to $\mathcal{X}$.

(i) We write (a) $x_n = \mathrm{o}(1)$, if $d(x_n, 0) \xrightarrow{n\to\infty} 0$, and (b) $x_n = \mathrm{O}(1)$, if $\sup_{n\in\mathbb{N}} d(x_n, 0) < \infty$, and analogously (a) $X_n = \mathrm{o}_{\mathbb{P}}(1)$, if $X_n \xrightarrow{\mathbb{P}} 0$, and (b) $X_n = \mathrm{O}_{\mathbb{P}}(1)$, if $(X_n)_{n\in\mathbb{N}}$ is bounded in probability

(ii) Let a_n, $n \in \mathbb{N}$, be strictly positive numbers. We write (a) $x_n = \mathrm{o}(a_n)$, if $d(x_n, 0)/a_n = \mathrm{o}(1)$, and that (b) $x_n = \mathrm{O}(a_n)$, if $d(x_n, 0)/a_n = \mathrm{O}(1)$, and analogously (a) $X_n = \mathrm{o}_{\mathbb{P}}(a_n)$, if $d(X_n, 0)/a_n = \mathrm{o}_{\mathbb{P}}(1)$, and (b) $X_n = \mathrm{O}_{\mathbb{P}}(a_n)$, if $d(X_n, 0)/a_n = \mathrm{O}_{\mathbb{P}}(1)$.

(iii) Let A_n, $n \in \mathbb{N}$, be strictly positive random variables on $(\Omega, \mathscr{A}, \mathbb{P})$. We write (a) $X_n = \mathrm{o}_{\mathbb{P}}(A_n)$, if $d(X_n, 0)/A_n = \mathrm{o}_{\mathbb{P}}(1)$, and (b) $X_n = \mathrm{O}_{\mathbb{P}}(A_n)$, if $d(X_n, 0)/A_n = \mathrm{O}_{\mathbb{P}}(1)$. □

§20.26 **Property** *(Exercise).* *For real random variables the following properties hold:*

(i) $\mathrm{o}_{\mathbb{P}}(1) + \mathrm{o}_{\mathbb{P}}(1) = \mathrm{o}_{\mathbb{P}}(1)$ *meaning if* $X_n = \mathrm{o}_{\mathbb{P}}(1)$ *and* $Y_n = \mathrm{o}_{\mathbb{P}}(1)$ *then* $X_n + Y_n = \mathrm{o}_{\mathbb{P}}(1)$*;*

(ii) $\mathrm{O}_{\mathbb{P}}(1) + \mathrm{o}_{\mathbb{P}}(1) = \mathrm{O}_{\mathbb{P}}(1)$*;*

(iii) $\mathrm{O}_{\mathbb{P}}(1) \cdot \mathrm{o}_{\mathbb{P}}(1) = \mathrm{o}_{\mathbb{P}}(1)$*;*

(iv) $(1 + \mathrm{o}_{\mathbb{P}}(1))^{-1} = \mathrm{O}_{\mathbb{P}}(1)$*;*

(v) $\mathrm{o}_{\mathbb{P}}(\mathrm{O}_{\mathbb{P}}(1)) = \mathrm{o}_{\mathbb{P}}(1)$ *meaning if* $X_n = \mathrm{O}_{\mathbb{P}}(1)$ *and* $Y_n = \mathrm{o}_{\mathbb{P}}(X_n)$ *then* $Y_n = \mathrm{o}_{\mathbb{P}}(1)$. □

§21 Conditional expectation

In the reminder of this section let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space, $\mathbb{E}$ be the expectation with respect to $\mathbb{P}$ and $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-field of $\mathscr{A}$.

§21.01 **Notation.** We write shortly $X \in \overline{\mathscr{A}}^+$, if X is a positive numerical random variable on $(\Omega, \mathscr{A})$, i.e., $X : \Omega \to \overline{\mathbb{R}}^+$ is a $\mathscr{A}$-$\overline{\mathscr{B}}^+$-measurable function. In particular, we have $\overline{\mathscr{F}}^+ \subseteq \overline{\mathscr{A}}^+$ and for $Y \in \overline{\mathscr{F}}^+$ its expectation $\mathbb{E}(Y)$ is well-defined. □

§21.02 **Property.** *For every $X \in \overline{\mathscr{A}}^+$ exists $Y \in \overline{\mathscr{F}}^+$ with $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for all $F \in \mathscr{F}$, where Y is unique up to $\mathbb{P}$-a.s. equality.* □

§21.03 **Definition.** A map $Y : \Omega \to \overline{\mathbb{R}}^+$ is called a (version of the) *conditional expectation* of $X \in \overline{\mathscr{A}}^+$ given $\mathscr{F}$, symbolically $\mathbb{E}(X|\mathscr{F}) := Y$, if

(CE1) Y is $\mathscr{F}$-$\overline{\mathscr{B}}^+$-measurable, hence $Y \in \overline{\mathscr{F}}^+$ and

(CE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for any $F \in \mathscr{F}$.

Any map $\mathbb{E}(\bullet|\mathscr{F}) : \overline{\mathscr{A}}^+ \to \overline{\mathscr{F}}^+$ with $X \mapsto \mathbb{E}(X|\mathscr{F})$ is called (version of the) *conditional expectation* with respect to $\mathbb{P}$ given $\mathscr{F}$. It implies a map $\mathbb{P}(\bullet|\mathscr{F}) : \mathscr{A} \to \overline{\mathscr{F}}^+$ with $A \mapsto \mathbb{P}(A|\mathscr{F}) := \mathbb{E}(\mathbb{1}_A|\mathscr{F})$ called (version of the) *conditional distribution* of $\mathbb{P}$ given $\mathscr{F}$. Exploiting (CE2) every version satisfies $\mathbb{E}(\mathbb{1}_F \mathbb{P}(A|\mathscr{F})) = \int_F \mathbb{P}(A|\mathscr{F}) d\mathbb{P} = \mathbb{P}(F \cap A)$ for all $F \in \mathscr{F}$ and $A \in \mathscr{A}$. □

§21.04 **Reminder.** Let $X \in \overline{\mathscr{A}}$ be a numerical random variable. Considering the decomposition $X = X^+ - X^-$ with $X^+, X^- \in \overline{\mathscr{A}}^+$ we define for X with $\mathbb{P}(|X|) < \infty$, hence $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$, the expectation $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$. Keep in mind that $\mathcal{L}_1(\mathscr{A}, \mathbb{P}) := \{X \in \overline{\mathscr{A}} : \mathbb{E}(|X|) < \infty\}$ and $\mathbb{E} : \mathcal{L}_1(\mathscr{A}, \mathbb{P}) \to \mathbb{R}$ denotes the uniquely determined expectation with respect to $\mathbb{P}$. Note that $\overline{\mathscr{F}} \subseteq \overline{\mathscr{A}}$ implies $\mathcal{L}_1(\mathscr{F}, \mathbb{P}) \subseteq \mathcal{L}_1(\mathscr{A}, \mathbb{P})$. Let $X \in \mathcal{L}_1(\mathscr{A}, \mathbb{P})$, and hence $\mathbb{E}(X^+) < \infty$ and for any version $\mathbb{E}\big(X^+\big|\mathscr{F}\big)$ holds (CE1), $\mathbb{E}\big(X^+\big|\mathscr{F}\big) \in \overline{\mathscr{F}}^+$ and (CE2), $\mathbb{E}(\mathbb{1}_F \mathbb{E}\big(X^+\big|\mathscr{F}\big)) = \mathbb{E}(\mathbb{1}_F X^+)$ for all $F \in \mathscr{F}$, in particular with $F = \Omega$ also $\mathbb{E}(\mathbb{E}\big(X^+\big|\mathscr{F}\big)) = \mathbb{E}(X^+) < \infty$. Therewith, $\mathbb{E}\big(X^+\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ and analogously also for any version $\mathbb{E}\big(X^-\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$. Consequently, $\mathbb{E}\big(X^+\big|\mathscr{F}\big) - \mathbb{E}\big(X^-\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ satisfies (CE2) too. □

§21.05 **Definition.** For $X \in \mathcal{L}_1(\mathscr{A}, \mathbb{P})$ and each version $\mathbb{E}\big(X^+\big|\mathscr{F}\big), \mathbb{E}\big(X^-\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ we call $\mathbb{E}\big(X\big|\mathscr{F}\big) := \mathbb{E}\big(X^+\big|\mathscr{F}\big) - \mathbb{E}\big(X^-\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ a (version of the) *conditional expectation* of X given $\mathscr{F}$. Any map

$$\mathbb{E}\big(\,\bullet\,\big|\mathscr{F}\big) : \mathcal{L}_1(\mathscr{A}, \mathbb{P}) \to \mathcal{L}_1(\mathscr{F}, \mathbb{P}) \text{ with } X \mapsto \mathbb{E}\big(X\big|\mathscr{F}\big) := \mathbb{E}\big(X^+\big|\mathscr{F}\big) - \mathbb{E}\big(X^-\big|\mathscr{F}\big)$$

is called a (version of the) *conditional expectation* with respect to $\mathbb{P}$ given $\mathscr{F}$. □

§21.06 **Remark.** Due to Property §21.02 versions of the conditional expectation of $X \in \overline{\mathscr{A}}^+$ or $X \in \mathcal{L}_1(\mathscr{A}, \mathbb{P})$ given $\mathscr{F}$ differ only on null sets. This property does in generally not extend to the version of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, since for each X we obtain a null set, and their union in general is not a null set. □

§21.07 **Definition.** Let $(\Omega_1, \mathscr{A}_1)$, $(\Omega_2, \mathscr{A}_2)$ be measurable spaces. A map $\kappa : \Omega_1 \times \mathscr{A}_2 \to \mathbb{R}^+$ is called *Markov kernel* (from $(\Omega_1, \mathscr{A}_1)$ to $(\Omega_2, \mathscr{A}_2)$), if

(MK1) $A_2 \mapsto \kappa(\omega_1, A_2)$ is for all $\omega_1 \in \Omega_1$ a probability measure on $(\Omega_2, \mathscr{A}_2)$, symbolically $\kappa(\omega_1, \bullet) \in \mathcal{W}(\mathscr{A}_2)$;

(MK2) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ is $\mathscr{A}_1$-$\mathscr{B}$-measurable for all $A_2 \in \mathscr{A}_2$, symbolically $\kappa(\bullet, A_2) \in \mathscr{A}^+{}_1$. □

§21.08 **Notation.** Consider a probability space $(\Omega_1, \mathscr{A}_1, \mathbb{P})$, a measurable space $(\Omega_2, \mathscr{A}_2)$ and a Markov kernel κ (from $(\Omega_1, \mathscr{A}_1)$ to $(\Omega_2, \mathscr{A}_2)$). Then there exists an unique probability measure $\kappa \odot \mathbb{P}$ on $(\Omega_2 \times \Omega_1, \mathscr{A}_2 \otimes \mathscr{A}_1)$ determined by

$$\kappa \odot \mathbb{P}(A_2 \times A_1) = \int_{A_1} \kappa(\omega_1, A_2) \mathbb{P}(d\omega_1), \quad \text{for all } A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2.$$

If $f \in \overline{\mathscr{A}_2 \otimes \mathscr{A}_1}^+$ or $f \in \mathcal{L}_1(\kappa \odot \mathbb{P})$ then

$$\kappa \odot \mathbb{P} f = \int_{\Omega_2 \times \Omega_1} f(\omega_2, \omega_1) \kappa \odot \mathbb{P}(d\omega_2, d\omega_1) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_2, \omega_1) \kappa(\omega_1, d\omega_2) \mathbb{P}(d\omega_1).$$

Furthermore, we denote by $\kappa\mathbb{P}$ the marginal distribution on $(\Omega_2, \mathscr{A}_2)$ induced by $\kappa \odot \mathbb{P}$, i.e. $\kappa\mathbb{P}(A_2) = \kappa \odot \mathbb{P}(A_2 \times \Omega_1) = \int_{\Omega_1} \kappa(\omega_1, A_2) \mathbb{P}(d\omega_1)$ for all $A_2 \in \mathscr{A}_2$. □

§21.09 **Definition.**

(a) $\mathbb{P}\big(\bullet\big|\mathscr{F}\big)$ is called *regular* (version of the) conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, if $(\omega, A) \mapsto \mathbb{P}\big(A\big|\mathscr{F}\big)(\omega)$ satisfies the conditions (MK1) and (MK2), i.e. $\mathbb{P}\big(\,\bullet\,\big|\mathscr{F}\big)$ is a Markov kernel (from $(\Omega, \mathscr{F})$ to $(\Omega, \mathscr{A})$).

(b) $\mathbb{E}\big(\,\bullet\,\big|\mathscr{F}\big)$ is called *regular* (version of the) conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, if the implied conditional distribution $\mathbb{P}\big(\,\bullet\,\big|\mathscr{F}\big)$ of $\mathbb{P}$ given $\mathscr{F}$ is regular, and for each $\omega \in \Omega$ is $X \mapsto \mathbb{E}\big(X\big|\mathscr{F}\big)(\omega)$ the expectation with respect to $\mathbb{P}\big(\,\bullet\,\big|\mathscr{F}\big)(\omega)$. □

§21.10 **Property**.

(i) *Each regular conditional distribution of* $\mathbb{P}$ *given* $\mathscr{F}$ *is implied by a regular conditional expectation with respect to* $\mathbb{P}$ *given* $\mathscr{F}$.

(ii) *For any probability measure* $\mathbb{P}$ *on a polish space* (Ω, d) *endowed with its Borel-*σ*-algebra* $\mathscr{B}_\Omega$ *and sub-*σ*-field* $\mathscr{F} \subseteq \mathscr{B}_\Omega$ *exists a regular conditional distribution of* $\mathbb{P}$ *given* $\mathscr{F}$. □

§21.11 **Notation**.

(i) Let X be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{X}, \mathscr{X})$. For $h \in \mathcal{L}_1(\mathscr{X}, \mathbb{P}^X)$ denotes $\mathbb{E}_x\big(h\big|\mathscr{F}\big) := \mathbb{E}\big(h(X)\big|\mathscr{F}\big) \in \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ a conditional expectation of $h(X)$ given $\mathscr{F}$ and $\mathbb{E}_x\big(\bullet\big|\mathscr{F}\big) : \mathcal{L}_1(\mathscr{X}, \mathbb{P}^X) \to \mathcal{L}_1(\mathscr{F}, \mathbb{P})$ with $h \mapsto \mathbb{E}_x\big(h\big|\mathscr{F}\big)$ a *(regular)* (version of the) *conditional expectation* with respect to $\mathbb{P}^X$ given $\mathscr{F}$.

(ii) Let S be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{S}, \mathscr{S})$. For $h \in \mathcal{L}_1(\overline{\mathscr{A}}, \mathbb{P})$ we call $\mathbb{E}\big(h\big|\sigma(S)\big) \in \mathcal{L}_1(\sigma(S), \mathbb{P})$ be a conditional expectation of h given $\mathscr{F} = \sigma(S)$. Keeping $\mathbb{E}\big(h\big|\sigma(S)\big) \in \overline{\sigma(S)}$ in mind and applying Property §19.02 (iv) there is $\varphi \in \overline{\mathscr{S}}$ with $\mathbb{E}\big(h\big|\sigma(S)\big) = \varphi(S)$, that is, $\mathbb{E}\big(h\big|\sigma(S)\big)(\omega) = \varphi(S(\omega))$, $\omega \in \Omega$. Then $\mathbb{E}\big(h\big|S\big) := \varphi \in \mathcal{L}_1(\mathscr{S}, \mathbb{P}^S)$ and $\mathbb{E}\big(h\big|S = s\big) := \varphi(s) \in \overline{\mathbb{R}}$ is called a (version of the) *conditional expectation* of h given S respectively $S = s$, and $\mathbb{E}\big(\bullet\big|S\big) : \mathcal{L}_1(\mathscr{A}, \mathbb{P}) \to \mathcal{L}_1(\mathscr{S}, \mathbb{P}^S)$ with $X \mapsto \mathbb{E}\big(X\big|S\big)$ a *(regular)* (version of the) *conditional expectation* with respect to $\mathbb{P}$ given S.

(iii) Let $(X, S) : (\Omega, \mathscr{A}) \to (\mathcal{X} \times \mathcal{S}, \mathscr{X} \otimes \mathscr{S})$ with joint distribution $\mathbb{P}^{(X,S)}$. We denote by $\Pi_x : \mathcal{X} \times \mathcal{S} \to \mathcal{X}$ and $\Pi_s : \mathcal{X} \times \mathcal{S} \to \mathcal{S}$ with $(x, s) \mapsto \Pi_x(x, s) := x$ and $(x, s) \mapsto \Pi_s(x, s) := s$, respectively, the corresponding coordinate maps. The marginal distribution of X respectively S is given by $\mathbb{P}^X = \mathbb{P} \circ X^{-1} = \mathbb{P} \circ \Pi_x^{-1}(X, S) = \mathbb{P}^{(X,S)} \circ \Pi_x^{-1}$ respectively $\mathbb{P}^S = \mathbb{P}^{(X,S)} \circ \Pi_s^{-1}$. For each version $\mathbb{P}^{(X,S)}\big(\bullet\big|\sigma(\Pi_s)\big)$ of the conditional distribution with respect to $\mathbb{P}^{(X,S)}$ given $\sigma(\Pi_S)$, the map

$$\mathbb{P}^X\big(\bullet\big|S\big) : \mathscr{X} \to \overline{\mathscr{S}} \text{ with } B \mapsto \mathbb{P}^X\big(B\big|S\big) := \varphi \text{ determined by}$$
$$\mathbb{P}^X\big(B\big|\sigma(\Pi_s)\big) = \mathbb{P}^{(X,S)}\big(\Pi_x^{-1}(B)\big|\sigma(\Pi_s)\big) = \varphi(\Pi_s)$$

and analogously $\mathbb{P}^X\big(\bullet\big|S = s\big)$ is called (version of the) *conditional distribution* of X given S respectively $S = s$. We call a version *regular*, if $(s, B) \mapsto \mathbb{P}^X\big(B\big|S = s\big)$ is a Markov kernel (from $(\mathcal{S}, \mathscr{S})$ to $(\mathcal{X}, \mathscr{X})$), where due to Definition §21.03 (CE2) $\mathbb{P}^X\big(\bullet\big|S\big) \odot \mathbb{P}^S = \mathbb{P}^{(X,S)}$ (see Notation §21.08). Analogously, for $h \in \mathcal{L}_1(\mathscr{X}, \mathbb{P}^X)$ we define a (regular) version $\mathbb{E}_x\big(h\big|S\big) \in \mathcal{L}_1(\mathscr{S}, \mathbb{P}^S)$ and $\mathbb{E}_x\big(h\big|S = s\big) \in \overline{\mathbb{R}}$ of the conditional expectation of h given S respectively $S = s$. If $\mathbb{P}^X\big(\bullet\big|S\big)$ is a regular conditional distribution of X given S and for $s \in \mathcal{S}$ the probability measure $\mathbb{P}^X\big(\bullet\big|S = s\big)$ has for example a finite first absolute moment, i.e., $\mathbb{P}^X\big(\bullet\big|S = s\big) \in \mathcal{W}_1(\mathscr{B}^n)$ (see Notation §19.05) then $\mathbb{E}\big(X\big|S = s\big) = \mathbb{E}_x\big(\mathrm{id}_\mathcal{X}\big|S = s\big) = \int_\mathcal{X} x \mathbb{P}^X\big(dx\big|S = s\big)$.

(iv) Suppose the joint distribution $\mathbb{P}^{(X,S)}$ is dominated by a product measure $\mu \otimes \nu$ where μ and ν is a σ-finite measure on $\mathscr{X}$ and $\mathscr{S}$, respecitively, $\mu \in \mathcal{M}_\sigma(\mathscr{X})$ and $\nu \in \mathcal{M}_\sigma(\mathscr{S})$ for short. Let $\mathfrak{f}^{(X,S)}$ denote a $(\mu \otimes \nu)$-density of $\mathbb{P}^{(X,S)}$. A μ- and ν-density of the marginal distribution $\mathbb{P}^X$ and $\mathbb{P}^S$ is given by $\mathfrak{f}^X : x \mapsto \int_\mathcal{S} \mathfrak{f}^{(X,S)}(x, s)\nu(ds)$ and $\mathfrak{f}^S : s \mapsto \int_\mathcal{X} \mathfrak{f}^{(X,S)}(x, s)\mu(dx)$, respectively. The $\mathfrak{f}^{X|S} : \mathcal{S} \times \mathcal{X} \to \overline{\mathbb{R}}^+$ with

$$(s, x) \mapsto \mathfrak{f}^{X|S=s}(x) = \frac{\mathfrak{f}^{(X,S)}(x, s)}{\mathfrak{f}^S(s)} \mathbb{1}_{\{\mathfrak{f}^S(s)>0\}} + \mathfrak{f}^X(x) \mathbb{1}_{\{\mathfrak{f}^S(s)=0\}}$$

belongs to $\overline{\mathscr{S}\otimes\mathscr{X}}^{+}$ and it is a *μ-density of the Markov kernel* $\mathbb{P}^{X|S}$ from $(\mathcal{S},\mathscr{S})$ to $(\mathcal{X},\mathscr{X})$ defined by $(s,B)\mapsto \mathbb{P}^{X|S=s}(B) := \int_B \mathbb{f}^{X|S=s}(x)\mu(dx)$. We call $\mathbb{f}^{X|S=s}$ *conditional density* of X given $S=s$.

(v) As an example let $(X,S)\in\mathscr{B}^{k+l}$ be multivariate normally distributed with $\mathbb{C}\mathrm{ov}(X,S)=\Sigma_{XS}$ and marginal distributions $X\sim \mathrm{N}_{(\mu_X,\Sigma_X)}$ and $S\sim \mathrm{N}_{(\mu_S,\Sigma_S)}$, i.e.,

$$\begin{pmatrix}X\\S\end{pmatrix}\sim \mathrm{N}_{(\mu,\Sigma)} \text{ with } \mu=\begin{pmatrix}\mu_X\\\mu_S\end{pmatrix}\in\mathbb{R}^{k+l} \text{ and } \Sigma=\begin{pmatrix}\Sigma_X & \Sigma_{XS}\\ \Sigma_{XS}^t & \Sigma_S\end{pmatrix}.$$

Assuming $\Sigma>0$ the joint distribution $\mathbb{P}^{(X,S)}$ admits a density with respect to the Lebesgue measure λ^{k+l} on $(\mathbb{R}^{k+l},\mathscr{B}^{k+l})$. For each $s\in\mathbb{R}^l$ the conditional density $\mathbb{f}^{X|S=s}$ as in (iv) is a density of the multivariate normal distribution $\mathrm{N}_{(\mu_{X|S=s},\Sigma_{X|S=s})}$-distribution with

$$\mu_{X|S=s} := \mu_X+\Sigma_{XS}\Sigma_S^{-1}(s-\mu_S)\in\mathbb{R}^k \text{ und } \Sigma_{X|S=s} := \Sigma_X-\Sigma_{XS}\Sigma_S^{-1}\Sigma_{SX}>0$$

which is thus a regular conditional distribution of X given $S=s$. □

§21.12 **Property.** *Let $X,Y\in\mathcal{L}_1(\mathscr{A},\mathbb{P})$ and $\mathscr{F}\subseteq\mathscr{A}$ be a sub-σ-field. Any version of the conditional expectation satisfies the following properties $\mathbb{P}$-a.s.:*

(i) *For all $a,b\in\mathbb{R}$ holds $\mathbb{E}\big(aX+bY\big|\mathscr{F}\big)=a\mathbb{E}\big(X\big|\mathscr{F}\big)+b\mathbb{E}\big(Y\big|\mathscr{F}\big)$;* *(linear)*

(ii) *For $X\leqslant Y$ holds $\mathbb{E}\big(X\big|\mathscr{F}\big)\leqslant\mathbb{E}\big(Y\big|\mathscr{F}\big)$;* *(monotone)*

(iii) $|\mathbb{E}\big(X\big|\mathscr{F}\big)|\leqslant\mathbb{E}\big(|X|\big|\mathscr{F}\big)$; *(triangular inequality)*

(iv) *For $S\in\overline{\mathscr{A}}$ with $\mathbb{E}\big(|S|\big|\mathscr{F}\big)<\infty$ holds $\mathbb{P}(|S|<\infty)=1$.* *(finite)*

(v) *For $\phi:\mathbb{R}\to\mathbb{R}$ convex with $\phi(X)\in\mathcal{L}_1(\overline{\mathscr{A}},\mathbb{P})$ holds $\phi\big(\mathbb{E}\big(X\big|\mathscr{F}\big)\big)\leqslant\mathbb{E}\big(\big(\phi(X)\big)\big|\mathscr{F}\big)$.* *(Jensen's inequality)*

(vi) *For $X_n\uparrow X$ $\mathbb{P}$-a.s. holds $\sup_{n\in\mathbb{N}}\mathbb{E}\big(X_n\big|\mathscr{F}\big)=\mathbb{E}\big(X\big|\mathscr{F}\big)$.* *(monotone convergence)*

(vii) *For $X_n\to X$ $\mathbb{P}$-a.s. with $|X_n|\leqslant Y$, $n\in\mathbb{N}$, holds $\lim_{n\to\infty}\mathbb{E}\big(X_n\big|\mathscr{F}\big)=\mathbb{E}\big(X\big|\mathscr{F}\big)$ $\mathbb{P}$-a.s. and in $\mathcal{L}_1(\mathscr{A},\mathbb{P})$.* *(dominated convergence)*

If the version is regular, i.e., $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big)(\omega)$ is an expectation for all $\omega\in\Omega$, then the statements (i)-(vii) holds for all $\omega\in\Omega$. □

§21.13 **Property.** *Let $X,Y\in\mathcal{L}_1(\mathscr{A},\mathbb{P})$ and $\mathscr{G}\subseteq\mathscr{F}\subseteq\mathscr{A}$ sub-σ-fields. Any version of the conditional expectation satisfies the following properties $\mathbb{P}$-a.s.:*

(i) *For $\mathbb{E}(|XY|)<\infty$ and $Y\in\mathscr{F}$ holds*

$$\mathbb{E}\big(XY\big|\mathscr{F}\big)=Y\mathbb{E}\big(X\big|\mathscr{F}\big) \text{ and } \mathbb{E}\big(Y\big|\mathscr{F}\big)=\mathbb{E}\big(Y\big|\sigma(Y)\big)=Y;$$

(ii) $\mathbb{E}\big(\mathbb{E}\big(X\big|\mathscr{F}\big)\big|\mathscr{G}\big)=\mathbb{E}\big(\mathbb{E}\big(X\big|\mathscr{G}\big)\big|\mathscr{F}\big)=\mathbb{E}\big(X\big|\mathscr{G}\big)$; *(tower property)*

(iii) *If $\sigma(X)$ and $\mathscr{F}$ are independent, then $\mathbb{E}\big(X\big|\mathscr{F}\big)=\mathbb{E}(X)$;* *(independence)*

(iv) $\mathbb{E}(\mathbb{E}\big(X\big|\mathscr{F}\big))=\mathbb{E}(X)$. *(total probability)*

(v) *For $\overline{\mathscr{T}}:=\big\{A\in\mathscr{A}\,\big|\,\mathbb{P}(A)\in\{0,1\}\big\}$ holds $\mathbb{E}\big(X\big|\overline{\mathscr{T}}\big)=\mathbb{E}(X)$.* □

§21.14 **Property.** *Let $\mathscr{F}\subseteq\mathscr{A}$ be a sub-σ-field and $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big)$ be a conditional expectation.*

(i) *$\mathbb{E}\big(\bullet\,\big|\overline{\mathscr{F}}\big):\mathcal{L}_2(\mathscr{A},\mathbb{P})\to\mathcal{L}_2(\mathscr{F},\mathbb{P})$ is an orthogonal projection, that is, for all $X\in\mathcal{L}_2(\mathscr{A},\mathbb{P})$ and $Y\in\mathcal{L}_2(\mathscr{F},\mathbb{P})$ holds*

$$\|X-Y\|^2_{\mathcal{L}_2(\mathbb{P})}=\mathbb{E}(|X-Y|^2)\geqslant\mathbb{E}(|X-\mathbb{E}\big(X\big|\mathscr{F}\big)|^2)=\|X-\mathbb{E}\big(X\big|\mathscr{F}\big)\|^2_{\mathcal{L}_2(\mathbb{P})},$$

where equality holds if and only if $Y=\mathbb{E}\big(X\big|\mathscr{F}\big)$ $\mathbb{P}$-a.s..

(ii) $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big) : \mathcal{L}_s(\mathscr{A},\mathbb{P}) \to \mathcal{L}_s(\mathscr{F},\mathbb{P})$ *is a contraction for* $s \in [1,\infty]$, *i.e.,* $\|\mathbb{E}\big(X\big|\mathscr{F}\big)\|_{\mathcal{L}_s(\mathbb{P})} \leqslant \|X\|_{\mathcal{L}_s(\mathbb{P})}$, *and thus bounded and continuous. If* $(X_n)_{n\in\mathbb{N}}$ *converges in* $\mathcal{L}_s(\mathscr{A},\mathbb{P})$, *then* $(\mathbb{E}\big(X_n\big|\mathscr{F}\big))_{n\in\mathbb{N}}$ *converges in* $\mathcal{L}_s(\mathscr{F},\mathbb{P})$. □

Bibliography

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999.

L. Birgé. An alternative point of view on Lepskij's method. In Monogr., editor, *State of the art in probability and statistics*, volume 36 of *IMS Lecture Notes*, pages 113–133. (Leiden 1999), 2001.

L. D. Brown and M. G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *The Annals of Statistics*, 24(6):2524–2535, 1996.

N. L. Carr. Kinetics of catalytic isomerization of n-pentane. *Industrial and Engineering Chemistry*, 52:391–396, 1960.

F. Comte. *Estimation non-paramétrique*. Spartacus-idh, Paris, 2015.

A. Goldenshluger and O. Lepskij. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39:1608–1632, 2011.

T. Kawata. *Fourier analysis in probability theory.* Academic Press, New York, 1972.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, 2005.

A. Klenke. *Probability theory. A comprehensive course.* London: Springer, 2008.

A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Spektrum, 3., überarbeitete und ergänzte Auflage, 2012.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

B. Laurent, C. Ludena, and C. Prieur. Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2(993-1020), 2008.

L. Le Cam. Convergence of Estimates Under Dimensionality Restrictions. *The Annals of Statistics*, 1(1):38–53, 1973.

O. Lepskij. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, 35(3):454–466, 1990.

O. Lepskij. Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory of Probability and its Applications*, 36(4):682–697, 1991.

O. Lepskij. Asymptotically minimax adaptive estimation. II: Schemes without optimal adaption: Adaptive estimators. *Theory of Probability and its Applications*, 37(3):433–448, 1992a.

O. Lepskij. On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, volume 12 of *Pap. Semin. Math. Stat.*, pages 87–106, Moscow/USSR, 1992b. Adv. Sov. Math.

H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1):50–60, 1947.

P. Massart. *Concentration inequalities and model selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003.* Lecture Notes in Mathematics 1896. Berlin: Springer., 2007.

M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126:505–563, 1996.

A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York, 2009.

A. W. van der Vaart. *Asymptotic statistics.* Cambridge University Press, 1998.

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

H. Witting. Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang. Stuttgart: B. G. Teubner, 1985.

H. Witting and U. Müller-Funk. *Mathematische Statistik II. Asymptotische Statistik: Parametrische Modelle und nichtparametrische Funktionale.* Stuttgart: B. G. Teubner, 1995.