



Skript zur Vorlesung

STATISTIK 1

Wintersemester 2021/22

Fassung Stand 7. Februar 2022

Falls Sie **Fehler im Skript** finden, teilen Sie mir diese bitte
per eMail an johannes@math.uni-heidelberg.de mit.

MATHEMATIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
Telefon: +49 6221 54.14.190 – Fax: +49 6221 54.14.101
eMail: johannes@math.uni-heidelberg.de
Webseite: sip.math.uni-heidelberg.de

Inhaltsverzeichnis

1	Statistische Inferenz im linearen Modell	1
§01	Das lineare Modell	1
§02	Methode der kleinsten Quadrate	5
§03	Das normale lineare Modell	8
§04	Asymptotische Theorie und Residuenanalyse	11
2	Entscheidungstheorie	15
§05	Formalisierung eines statistischen Problem	15
§06	Minimax- und Bayes-Ansatz	18
§07	Das Stein-Phänomen	23
3	Schätztheorie	25
§08	Dominierte Modelle	25
§09	Erschöpfende Statistik	26
§10	Exponentialfamilien	32
§11	Vollständige Statistik	34
§12	Erwartungstreue Schätzer	36
§13	Informationsungleichungen	37
§14	Translations-äquivalente Schätzer	41
4	Allgemeine Schätzmethoden	43
§15	Momentenschätzer	43
§16	Maximum-Likelihood-Schätzer	45
§17	Minimum-Kontrast-Schätzer	47
5	Testtheorie	53
§18	Neyman-Pearson-Theorie	53
§19	Bedingte Tests	60
§20	Likelihood-Quotienten-Test	65
§21	Rangtest	69
6	Nonparametric estimation	73
§22	Introduction	73
§23	Kernel density estimation	74
§24	Nonparametric regression by local smoothing	78
A	Grundlagen der Wahrscheinlichkeitstheorie	83
A01	Wahrscheinlichkeitsraum	83
A02	Zufallsvariablen	86
A03	Unabhängigkeit	88
A04	Erwartung	89
A05	Multivariate Normalverteilung	93
A06	Grenzwertsätze	97
B	Elementare Maß- und Integrationstheorie	101
B01	Maßraum	101

B02	Integral	102
B03	Konvergenzarten	105
B04	Maße mit Dichten - Satz von Radon-Nikodym	106
B05	Endliche Produktmaße	108
B06	Übergangskerne	109
C	Bedingte Erwartung	111
C01	Diskret- oder stetig-verteilte Zufallsvariablen	111
C02	Positive numerische Zufallsvariablen	113
C03	Integrierbare Zufallsvariablen	121

1 Statistische Inferenz im linearen Modell

§01 Das lineare Modell

§01.01 **Beispiel.** In der folgenden Tabelle ist ein Auszug des „Cars93“ Datensatzes aus dem Statistikpaket R Core Team [2015] (library {MASS}) angegeben. Der Datensatz umfasst unter anderem den Preis, die Anzahl der Zylinder (Zyl.), den Hubraum (Hub.), die Breite sowie das Herkunftsland für 93 in den USA im Jahr 1993 verkauften Autos.

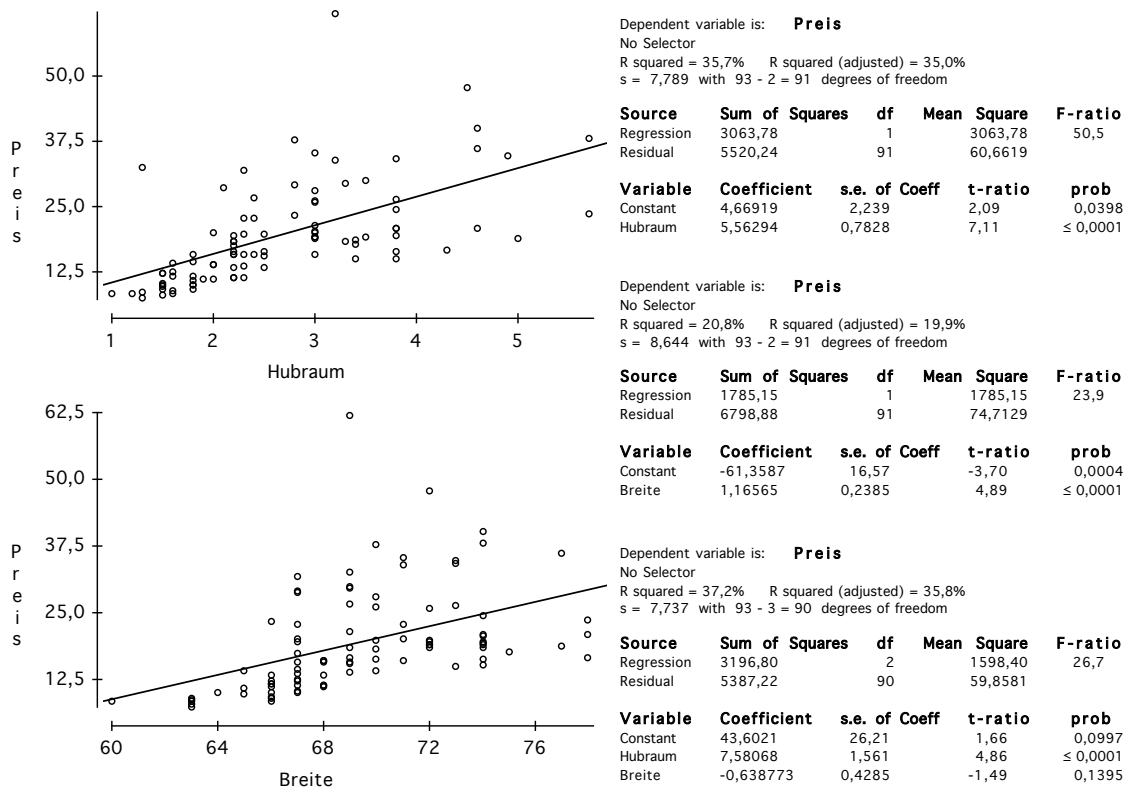
Preis	Zyl.	Hub.	Breite	Herkunft	Preis	Zyl.	Hub.	Breite	Herkunft
15.9	4	1.8	68	non-USA	10	4	1.5	64	non-USA
33.9	6	3.2	71	non-USA	13.9	4	2	69	non-USA
29.1	6	2.8	67	non-USA	47.9	8	4.5	72	non-USA
37.7	6	2.8	70	non-USA	28	6	3	70	non-USA
30	4	3.5	69	non-USA	35.2	6	3	71	non-USA
15.7	4	2.2	69	USA	34.3	6	3.8	73	USA
20.8	6	3.8	74	USA	36.1	8	4.6	77	USA
23.7	6	5.7	78	USA	8.3	4	1.6	66	non-USA
26.3	6	3.8	73	USA	11.6	4	1.8	66	non-USA
34.7	8	4.9	73	USA	16.5	4	2.5	69	non-USA
40.1	8	4.6	74	USA	19.1	6	3	72	non-USA
11.4	4	2.2	68	USA	31.9	4	2.3	67	non-USA
15.1	6	3.4	74	USA	61.9	6	3.2	69	non-USA
15.9	4	2.2	71	USA	14.1	4	1.6	65	USA
16.3	6	3.8	74	USA	14.9	6	3.8	73	USA
16.6	6	4.3	78	USA	10.3	4	1.5	67	non-USA

Preis, Anzahl der Zylinder (Zyl.), Hubraum (Hub.), Breite sowie Herkunftsland von in den USA verkauften Autos.

Sei Y_i der Preis des i -ten Autos mit Hubraum z_{1i} und Breite z_{2i} . Wir nehmen an, die Autos seien austauschbar und es existiert ein linearer Zusammenhang (vgl. nachfolgende Graphik) zwischen dem erwarteten Verhalten des Preises und den erklärenden Variablen Hubraum und Breite:

$$\mathbb{E}Y_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i}, \quad i = 1, \dots, 93.$$

Wir möchten statistische Aussagen über die Parameter β_1 und β_2 treffen, wie zum Beispiel die Werte der Parameter schätzen, Hypothesen der Form $\beta_1 = 0$ oder $\beta_2 = 0$ verifizieren oder den zu Grunde gelegten linearen Zusammenhang überprüfen.



Preis in Abhängigkeit des Hubraumes bzw. der Breite des Autos.

□

§01.02 **Einfache lineare Regression.** Zu einem vorgegeben (nicht zufälligem) Versuchsplan $z_i \in \mathbb{R}$, $i \in \llbracket n \rrbracket := [1, n] \cap \mathbb{N}$ beobachten wir Realisierungen der reellen Zufallsvariablen

$$Y_i = a + bz_i + \varepsilon_i, \quad i \in \llbracket n \rrbracket,$$

wobei die Zufallsvariablen $(\varepsilon_i)_{i \in \llbracket n \rrbracket}$ Messfehler modellieren und $a, b \in \mathbb{R}$ unbekannte Parameter sind. Man denke z.B. an Messungen der Leitfähigkeit Y_i eines Stoffes in Abhängigkeit der Temperatur z_i , eines Effektes Y_i in Abhängigkeit einer Dosierung z_i oder eines Klausurergebnisses Y_i in Abhängigkeit der Klassengröße z_i . Sind die Messfehler $\varepsilon_i \in \mathcal{L}_1$, $i \in \llbracket n \rrbracket$, zentriert, das heißt $\mathbb{E}(\varepsilon_i) = 0$, $i \in \llbracket n \rrbracket$, so gilt offensichtlich

$$\mathbb{E}(Y_i) = a + bz_i, \quad i \in \llbracket n \rrbracket,$$

so dass ein linearer Zusammenhang nur zwischen der erklärenden Variable z_i und der Erwartung der zu erklärenden zufälligen Größe Y_i zu Grunde gelegt wird. Betrachten wir weiterhin die \mathbb{R}^n -wertigen Zufallsvektoren $Y = (Y_1, \dots, Y_n)^t$ und $\varepsilon := (\varepsilon_1, \dots, \varepsilon_n)^t$ (aufgefasst als Spaltenvektoren), kurz $Y, \varepsilon \in \mathcal{A}^n$, den unbekanntem Parametervektor $\beta = (a, b)^t \in \mathbb{R}^2$ sowie die vorgegebene (Design-)Matrix $X = (x_1, \dots, x_n)^t \in \mathbb{R}^{(n,2)}$ mit Zeilen $x_i^t = (1, z_i)$, $i \in \llbracket n \rrbracket$, dann lässt sich die einfache lineare Regression kompakt in der Form $Y = X\beta + \varepsilon$ schreiben. Für $\varepsilon \in \mathcal{L}_2$ und somit $Y \in \mathcal{L}_2$ bezeichnen wir weiterhin mit $\Sigma = \text{Cov}(\varepsilon) = \Sigma(Y) \in \mathbb{R}^{(n,n)}$ die Kovarianzmatrix von ε und Y , d.h. für den Eintrag Σ_{ij} in der i -ten Zeile und j -ten Spalte von $\Sigma := (\Sigma_{ij})_{i,j \in \llbracket 1,n \rrbracket}$ gilt $\Sigma_{ij} = \text{Cov}(Y_i, Y_j) = \text{Cov}(\varepsilon_i, \varepsilon_j) = \mathbb{E}(\varepsilon_i \varepsilon_j)$. Bezeichnet $\langle v, w \rangle = w^t v$ für $v, w \in \mathbb{R}^2$ das euklidische Skalarprodukt, dann gilt $\text{Cov}(\langle \varepsilon, v \rangle, \langle \varepsilon, w \rangle) = \langle \Sigma v, w \rangle$ (siehe Definition A04.07).

□

§01.03 **Erinnerung.** Wir schreiben $\Sigma > 0$ oder $\Sigma \in \mathbb{R}_{>}^{(n,n)}$, falls $\Sigma \in \mathbb{R}^{(n,n)}$ eine symmetrische, strikt positiv-definite Matrix ist. Insbesondere, ist dann Σ diagonalisierbar mit $\Sigma = U\Lambda U^t$ für eine

Diagonalmatrix $\Lambda = \text{diag}(\lambda)$ mit Diagonaleinträgen $\lambda = (\lambda_i)_{i \in \llbracket 1, n \rrbracket}$ aus $\mathbb{R}_0^+ := (0, \infty)$ und eine unitäre Matrix U . Für $s \in \mathbb{R}$ setzen wir $\Sigma^s = U\Lambda^s U^t$ wobei $\Lambda^s = \text{diag}(\lambda^s)$ mit $\lambda^s = (\lambda_i^s)_{i \in \llbracket 1, n \rrbracket}$. Wie erwartet, gilt $(\Sigma^{-1/2})^2 = \Sigma^{-1}$ und somit $\|\Sigma^{-1/2}v\|^2 = \langle \Sigma^{-1}v, v \rangle$. \square

§01.04 **Schreibweise.** Wir haben im **Beispiel** §01.02 angenommen, dass die unbekannte gemeinsame Verteilung \mathbb{P} des Zufallsvektors $Y \in \mathcal{X}^n$, kurz $Y \sim \mathbb{P}$, ein endliches zweites Moment besitzt, also $\mathbb{P} \in \mathcal{W}_2(\mathcal{B}^n)$ gilt (vgl. **Schreibweise** A04.08). Wir schreiben in dieser Situation abkürzend $Y \odot \mathcal{W}_2(\mathcal{B}^n)$. Weiterhin erfüllt Y gerade $\mathbb{E}_{\mathbb{P}}(Y) \in \text{Bild}(X) = X\mathbb{R}^p$, also $\mathbb{E}_{\mathbb{P}}(Y) = X\beta$ für ein $\beta \in \mathbb{R}^p$, sowie $\text{Cov}_{\mathbb{P}}(Y) \in \mathbb{R}_>^{(n,n)}$. Wir fassen unter $\text{IM}_{(X,\beta,\Sigma)}$ alle $\mathbb{P} \in \mathcal{W}_2(\mathcal{B}^n)$ mit $\mathbb{E}_{\mathbb{P}}(Y) = X\beta$ und $\text{Cov}_{\mathbb{P}}(Y) = \Sigma$ zusammen und die abkürzende Schreibweise $Y \odot \text{IM}_{(X,\beta,\Sigma)}$ ist so zu verstehen, dass es ein $\mathbb{P} \in \text{IM}_{(X,\beta,\Sigma)}$ mit $Y \sim \mathbb{P}$ gibt. Setzen wir weiterhin $\text{IM}_{X\mathbb{R}^p \times \mathbb{R}_>^{(n,n)}} := (\text{IM}_{(X,\beta,\Sigma)})_{(\beta,\Sigma) \in \mathbb{R}^p \times \mathbb{R}_>^{(n,n)}}$, so meint die Schreibweise $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_>^{(n,n)}}$, dass es $\beta \in \mathbb{R}^p$ und $\Sigma \in \mathbb{R}_>^{(n,n)}$ mit $Y \odot \text{IM}_{(X,\beta,\Sigma)}$ gibt. \square

§01.05 **Definition.** Ein *lineares Modell* beschreibt *adäquat* den Zusammenhang zwischen einem zu erklärenden Zufallsvektors (Zielgröße) $Y \in \mathcal{X}^n$ in \mathcal{L}_2 und einer erklärenden, vorgegebenen Matrix $X \in \mathbb{R}^{(n,p)}$, der *Designmatrix* oder Matrix der Effekte, falls ein Parametervektor $\beta \in \mathbb{R}^p$ existiert, so dass $\mathbb{E}(Y) = X\beta$ gilt. Die Kovarianzmatrix $\Sigma = \text{Cov}(\varepsilon) \in \mathbb{R}^{(n,n)}$ des zentrierten zufälligen Vektors $\varepsilon := Y - X\beta$, den *Fehler- oder Störgrößen*, sowie der Vektor $\beta \in \mathbb{R}^p$ sind unbekannte Parameter in einem linearem Modell. Beobachtet wird eine Realisierung von Y und die Designmatrix X und wir schreiben abkürzend $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_>^{(n,n)}}$. In einem *gewöhnlichen linearen Modell* gilt weiterhin $\Sigma = \sigma^2 E_n$ für ein Fehlerniveau $\sigma \in \mathbb{R}_0^+$, wobei $E_n \in \mathbb{R}^{(n,n)}$ die Einheitsmatrix bezeichnet. Wir schreiben abkürzend $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+} := (\text{IM}_{(X,\beta,\sigma^2 E_n)})_{(\beta,\sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+}$. \square

§01.06 **Beispiel.**

(a) Ein Zufallsvektor $Y \in \mathcal{A}^n$ folgt einem *Lokations-Skalen-Modell*, falls $\mathbb{E}(Y) = \mu \mathbb{1}_n$ mit $\mathbb{1}_n := (1, \dots, 1)^t \in \mathbb{R}^n$ und $\text{Cov}(Y) = \sigma^2 E_n$ gilt. Die unbekannt Parameter sind $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_0^+$. Wir schreiben $Y \odot \text{IM}_{\mathbb{1}_n \mathbb{R} \times \mathbb{R}_0^+} := (\text{IM}_{(\mu \mathbb{1}_n, \sigma^2 E_n)})_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$. Sind die Koordinaten von Y zusätzlich unabhängige und identisch \mathbb{P} -verteilte (u.i.v.) reelle Zufallsvariablen, so ist die Verteilung von Y durch das Produkt $\mathbb{P}^n := \bigotimes_{i \in \llbracket n \rrbracket} \mathbb{P}$ der identischen eindimensionalen Randverteilungen gegeben und wir schreiben $\text{IM}_{\mu, \sigma^2}^n := \{\mathbb{P}^n \in \text{IM}_{(\mu \mathbb{1}_n, \sigma^2 E_n)}\}$ sowie $Y \odot \text{IM}_{\mathbb{R} \times \mathbb{R}_0^+}^n := (\text{IM}_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$. Wird die Varianz $\sigma_o^2 \in \mathbb{R}_0^+$ der Beobachtungen als bekannt vorausgesetzt, so erfüllt der zufällige Vektor Y ein *Lokations-Modell* und wir schreiben abkürzend $Y \odot \text{IM}_{\mathbb{1}_n \mathbb{R} \times \{\sigma_o^2\}}$ oder $Y \odot \text{IM}_{\mathbb{R} \times \{\sigma_o^2\}}^n$. Wird dagegen der Erwartungswert $\mu_o \in \mathbb{R}$ als bekannt vorausgesetzt, so folgt der Zufallsvektor Y einem *Skalen-Modell* und wir schreiben abkürzend $Y \odot \text{IM}_{\{\mu_o \mathbb{1}_n\} \times \mathbb{R}_0^+}$ oder $Y \odot \text{IM}_{\{\mu_o\} \times \mathbb{R}_0^+}^n$. Setzen wir $\beta = \mu$ und $X = \mathbb{1}_n$ so sind die drei Modelle offensichtlich (gewöhnliche) lineare Modelle.

(b) *Varianzanalyse mit einem Faktor.* Es werden p Proben an q Labore geschickt, wir erhalten zu jeder Probe einen Messwert, den wir als Realisierung von Zufallsvariablen

$$Y_{jk} = \mu_j + \varepsilon_{jk}, \quad j \in \llbracket p \rrbracket, \quad k \in \llbracket q \rrbracket,$$

auffassen. Ein Anordnen der Zufallsvariablen als $n = pq$ dimensionalen Vektor, $Y = (Y_1, \dots, Y_n)^t \in \mathcal{A}^n$ mit $Y_i = Y_{jk}$ und $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t \in \mathcal{A}^n$ mit $\varepsilon_i = \varepsilon_{jk}$ für $i = k + (j - 1)q$ erlaubt es uns, kompakt $Y = X\beta + \varepsilon$ zu schreiben, wobei $\beta := (\mu_1, \dots, \mu_p)^t \in \mathbb{R}^p$ und $X = E_p \otimes \mathbb{1}_q \in \mathbb{R}^{(n,p)}$. Hier bezeichnet \otimes das Kronecker-Produkt, d.h. $A \otimes B := (a_{ij} B)$ für zwei Matrizen $A = (a_{ij})$ und B . Insbesondere, folgt also der zufällige Vektor Y einem linearen Modell.

- (c) Der Zusammenhang zwischen vorgegebenen Designpunkten $(z_i)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$ und einem Zufallsvektor $Y \in \mathcal{A}^n$ wird durch eine *polynomiale Regression* beschrieben, falls Parameter $(a_i)_{i \in \llbracket 0, p-1 \rrbracket} \in \mathbb{R}^p$ existieren, so dass

$$\mathbb{E}(Y_i) = a_0 + a_1 z_i + a_2 z_i^2 + \dots + a_{p-1} z_i^{p-1}, \quad i \in \llbracket n \rrbracket,$$

gilt. Bezeichnen wir mit $\beta = (a_i)_{i \in \llbracket 0, p-1 \rrbracket}^t$ den (Spalten-)Vektor der unbekannt Parameter und mit

$$X = \begin{pmatrix} 1 & z_1 & z_1^2 & \dots & z_1^{p-1} \\ 1 & z_2 & z_2^2 & \dots & z_2^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & z_n^2 & \dots & z_n^{p-1} \end{pmatrix}$$

die Designmatrix vom Vandermonde-Typ, so gilt $\mathbb{E}(Y) = X\beta$ und es liegt somit ein lineares Modell vor. Für $p > 2$ ist der Zusammenhang zwischen den Designpunkten $\{z_i\}$ und den Beobachtungen $\{Y_i\}$ insbesondere nichtlinear. Auf Grund der linearen Abhängigkeit vom Parametervektor β wird das Modell linear genannt. Eine natürliche Verallgemeinerung der Modellierung eines nichtlinearer Zusammenhang zwischen den Designpunkten $\{z_i\}$ und den Beobachtungen $\{Y_i\}$ ist

$$\mathbb{E}(Y_i) = \beta_1 \psi_1(z_i) + \dots + \beta_p \psi_p(z_i), \quad i \in \llbracket n \rrbracket,$$

mit unbekanntem Parametervektor $\beta = (\beta_i)_{i \in \llbracket p \rrbracket}^t$ und vorgegebene Basisfunktionen $\{\psi_j\}$, zum Beispiel Splinefunktionen. Setzen wir $X := (\psi_k(z_j))_{jk}$ so gilt erneut $\mathbb{E}(Y) = X\beta$ und das zugrunde liegende Modell ist linear. □

§01.07 **Erinnerung.** Ausgehend von $\text{IM}_{X \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}} = (\text{IM}_{(X,\Sigma)})_{(\beta,\Sigma) \in \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ bezeichnen wir eine Abbildung $\gamma : \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)} \rightarrow \Gamma$ als *abgeleiteten* oder *interessierenden Parameter* und für jedes $(\beta, \Sigma) \in \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}$ wird $\gamma(\beta, \Sigma)$ *abgeleiteter* oder *interessierender Parameterwert* genannt. Häufig benutzen wir das Symbol γ sowohl für den abgeleiteten Parameter, also die Abbildung, als auch für die Elemente von Γ , also die Parameterwerte. □

§01.08 **Definition.** In einem linearen Modell $Y \odot \text{IM}_{X \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ heißt der Parameter $\beta \in \mathbb{R}^p$ oder allgemeiner ein abgeleiteter Parameter $\gamma : \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)} \rightarrow \Gamma$ *identifizierbar*, falls für beliebige $(\beta_1, \Sigma_1), (\beta_2, \Sigma_2) \in \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}$ aus $\gamma(\beta_1, \Sigma_1) \neq \gamma(\beta_2, \Sigma_2)$ folgt $\text{IM}_{(X,\beta_1,\Sigma_1)} \neq \text{IM}_{(X,\beta_2,\Sigma_2)}$. □

§01.09 **Lemma.** Sei $Y \odot \text{IM}_{X \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ und $C \in \mathbb{R}^{(q,p)}$ eine vorgegebene Matrix. Der *abgeleitete lineare Parameter* $\gamma : \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)} \rightarrow \mathbb{R}^q$ mit $(\beta, \Sigma) \mapsto \gamma(\beta, \Sigma) := C\beta$ ist genau dann identifizierbar, wenn eine Matrix $A \in \mathbb{R}^{(q,n)}$ mit $C = AX$ existiert.

§01.10 **Beweis von Lemma §01.09.** In der Vorlesung. □

§01.11 **Korollar.** In einem linearen Modell $Y \odot \text{IM}_{X \mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ ist der Parameter $\beta \in \mathbb{R}^p$ genau dann identifizierbar, wenn die Designmatrix $X \in \mathbb{R}^{(n,p)}$ den Rang $\text{rg}[X] = p$ besitzt.

§01.12 **Beweis von Korollar §01.11.** In der Vorlesung. □

§01.13 **Bemerkung.** Sei $Y \odot \text{IM}_{(X,\Sigma)}$. Besitzt die Designmatrix X den Rang $\text{rg}[X] = r < p$, so lässt sich durch eine geeignete Transformation $\gamma = C\beta$ und $\tilde{X} = XU$ für $C \in \mathbb{R}^{r \times p}$ und $U \in \mathbb{R}^{p \times r}$ erreichen, dass $\gamma \in \mathbb{R}^r$ in dem reparametrisierten linearen Modell $Y \odot \text{IM}_{\tilde{X} \mathbb{R}^r \times \mathbb{R}_{>}^{(n,n)}} = (\text{IM}_{\tilde{X},\Sigma})_{(\gamma,\Sigma) \in \mathbb{R}^r \times \mathbb{R}_{>}^{(n,n)}}$ identifizierbar ist. Dies ist genau dann der Fall, wenn $XUC = X$ und $\text{rg}[XU] = r$ gilt. □

§01.14 **Beispiel.**

- (a) (*Einfache lineare Regression Beispiel §01.02 fortgesetzt.*) Die Parameter a und b sind identifizierbar, falls mindestens zwei Effekte des Versuchsplans $\{z_i\}$ verschieden sind.
- (b) (*Polynomiale Regression Beispiel §01.06 (c) fortgesetzt.*) Die Determinante einer Matrix vom Vandermonde-Typ ist im Fall $p = n$ gegeben durch $\prod_{p \geq k > j \geq 1} (z_k - z_j)$. Damit ist eine hinreichende und notwendige Bedingung für die Identifizierbarkeit des Parameters β , dass mindestens p verschiedene Effekte existieren. \square

§01.15 **Bemerkung.** Es gibt wichtige *Verallgemeinerungen linearer Modelle* (GLM für *Generalized Linear Model*). Der Zusammenhang zwischen einem zufälligem Vektor $Y \in \mathcal{A}^n$ und einer Designmatrix $X = (x_1, \dots, x_n)^t \in \mathbb{R}^{(n,p)}$ ist durch ein *verallgemeinertes lineares Modell* mit vorgegebener Linkfunktion ℓ beschrieben, falls ein Parametervektor $\beta \in \mathbb{R}^p$ existiert, so dass $\mathbb{E}(Y_i) = \ell(x_i^t \beta)$, $i \in \llbracket n \rrbracket$, gilt. Nehmen wir an, dass die Zufallsvariable Y_i das Auftreten eines positiven oder negativen Effektes nach Verabreichung eines Medikamentes wiedergibt. In diesem Fall ist $Y_i \sim B_{\pi_i}$ eine Bernoulli-Zufallsvariable und die Erfolgswahrscheinlichkeit π_i der unbekanntem Parameter. Eine *logistische Regression* liegt nun vor, falls ein Parametervektor $\beta \in \mathbb{R}^p$ existiert, so dass $\log(\pi_i/(1 - \pi_i)) = x_i^t \beta$ oder äquivalent $\pi_i = \{1 + \exp(-x_i^t \beta)\}^{-1}$ für $i \in \llbracket n \rrbracket$ gilt. Die Linkfunktion $\ell(x) = \{1 + \exp(-x)\}^{-1}$, $x \in \mathbb{R}$, entspricht gerade der logistischen Verteilungsfunktion, so dass wir auch von einem Logitmodell sprechen. Ein weiteres Beispiel, ist das Probitmodell, in dem ℓ der Verteilungsfunktion einer Standardnormalverteilung entspricht. \square

§02 Methode der kleinsten Quadrate

Zur Erinnerung, im Sinne des mittleren quadratischen Fehlers (MSE für *mean squared error*) ist die beste konstante Approximation einer reellen Zufallsvariablen Z in \mathcal{L}_1 ihr Erwartungswert $\mu = \mathbb{E}(Z)$, d.h. $\mathbb{E}(Z - \mu)^2 = \min_{a \in \mathbb{R}} \mathbb{E}(Z - a)^2$. Das folgende Lemma verallgemeinert diesen Sachverhalt und motiviert zudem die Methode der kleinsten Quadrate.

§02.01 **Lemma.** Für $Y \odot \text{IM}_{X,\beta,\Sigma}$ gilt

$$\beta \in \arg \min_{b \in \mathbb{R}^p} \mathbb{E} \|\Sigma^{-1/2}(Y - Xb)\|^2$$

$$:\Leftrightarrow \mathbb{E} \|\Sigma^{-1/2}(Y - X\beta)\|^2 = \min_{b \in \mathbb{R}^p} \mathbb{E} \|\Sigma^{-1/2}(Y - Xb)\|^2. \quad (02.01)$$

§02.02 **Beweis** von Lemma §02.01. In der Vorlesung. \square

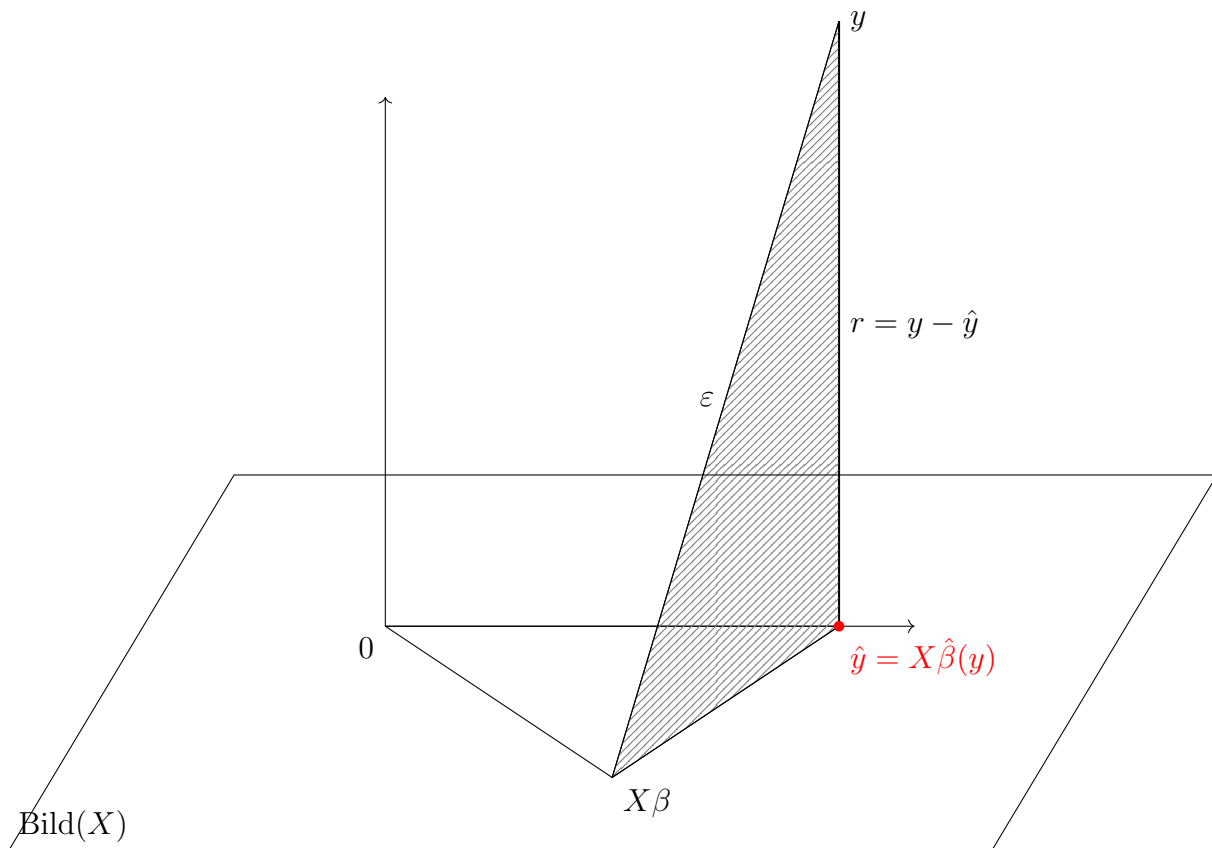
§02.03 **Erinnerung.** In einem linearen Modell $Y \odot \text{IM}_{X_{\mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ heißt jede Zufallsvariable $\hat{\beta} : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^p, \mathcal{B}^p)$ *Schätzfunktion*, kurz *Schätzer*, für den Parameter $\beta \in \mathbb{R}^p$. Für eine Stichprobe $y \in \mathbb{R}^n$ wird $\hat{\beta}(y)$ *Schätzwert* genannt. \square

§02.04 **Definition.** In einem linearen Modell $Y \odot \text{IM}_{X_{\mathbb{R}^p \times \{\Sigma\}}$ heißt jede (messbare) Wahl von $\hat{\beta}$, so dass

$$\hat{\beta} \in \arg \min_{b \in \mathbb{R}^p} \|\Sigma^{-1/2}(Y - Xb)\|^2 \quad (02.02)$$

verallgemeinerter Kleinste-Quadrate-Schätzer (vKQS oder GLSE für *generalized least squares estimator*) des unbekanntem Parametervektors β . Im gewöhnlichen Fall ($\Sigma = \sigma^2 E_n$) bezeichnen wir $\hat{\beta}$ als *gewöhnlichen Kleinste-Quadrate-Schätzer* (gKQS oder OLSE für *ordinary least squares estimator*). \square

§02.05 **Geometrische Interpretation.** Betrachten wir eine Realisierung y der Beobachtung Y als einen Punkt im n -dimensionalen Raum \mathbb{R}^n und variieren wir den Parameter β , so beschreibt $X\beta$ den k -dimensionalen Unterraum $\text{Bild}(X) = X\mathbb{R}^p$, d.h. eine k -dimensionale Hyperebene durch den Ursprung im \mathbb{R}^n . Der gewöhnliche Kleinste-Quadrate-Schätzwert $\hat{\beta}(y)$ gibt uns nun den Punkt $X\hat{\beta}(y)$ auf der Hyperebene, der der Beobachtung y am nächsten liegt. Da die euklidische Norm durch ein Skalarprodukt $\langle \cdot, \cdot \rangle$ induziert ist, bedeutet die Wahl der euklidischen Norm als Abstand im \mathbb{R}^n , geometrisch, dass wir y orthogonal bzgl. des Skalarproduktes $\langle \cdot, \cdot \rangle$ auf diese Hyperebene projizieren.



□

§02.06 **Bemerkung.** Formal ist der in (02.02) angegebene verallgemeinerte Kleinste-Quadrate-Schätzer nur in dem linearen Modell $Y \odot \text{IM}_{X\mathbb{R}^p \times \{\Sigma\}}$ mit bekannter Kovarianzmatrix $\Sigma \in \mathbb{R}_{>}^{(n,n)}$ ein Schätzer, da seine Definition die Kenntnis von Σ voraussetzt. Wir werden sehen (vgl. Lemma §02.07), dass ein gewöhnlicher Kleinste-Quadrate-Schätzer nicht von σ^2 abhängt und somit ein Schätzer im gewöhnlichen linearen Modell $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ ist. Im allgemeinen Fall $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_{>}^{(n,n)}}$ wird zum Beispiel ein zwei-stufiges Verfahren betrachtet, in dem zuerst Σ geschätzt wird, und dann im zweiten Schritt ein verallgemeinerter Kleinste-Quadrate-Schätzer bezüglich der geschätzten Kovarianzmatrix bestimmt wird. □

§02.07 **Lemma.** Setze $\tilde{X} := \Sigma^{-1/2}X$ sowie $\tilde{Y} := \Sigma^{-1/2}Y$. Bezeichne mit $\Pi_{\tilde{X}\mathbb{R}^p}$ die orthogonale Projektion von \mathbb{R}^n auf $\text{Bild}(\tilde{X}) = \tilde{X}\mathbb{R}^p$. Dann sind die folgenden Aussagen äquivalent: (i) $\hat{\beta}$ ist vKQS, d.h. $\hat{\beta}$ erfüllt (02.02), (ii) $\tilde{X}\hat{\beta} = \Pi_{\tilde{X}\mathbb{R}^p}\tilde{Y}$, (iii) $\tilde{X}^t\tilde{X}\hat{\beta} = \tilde{X}^t\tilde{Y}$ („Normalgleichungen“). Insbesondere existiert der vKQS. □

§02.08 **Beweis von Lemma §02.07.** In der Vorlesung. □

§02.09 **Korollar.** Sei X eine Designmatrix mit $\text{rg}[X] = p$, dann gilt $\Pi_{\tilde{X}\mathbb{R}^p} = \tilde{X}(\tilde{X}^t\tilde{X})^{-1}\tilde{X}^t$ und $\hat{\beta} = (\tilde{X}^t\tilde{X})^{-1}\tilde{X}^t\tilde{Y} = (X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}Y$ ist der eindeutige vKQS. Weiterhin ist im gewöhnlichen linearen Modell der gKQS $\hat{\beta} = (X^tX)^{-1}X^tY$ eindeutig und unabhängig von der Kenntnis von σ^2 .

§02.10 **Beweis** von **Korollar** §02.09. Die Aussage folgt direkt aus **Lemma** §02.07. \square

§02.11 **Bemerkung.** Die Matrix $\tilde{X}^+ := (\tilde{X}^t\tilde{X})^{-1}\tilde{X}^t$ heißt auch **Moore-Penrose-Inverse** von \tilde{X} und für die vKQS gilt $\hat{\beta} = \tilde{X}^+\tilde{Y}$. \square

§02.12 **Einfache lineare Regression (Beispiel §01.02 fortgesetzt).** Wir wählen eine alternative Parametrisierung $\beta_1 := a + b\bar{z}$ sowie $\beta_2 := b$ mit $\bar{z} = n^{-1}\sum_{i \in \llbracket n \rrbracket} z_i$. Dann gilt

$$Y_i = \beta_1 + \beta_2(z_i - \bar{z}) + \varepsilon_i, \quad i \in \llbracket n \rrbracket.$$

Setze weiterhin $x_i := (1, z_i - \bar{z})^t$, $i \in \llbracket n \rrbracket$ und $X = (x_i)_{i \in \llbracket n \rrbracket}^t$, so dass $\mathbb{E}(Y) = X\beta$ mit $\beta = (\beta_1, \beta_2)^t$. Wir bestimmen im Folgenden einen gKQS von β , dazu setze $\bar{Y} := n^{-1}\sum_{i \in \llbracket n \rrbracket} Y_i$, $S_{zY} := \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z})Y_i = \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z})(Y_i - \bar{Y})$ und $S_{zz} := \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z})^2$, dann gilt

$$\begin{aligned} X^tY &= \sum_{i \in \llbracket n \rrbracket} x_i Y_i = \begin{pmatrix} \sum_{i \in \llbracket n \rrbracket} Y_i \\ \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z})Y_i \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ S_{zY} \end{pmatrix} \\ X^tX &= \sum_{i \in \llbracket n \rrbracket} x_i x_i^t = \begin{pmatrix} n & \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z}) \\ \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z}) & \sum_{i \in \llbracket n \rrbracket} (z_i - \bar{z})^2 \end{pmatrix} = \begin{pmatrix} n & 0 \\ 0 & S_{zz} \end{pmatrix}. \end{aligned}$$

Somit hat X^tX den vollen Rang falls mindestens zwei $\{z_i\}$ verschieden sind. In dieser Situation ist nach **Korollar** §02.09 der gKQS eindeutig gegeben durch $\hat{\beta} = (X^tX)^{-1}X^tY = (\bar{Y}, S_{zz}^{-1}S_{zY})^t$ und somit sind $\hat{a} = \bar{Y} - \hat{b}\bar{z}$ und $\hat{b} = S_{zz}^{-1}S_{zY}$ die gKQS von a und b . \square

§02.13 **Varianzanalyse mit einem Faktor (Beispiel §01.06 (b) fortgesetzt).** Wir bestimmen im Folgenden die gKQS der unbekannt Parameter $(\mu_i)_{i \in \llbracket p \rrbracket}^t$. Bezeichnet $\bar{Y}_{j\bullet} := q^{-1}\sum_{k \in \llbracket q \rrbracket} Y_{jk}$, $j \in \llbracket 1, p \rrbracket$, dann gilt $X^tY = (q\bar{Y}_{i\bullet})_{i \in \llbracket p \rrbracket}^t$ und $X^tX = qE_p$. Offensichtlich hat X^tX den vollen Rang, so dass $\hat{\beta} = (X^tX)^{-1}X^tY = (\bar{Y}_{i\bullet})_{i \in \llbracket p \rrbracket}^t$ nach **Korollar** §02.09 der eindeutige gKQS von $\beta = (\mu_i)_{i \in \llbracket p \rrbracket}^t$ ist. \square

§02.14 **Erinnerung..** In einem linearen Modell $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_>^{(n,n)}} = (\text{IM}_{(X,\Sigma)})_{(\beta,\Sigma) \in \mathbb{R}^p \times \mathbb{R}_>^{(n,n)}}$ sei $\hat{\beta} := \hat{\beta}(Y)$ ein Schätzer für den Parameter β . Ist $\hat{\beta} \in \mathcal{L}_1(\mathbb{P}_{(\beta,\Sigma)})$, also $\mathbb{E}_{(\beta,\Sigma)}(|\hat{\beta}|) < \infty$, für jedes $\mathbb{P}_{(\beta,\Sigma)} \in \text{IM}_{X,\Sigma}$ und alle $(\beta, \Sigma) \in \mathbb{R}^p \times \mathbb{R}_>^{(n,n)}$, so wird $\text{bias}_{(\beta,\Sigma)}(\hat{\beta}) := \mathbb{E}_{(\beta,\Sigma)}(\hat{\beta} - \beta)$ **Verzerrung** oder **Bias** von $\hat{\beta}$ genannt. Der Schätzer $\hat{\beta}$ heißt **erwartungstreu** oder **unverfälscht** für den Parameter β , wenn $\text{bias}_{(\beta,\Sigma)}(\hat{\beta}) = 0$ für jedes $\mathbb{P}_{(\beta,\Sigma)} \in \text{IM}_{X,\Sigma}$ und alle $(\beta, \Sigma) \in \mathbb{R}^p \times \mathbb{R}_>^{(n,n)}$ gilt. \square

§02.15 **Satz von Gauß-Markov.** Besitzt die Designmatrix X den Rang $\text{rg}[X] = p$, so gelten im gewöhnlichen linearen Modell $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_>^+}$ die folgenden Aussagen:

(i) Der gKQS $\hat{\beta} = (X^tX)^{-1}X^tY$ ist ein erwartungstreuer Schätzer für β .

(ii) (**Satz von Gauß-Markov**) Unter allen Schätzern des abgeleiteten linearen Parameters $\gamma = \langle \beta, v \rangle$ für ein $v \in \mathbb{R}^p$, die (afin-)linear (in den Daten Y) und erwartungstreu sind, besitzt der lineare und erwartungstreu Schätzer $\hat{\gamma} = \langle \hat{\beta}, v \rangle$ eine minimale Varianz, nämlich $\text{Var}(\hat{\gamma}) = \sigma^2 \|X(X^tX)^{-1}v\|^2$.

(iii) Bezeichnet $R := Y - X\hat{\beta}$ den Residuenvektor, so ist die geeignet normalisierte Stichprobenvarianz $\hat{\sigma}^2 := \frac{1}{n-p}\|R\|^2 = \frac{1}{n-p}\|Y - X\hat{\beta}\|^2$ ein erwartungstreuer Schätzer für σ^2 .

§02.16 **Beweis** von Satz §02.15. In der Vorlesung. □

§02.17 **Bemerkung.**

- (i) Der Schätzer $\hat{\gamma}$ im Satz von Gauß-Markov wird bester linearer erwartungstreuer Schätzer (*BLUE* für best linear unbiased estimator) genannt. Verzichtet man auf die Linearität oder Erwartungstreue des Schätzers, so gibt es im Allgemeinen bessere Schätzer im Sinne des mittleren quadratischen Fehlers, zumindest für ausgewählte unbekannte Parameter β bzw. γ . Ein einfacher linearer aber nicht erwartungstreuer Schätzer ist $\tilde{\gamma} = 0$. Offensichtlich gilt für seinen MSE $\mathbb{E}(\tilde{\gamma} - \gamma)^2 = \gamma^2$, so dass für alle unbekannt Parameter in einer hinreichend kleine Umgebung um die Null, der MSE von $\tilde{\gamma}$ strikt kleiner als der MSE des BLUE $\hat{\gamma}$ ist.
- (ii) Aus Satz §02.15 (ii) folgt direkt, dass der gKQS $\hat{\beta} = (X^t X)^{-1} X^t Y$ unter allen Schätzern für β , die (afin-)linear (in den Daten Y) und erwartungstreu sind, einen minimalen mittleren quadratischen Fehler besitzt, nämlich $\mathbb{E}\|\hat{\beta} - \beta\|^2 = \sigma^2 \text{Spur}((X^t X)^{-1})$.
- (iii) Häufig sind wir nicht am MSE für eine Parameterschätzung im zu Grunde liegenden Modell interessiert, sondern an dem Vorhersagefehler $\|X\hat{\beta} - X\beta\|^2$. In der Situation einer gewöhnlichen linearen Regression entspricht dies der quadrierten Differenz der vorhergesagten und wahren Werte an den Designpunkten. Die Koordinaten des Vektors $\hat{Y} = X\hat{\beta}$ werden angepasste Werte (*fitted values*) genannt. Für den mittleren Vorhersagefehler (MPE für *mean prediction error*) lässt sich nun leicht nachprüfen, dass

$$\mathbb{E}\|X\hat{\beta} - X\beta\|^2 = \mathbb{E}\|\Pi_{X\mathbb{R}^p} Y - \Pi_{X\mathbb{R}^p} X\beta\|^2 = \mathbb{E}\|\Pi_{X\mathbb{R}^p} \varepsilon\|^2 = \sigma^2 p.$$

Insbesondere wächst der Vorhersagefehler linear in der Dimension p des Parameterraumes.

- (iv) Eine entsprechende Aussage des Satzes von Gauß-Markov gilt auch für den vKQS im linearen Modell $Y \odot \text{IM}_{X\mathbb{R}^p \times \{\Sigma\}}$ (nachrechnen!). □

§03 Das normale lineare Modell

§03.01 **Definition.** Ein *normales lineares Modell* bezeichnet ein lineares Modell in dem der zu erklärende zufällige Vektor eine multivariate Normalverteilung besitzt, so dass wir abkürzend schreiben $Y \odot N_{X\mathbb{R}^p \times \mathbb{R}_>^{(n,n)}} := (N_{(X\beta, \Sigma)})_{(\beta, \Sigma) \in \mathbb{R}^p \times \mathbb{R}_>^{(n,n)}}$. In einem *gewöhnlichen normalen linearen Modell* $Y \odot N_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ gilt weiterhin $\Sigma = \sigma^2 E_n$ für ein Fehlerniveau $\sigma \in \mathbb{R}_0^+$. Im gewöhnlichen Fall sind die Koordinaten des zentrierten Fehlervektors $\varepsilon := Y - X\beta$ unabhängig und identisch $N_{(0, \sigma^2)}$ -verteilt, d.h. $\varepsilon \sim N_{(0, \sigma^2)}^n$. □

§03.02 **Satz.** Besitzt die Designmatrix X den Rang $\text{rg}[X] = p$, so gelten im gewöhnlichen normalen linearen Modell $Y \odot N_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ die folgenden Aussagen:

- (i) Der gKQS ist normalverteilt:

$$\hat{\beta} \sim N_{(\beta, \sigma^2 (X^t X)^{-1})}.$$

- (ii) Die Stichprobenvarianz $\hat{\sigma}^2 = \frac{1}{n-p}\|Y - X\hat{\beta}\|^2$ ist nach geeigneter Normalisierung χ^2 -verteilt mit $n - p$ Freiheitsgraden:

$$(n - p) \hat{\sigma}^2 / \sigma^2 \sim \chi_{n-p}^2.$$

- (iii) Der gKQS $\hat{\beta}$ und die Stichprobenvarianz $\hat{\sigma}^2$ sind unabhängig.

- (iv) Der zentrierte und geeignet normalisierte gKQS $\widehat{\beta}$ hat eine t_k -Verteilung mit $n - p$ Freiheitsgraden: mit $\widehat{\sigma} := \sqrt{\widehat{\sigma}^2}$ für $v \in \mathbb{R}^p$

$$\frac{\langle \widehat{\beta} - \beta, v \rangle}{\widehat{\sigma} | \langle (X^t X)^{-1} v, v \rangle |^{1/2}} \sim t_{n-p}.$$

- (v) Der Vorhersagefehler $\|X(\beta - \widehat{\beta})\|^2$ ist nach geeigneter Normalisierung $F_{n,k}$ -verteilt mit p und $n - p$ Freiheitsgraden:

$$\frac{\|X(\widehat{\beta} - \beta)\|^2}{p\widehat{\sigma}^2} \sim F_{p,n-p}.$$

§03.03 **Beweis** von Satz §03.02. Übung. □

§03.04 **Erinnerung.** Im Folgenden beschreibe $Y \odot N_{X\mathbb{R}^p \times \mathbb{R}_0^+} := (N_{(X\beta, \sigma^2 E_n)})_{(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+}$ adäquat den linearen Zusammenhang. Sei $\gamma : \mathbb{R}^p \times \mathbb{R}_0^+ \rightarrow \Gamma$ ein identifizierbarer interessierender Parameter sowie $\{\mathcal{H}^0, \mathcal{H}^1\}$ eine Partition der Menge Γ , also $\mathcal{H}^0 \uplus \mathcal{H}^1 = \Gamma$. Wir sagen, die **Nullhypothese** $H_0 : \mathcal{H}^0$ ist erfüllt, wenn für ein $(\beta, \sigma) \in \gamma^{-1}(\mathcal{H}^0) \subseteq \mathbb{R}^p \times \mathbb{R}_0^+$, also $\gamma(\beta, \sigma) \in \mathcal{H}^0$, gilt $Y \sim N_{(X\beta, \sigma^2 E_n)}$, andernfalls ist die **Alternativhypothese** $H_1 : \mathcal{H}^1$ erfüllt. Die Entscheidung, ob die Nullhypothese $H_0 : \mathcal{H}^0$ oder die Alternativhypothese $H_1 : \mathcal{H}^1$ vorliegt, wird (**statistisches**) **Testproblem** genannt. Eine (messbare) Entscheidungsfunktion $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$, also Entscheidungen sind nur anhand einer Stichprobe $y \in \mathbb{R}^n$ erlaubt, heißt (**statistischer**) **Hypothesentest**, kurz **Test**. Wir vereinbaren, dass $\varphi^{-1}(\{1\}) \in \mathcal{B}^n$ das Ereignis aller Stichproben ist, die zu einer Entscheidung gegen die Nullhypothese H_0 , also für die Alternativhypothese H_1 führen, sowie dass $\varphi^{-1}(\{0\}) \in \mathcal{B}^n$ das Ereignis aller Stichproben ist, die zu einer Entscheidung für die Nullhypothese H_0 führen, Im Fall $\Gamma = (a, b) \subseteq \overline{\mathbb{R}}$ mit

- (i) $\mathcal{H}_{\gamma_o}^0 = (a, \gamma_o)$ und $\mathcal{H}_{\gamma_o}^1 = (\gamma_o, b)$ (bzw. $\mathcal{H}_{\gamma_o}^0 = [\gamma_o, b)$ und $\mathcal{H}_{\gamma_o}^1 = (a, \gamma_o)$) für ein $\gamma_o \in \Gamma$ wird das Testproblem **einseitig** genannt und wir schreiben es auch in der Form: Nullhypothese $H_0 : \gamma \leq \gamma_o$ gegen Alternativhypothese $H_1 : \gamma > \gamma_o$ (bzw. $H_0 : \gamma \geq \gamma_o$ gegen $H_1 : \gamma < \gamma_o$).
- (ii) $\mathcal{H}_{\gamma_o}^0 = \{\gamma_o\}$ und $\mathcal{H}_{\gamma_o}^1 = (a, \gamma_o) \cup (\gamma_o, b)$ für ein $\gamma_o \in \Gamma$ wird das Testproblem **zweiseitig** genannt und wir schreiben es auch in der Form: Nullhypothese $H_0 : \gamma = \gamma_o$ gegen Alternativhypothese $H_1 : \gamma \neq \gamma_o$.

Wir folgen der Konvention, dass $\mathcal{A} = \{\varphi = 1\} = \varphi^{-1}(\{1\})$ der Ablehnbereich eines Tests φ ist, also $\varphi = \mathbb{1}_{\mathcal{A}}$ ist und die **Nullhypothese abgelehnt** wird, wenn φ den Wert eins annimmt. Die Messbarkeit eines Tests garantiert dabei die Messbarkeit des assoziierten Ablehnbereiches. Weiterhin bezeichnet

Fehler 1. Art: die Wahrscheinlichkeit $N_{(X\beta, \sigma^2 E_n)}^\varphi(\{1\}) = N_{(X\beta, \sigma^2 E_n)}(\varphi = 1) = N_{(X\beta, \sigma^2 E_n)}(\mathcal{A})$ die Nullhypothese abzulehnen, sich also für die Alternativhypothese zu entscheiden, obwohl $(\beta, \sigma) \in \gamma^{-1}(\mathcal{H}^0)$ vorliegt, also die Nullhypothese adäquat ist;

Fehler 2. Art: die Wahrscheinlichkeit $N_{(X\beta, \sigma^2 E_n)}^\varphi(\{0\}) = N_{(X\beta, \sigma^2 E_n)}(\varphi = 0) = N_{(X\beta, \sigma^2 E_n)}(\mathcal{A}^c)$ die Nullhypothese nicht abzulehnen, sich also für Nullhypothese zu entscheiden, obwohl $(\beta, \sigma) \in \gamma^{-1}(\mathcal{H}^1)$ vorliegt, also die Alternativhypothese adäquat ist. □

Ein Test φ hält das (**Signifikanz**-) **Niveau** $\alpha \in [0, 1]$ ein, wenn für jedes $(\beta, \sigma) \in \gamma^{-1}(\mathcal{H}^0)$ der Fehler 1. Art $N_{(X\beta, \sigma^2 E_n)}(\varphi = 1) \leq \alpha$ erfüllt. In diesem Fall wird φ kurz **α -Test** genannt. □

§03.05 **Korollar.** Unter den Annahmen und den Notationen des Satz §03.02 kann für ein $r \in \mathbb{R}$ (a) die Hypothese $H_0 : \langle \beta, v \rangle \leq r$ gegen die Alternative $H_1 : \langle \beta, v \rangle > r$; (b) $H_0 : \langle \beta, v \rangle \geq r$ gegen

$H_1 : \langle \beta, v \rangle < r$ sowie (c) $H_0 : \langle \beta, v \rangle = r$ gegen $H_1 : \langle \beta, v \rangle \neq r$ mit Hilfe der Teststatistik $T := \frac{\langle \hat{\beta}, v \rangle - r}{\hat{\sigma} |((X^t X)^{-1} v, v)|^{1/2}}$ und den Entscheidungsregeln

- (i) lehne die Hypothese H_0 ab, falls $T \geq t_{n-p, 1-\alpha}$;
- (ii) lehne die Hypothese H_0 ab, falls $T \leq -t_{n-p, 1-\alpha}$;
- (iii) lehne die Hypothese H_0 ab, falls $|T| \geq t_{n-p, 1-\alpha/2}$;

unter Einhaltung des vorgegebenen Niveau $\alpha \in (0, 1)$ getestet werden.

§03.06 **Beweis** von **Korollar** §03.05. Die Aussage folgt direkt aus **Satz** §03.02. □

§03.07 **Beispiel** (Beispiel §01.06 (a) fortgesetzt). In einem *normalen Lokations-Skalen-Modell* $Y \odot N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ kann (a) die *Hypothese* $H_0 : \mu \leq \mu_o$ gegen die Alternative $H_1 : \mu > \mu_o$; (b) $H_0 : \mu \geq \mu_o$ gegen $H_1 : \mu < \mu_o$ sowie (c) $H_0 : \mu = \mu_o$ gegen $H_1 : \mu \neq \mu_o$ mit Hilfe der Entscheidungsregeln des (Student-)t-Tests für den Erwartungswert

- (a) lehne die Hypothese H_0 ab, falls $\bar{Y} - \mu_o \geq t_{n-1, 1-\alpha} n^{-1/2} \hat{\sigma}$;
- (b) lehne die Hypothese H_0 ab, falls $\bar{Y} - \mu_o \leq -t_{n-1, 1-\alpha} n^{-1/2} \hat{\sigma}$;
- (c) lehne die Hypothese H_0 ab, falls $|\bar{Y} - \mu_o| \geq t_{n-1, 1-\alpha/2} n^{-1/2} \hat{\sigma}$;

unter Einhaltung des vorgegebenen Niveau $\alpha \in (0, 1)$ getestet werden. □

§03.08 **Erinnerung** (Erinnerung §03.04 fortgesetzt). Eine Abbildung $B : \mathbb{R}^n \rightarrow 2^\Gamma$ heißt *Bereichsschätzfunktion* (BSF) für Γ , wenn $\{\gamma \in B\} := \{y \in \mathbb{R}^n \mid \gamma \in B(y)\} \in \mathcal{B}^n$ für alle Parameterwerte $\gamma \in \Gamma$ gilt. Für alle $\gamma \in \Gamma$ und $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$ wird $N_{(X\beta, \sigma^2 E_n)}(\gamma \in B)$ *Überdeckungswahrscheinlichkeit* von γ genannt. Für jedes $\gamma \in \Gamma$ sei $\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\}$ eine Partition der Menge der interessierenden Parameterwerte Γ , also $\mathcal{R}_\gamma \uplus \mathcal{F}_\gamma = \Gamma$. Für alle $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$ fassen wir $\mathcal{R}_{\gamma(\beta, \sigma)}$ und $\mathcal{F}_{\gamma(\beta, \sigma)}$ als Menge der „richtigen“ bzw. der „falschen“ abgeleiteten Parameterwerte auf. Im Fall $\Gamma = \mathbb{R}$ sind insbesondere von Interesse

- (i) $\mathcal{R}_\gamma = \{\gamma\}$ und $\mathcal{F}_\gamma = \{\gamma\}^c = (-\infty, \gamma) \cup (\gamma, \infty) = \mathbb{R} \setminus \{\gamma\} =: \mathbb{R}_\gamma$;
- (ii) $\mathcal{R}_\gamma = (-\infty, \gamma]$ und $\mathcal{F}_\gamma = (-\infty, \gamma]^c = (\gamma, \infty)$;
- (iii) $\mathcal{R}_\gamma = [\gamma, \infty)$ und $\mathcal{F}_\gamma = [\gamma, \infty)^c = (-\infty, \gamma)$.

Für ein $\alpha \in (0, 1)$ heißt eine Bereichsschätzfunktion B *Konfidenzbereich zum Niveau $1 - \alpha$* für $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$, wenn $N_{(X\beta, \sigma^2 E_n)}(\tilde{\gamma} \in B) \geq 1 - \alpha$ für alle $\tilde{\gamma} \in \mathcal{R}_{\gamma(\beta, \sigma)}$ und für alle $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$ gilt. In diesem Fall wird B kurz *(1- α)-Konfidenzbereich* (für $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$) genannt.

(i) Ist $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ eine Familie von Partitionen der Menge der interessierenden Parameterwerte Γ , so definiert $\mathcal{H}_\gamma^0 := \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\}$ und $\mathcal{H}_\gamma^1 := \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{F}_{\tilde{\gamma}}\}$ für $\gamma \in \Gamma$ eine *assoziierte Familie* $(\{\mathcal{H}_\gamma^0, \mathcal{H}_\gamma^1\})_{\gamma \in \Gamma}$ *von Partitionen* der Parametermenge Γ , dabei fassen wir \mathcal{H}_γ^0 als Nullhypothese und \mathcal{H}_γ^1 als Alternativhypothese eines Testproblems auffassen. Für $\gamma \in \Gamma = \mathbb{R}$ sind insbesondere assoziiert:

- $\mathcal{R}_\gamma = \{\gamma\}$ sowie $\mathcal{H}_\gamma^0 = \{\gamma\}$;
- $\mathcal{R}_\gamma = (-\infty, \gamma]$ sowie $\mathcal{H}_\gamma^0 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\} = \{\tilde{\gamma} \in \Gamma \mid \gamma \leq \tilde{\gamma}\} = [\gamma, \infty)$;
- $\mathcal{R}_\gamma = [\gamma, \infty)$ sowie $\mathcal{H}_\gamma^0 = (-\infty, \gamma]$.

(ii) Da definitionsgemäß $\mathcal{A}_\gamma := \{\gamma \notin B\} \in \mathcal{F}$ für jeden Parameterwert $\gamma \in \Gamma$ gilt, definiert eine Bereichsschätzfunktion B eine *assoziierte Familie von Tests* $(\varphi_\gamma = \mathbb{1}_{\mathcal{A}_\gamma})_{\gamma \in \Gamma}$ mit Ablehnbereich $\{\varphi_\gamma = 1\} = \mathcal{A}_\gamma$.

(iii) Für eine Familie $(\varphi_\gamma = \mathbb{1}_{\mathcal{A}_\gamma})_{\gamma \in \Gamma}$ von Tests mit Ablehnbereich $\mathcal{A}_\gamma = \{\varphi_\gamma = 1\} \in \mathcal{F}$ definiert $B : \mathbb{R}^n \rightarrow 2^\Gamma$ mit $B(y) := \{\gamma \in \Gamma \mid y \notin \mathcal{A}_\gamma\}$ eine *assoziierte Bereichsschätzfunktion*,

da in der Tat $\{\gamma \in B\} = \{y \in \mathbb{R}^n \mid \gamma \in B(y)\} = \{y \in \mathbb{R}^n \mid y \notin \mathcal{A}_\gamma\} = \mathcal{A}_\gamma^c \in \mathcal{F}$ für jeden abgeleiteten Parameterwert $\gamma \in \Gamma$ gilt.

Dann gilt für eine Familie $(\varphi_\gamma)_{\gamma \in \Gamma}$ von Tests, dass φ_γ genau dann ein α -Test der Nullhypothese $H_0 : \mathcal{H}_\gamma^0$ gegen die Alternativhypothese $H_1 : \mathcal{H}_\gamma^1$ für jedes $\gamma \in \Gamma$ ist, wenn die assoziierte Bereichsschätzfunktion B für $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ ein $(1-\alpha)$ -Konfidenzbereich ist. \square

§03.09 **Korollar.** Unter den Annahmen und den Notationen des Satz §03.02 gelten folgende Konfidenzaussagen für gegebenes $\alpha \in (0, 1)$:

(i) **Konfidenzbereich für β :** Bezeichnet $F_{p,n-p,1-\alpha}$ das $(1-\alpha)$ -Quantil einer $F_{p,n-p}$ -Verteilung mit p und $n-p$ Freiheitsgraden, so ist

$$C_\alpha = \{\beta \in \mathbb{R}^p : \|X(\hat{\beta} - \beta)\|^2 \leq p \hat{\sigma}^2 F_{p,n-p,1-\alpha}\}$$

ein Konfidenzellipsoid zum Niveau $1-\alpha$ für die Menge der richtigen Parameter $\mathcal{R}_\beta = \{\beta\}$ sowie der falschen Parameter $\mathcal{F}_\beta = \mathbb{R}_\beta^p$.

(ii) **Konfidenzbereich für $\langle \beta, v \rangle$:** Bezeichnet $t_\alpha := t_{n-p,1-\alpha/2} = -t_{n-p,\alpha/2}$ das $(1-\alpha/2)$ -Quantil einer t_k -Verteilung mit $n-p$ Freiheitsgraden, so ist

$$[\langle \hat{\beta}, v \rangle \pm t_\alpha \hat{\sigma} | \langle (X^t X)^{-1} v, v \rangle |^{1/2}]$$

ein Konfidenzintervall zum Niveau $1-\alpha$ für die Menge der richtigen Parameter $\mathcal{R}_\beta = \{\langle \beta, v \rangle\}$ sowie der falschen Parameter $\mathcal{F}_\beta = \mathbb{R}_{\langle \beta, v \rangle}$ mit $\beta \in \mathbb{R}^p$.

§03.10 **Beweis von Korollar §03.09.** Die Aussage folgt direkt aus Satz §03.02. \square

§03.11 **Beispiel (Beispiel §03.07 fortgesetzt).** In einem normalen Lokations-Skalen-Modell $Y \odot \mathbb{N}_{\mathbb{R} \times \mathbb{R}_0^+}^n$ ist

$$[\bar{Y} \pm t_{n-1,1-\alpha/2} n^{-1/2} \hat{\sigma}]$$

mit $\hat{\sigma}^2 = \frac{1}{n-1} \|Y - \bar{Y} \mathbf{1}_n\|^2$ ein Konfidenzintervall zum Niveau $1-\alpha$ für die Menge der richtigen Parameter $\mathcal{R}_\mu = \{\mu\}$ sowie der falschen Parameter $\mathcal{F}_\mu = \mathbb{R}_{\mu}$ mit $\mu \in \mathbb{R}$. Dies folgt direkt aus Korollar §03.09 (ii) mit $p=1$, $v=1$ und $\gamma=\mu$. \square

§04 Asymptotische Theorie und Residuenanalyse

Wir untersuchen nun die Verteilung des Kleinste-Quadrate-Schätzers im Grenzfall, in dem die Anzahl der Beobachtungen gegen unendlich geht. Dazu sei $(Y_n)_{n \in \mathbb{N}}$ eine Folge von zufälligen Zielgrößen und $(x_n)_{n \in \mathbb{N}}$ eine Folge von erklärenden Effekten. Wir nehmen an, dass für alle $n \geq n_0$ der Zusammenhang zwischen dem zufälligen (Spalten-)Vektor $Y_{(n)} := (Y_i)_{i \in \llbracket n \rrbracket}^t$ und der Designmatrix $X_{(n)} = (x_1 \cdots x_n)^t$ adäquat durch ein gewöhnliches lineares Modell beschrieben ist. Insbesondere fordern wir, dass $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_0^+$ existieren, so dass $Y_{(n)} \odot \text{IM}_{X_{(n)}\beta, \sigma^2 E_n}$ für alle $n \in \mathbb{N}$ adäquat ist. In dieser Situation gilt für $\varepsilon_{(n)} := Y_{(n)} - X_{(n)}\beta \odot \text{IM}_{0\mathbf{1}_n, \sigma^2 E_n}$ für alle $n \in \mathbb{N}$.

§04.01 **Erinnerung.** Sei $\gamma : \mathbb{R}^p \times \mathbb{R}_0^+ \rightarrow \Gamma \subseteq \mathbb{R}^q$ für jedes $n \in \mathbb{N}$ ein identifizierbarer Parameter und für jedes $n \in \mathbb{N}$ sei $\hat{\gamma}_n : \mathbb{R}^n \rightarrow \Gamma$ ein Schätzer für γ , also $\hat{\gamma}_n \in (\mathcal{B}^n)^q$. Die Folge von Schätzern $(\hat{\gamma}_n)_{n \in \mathbb{N}}$ heißt (*schwach*) *konsistent*, wenn $\|\hat{\gamma}_n - \gamma(\beta, \sigma)\| \xrightarrow{\mathbb{P}} 0$ für jedes $\mathbb{P}_n \in \text{IM}_{X_{(n)}\beta, \sigma^2 E_n}$, $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$, $n \in \mathbb{N}$, gilt. Zum Beispiel meint „ $\hat{\gamma}_n$ ist konsistent“ stets, dass die Folge von Schätzern $(\hat{\gamma}_n)_{n \in \mathbb{N}}$ konsistent ist. \square

§04.02 **Satz.** Für ein $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$ sei $Y_{(n)} \odot \text{IM}_{X_{(n)}\beta, \sigma^2 E_n}$ mit $\text{rg}[X_{(n)}] = p$ für $n \in \llbracket n_0, \infty \rrbracket$. Gelten die folgenden drei Bedingungen:

(B1) $\varepsilon_{(n)} := (\varepsilon_i)_{i \in \llbracket n \rrbracket} = Y_{(n)} - X_{(n)}\beta \stackrel{D}{\sim} \text{IM}_{0, \sigma^2}^n$ sind unabhängige und identisch verteilte (u.i.v.) Zufallsvariablen.

(B2) Für den kleinsten Eigenwert $\lambda_{(n)}$ der Matrix $X_{(n)}^t X_{(n)}$ gilt $\lim_{n \rightarrow \infty} \lambda_{(n)} = \infty$.

(B3) Für die Diagonalelemente der Projektionsmatrix $\Pi_{(n)} := X_{(n)}(X_{(n)}^t X_{(n)})^{-1} X_{(n)}^t$ gilt $\lim_{n \rightarrow \infty} \max_{j \in \llbracket n \rrbracket} [\Pi_{(n)}]_{jj} = 0$.

Dann ist der Kleinst-Quadrate-Schätzer $\widehat{\beta}_{(n)} := (X_{(n)}^t X_{(n)})^{-1} X_{(n)}^t Y_{(n)}$ konsistent für β und

$$\frac{1}{\sigma} (X_{(n)}^t X_{(n)})^{1/2} (\widehat{\beta}_{(n)} - \beta) \xrightarrow{D} N_{(0, E_p)}$$

(konvergiert in Verteilung gegen eine p -dimensionale Standardnormalverteilung) und weiterhin gilt für $v \in \mathbb{R}^p$

$$\frac{\langle \widehat{\beta}_{(n)} - \beta, v \rangle}{\sigma | \langle (X_{(n)}^t X_{(n)})^{-1} v, v \rangle |^{1/2}} \xrightarrow{D} N_{(0,1)}.$$

Gilt zusätzlich $\mathbb{E}(\varepsilon_1^4) < \infty$, so ist $\widehat{\sigma}^2 = \frac{1}{n-p} \|Y_{(n)} - X_{(n)} \widehat{\beta}_{(n)}\|^2$ ein konsistenter Schätzer für σ^2 .

§04.03 **Beweis** von Satz §04.02. In der Vorlesung. □

§04.04 **Bemerkung.** Die Bedingung Satz §04.02 (B2) besagt, dass man mit wachsendem n immer mehr Information bekommt. Weiterhin dominiert kein Vektor von Effekten x_j die anderen unter der Bedingung Satz §04.02 (B3). □

§04.05 **Erinnerung.** Seien $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ eine Familie von Partitionen der Menge der interessierenden Parameterwerte und B^n eine Bereichsschätzfunktion auf $(\mathbb{R}^n, \mathcal{B}^n)$ für jedes $n \in \mathbb{N}$. Die Folge von Bereichsfunktionen $(B^n)_{n \in \mathbb{N}}$ hält für $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ das **Niveau $1-\alpha$ asymptotisch** ein, wenn $\liminf_{n \rightarrow \infty} \mathbb{P}_{(n)}(\tilde{\gamma} \in B^n) \geq 1 - \alpha$ für alle $\mathbb{P}_{(n)} \in \text{IM}_{X_{(n)}, \beta, \sigma^2 E_n}$, $\tilde{\gamma} \in \mathcal{R}_{\gamma(\beta, \sigma)}$ und $(\beta, \sigma) \in \mathbb{R}^p \times \mathbb{R}_0^+$ gilt. □

§04.06 **Korollar.** Unter den Annahmen und den Notationen des Satz §04.02 gilt folgende asymptotische Konfidenzaussage für gegebenes $\alpha \in (0, 1)$. Bezeichnet $z_{1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil einer $N_{(0,1)}$ -Verteilung, so ist

$$[\langle \widehat{\beta}_{(n)}, v \rangle \pm z_{1-\alpha/2} \widehat{\sigma} | \langle (X_{(n)}^t X_{(n)})^{-1} v, v \rangle |^{1/2}]$$

ein Konfidenzintervall zum asymptotischen Niveau $1 - \alpha$ für $\langle \beta, v \rangle$.

§04.07 **Beweis** von Korollar §04.06. Die Aussage folgt direkt aus Satz §04.02. □

§04.08 **Beispiel** (Beispiel §01.06 (a) fortgesetzt). Für jedes $n \in \mathbb{N}$ sei $Y_{(n)} \stackrel{D}{\sim} \text{IM}_{\mathbb{R} \times \mathbb{R}_0^+}^n$ durch ein **Lokations-Skalen-Modell** mit u.i.v. Koordinaten adäquat beschrieben, dann ist

$$[\bar{Y}_{(n)} \pm z_{1-\alpha/2} n^{-1/2} \sqrt{\widehat{\sigma}_{(n)}^2}]$$

mit $\bar{Y}_{(n)} = \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} Y_i$ und $\widehat{\sigma}_{(n)}^2 = \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} (Y_i - \bar{Y}_{(n)})^2$ ein Konfidenzintervall zum asymptotischen Niveau $1 - \alpha$ für den unbekannt Parameter μ . Dies folgt direkt aus Korollar §04.06 mit $v = 1$. □

§04.09 **Erinnerung.** Sei $\{\mathcal{H}^0, \mathcal{H}^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ , und $\varphi^n : \mathbb{R}^n \rightarrow \{0, 1\}$ für jedes $n \in \mathbb{N}$ ein Test mit Gütefunktion $\beta_{\varphi^n}(\mathbb{P}_{(n)}) := \mathbb{E}_{\mathbb{P}_{(n)}}(\varphi^n)$. Für das Testproblem der Nullhypothese $H_0 : \mathcal{H}^0$ gegen die Alternativhypothese $H_1 : \mathcal{H}^1$ hält die Folge von Tests $(\varphi^n)_{n \in \mathbb{N}}$ **asymptotisch** das **Signifikanz-Niveau $\alpha \in (0, 1)$** ein, wenn $\limsup_{n \rightarrow \infty} \beta_{\varphi^n}(\mathbb{P}_{(n)}) \leq \alpha$ für alle $\mathbb{P}_{(n)} \in \text{IM}_{X_{(n)}, \beta, \sigma^2 E_n}$ und $(\beta, \sigma) \in \gamma^{-1}(\mathcal{H}^0)$ gilt.

- (i) Für $\gamma \in \Gamma$ sei \mathcal{R}_γ und \mathcal{F}_γ eine Partition in richtige und falsche interessierende Parameterwerte, dann bezeichnen $\mathcal{H}_\gamma^0 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\}$ und $\mathcal{H}_\gamma^1 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{F}_{\tilde{\gamma}}\}$ die *assozierte Null- bzw. Alternativhypothese* der interessierende Parameter.
- (ii) Für $n \in \mathbb{N}$ sei B^n eine Bereichsschätzfunktion, dann bezeichnet $(\varphi_\gamma^n)_{\gamma \in \Gamma}$ die *assozierte Familie von Tests* mit Ablehnbereich $\{\varphi_\gamma^n = 1\} = \{\gamma \notin B^n\}$.
- (iii) Für $n \in \mathbb{N}$ sei $(\varphi_\gamma^n)_{\gamma \in \Gamma}$ einer Familie von Tests, dann bezeichnet B^n die *assozierte Bereichsschätzfunktion* mit $B^n(x) := \{\gamma \in \Gamma \mid \varphi_\gamma^n(x) = 0\}$.

Dann gilt für eine Familie $((\varphi_\gamma^n)_{n \in \mathbb{N}})_{\gamma \in \Gamma}$ von Testfolgen, dass φ_γ^n ein *asymptotischer α -Test* der Nullhypothese $H_0 : \mathcal{H}_\gamma^0$ gegen die Alternativhypothese $H_1 : \mathcal{H}_\gamma^1$ für jedes $\gamma \in \Gamma$ ist, genau dann wenn die assoziierte Folge von Bereichsfunktionen $(B^n)_{n \in \mathbb{N}}$ für $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ ein *asymptotischer $1-\alpha$ -Konfidenzbereich* ist. \square

§04.10 **Korollar.** *Unter den Annahmen und den Notationen des Satz §04.02 kann für ein $r \in \mathbb{R}$ (a) die Hypothese $H_0 : \langle \beta, v \rangle \leq r$ gegen die Alternative $H_1 : \langle \beta, v \rangle > r$; (b) $H_0 : \langle \beta, v \rangle \geq r$ gegen $H_1 : \langle \beta, v \rangle < r$ sowie (c) $H_0 : \langle \beta, v \rangle = r$ gegen $H_1 : \langle \beta, v \rangle \neq r$ mit Hilfe der Teststatistik $T(n) := \frac{\langle \hat{\beta}, v \rangle - r}{\hat{\sigma} | \langle (X^t X)^{-1} v, v \rangle |^{1/2}}$ und den Entscheidungsregeln*

- (i) lehne die Hypothese H_0 ab, falls $T(n) \geq z_{1-\alpha}$;
- (ii) lehne die Hypothese H_0 ab, falls $T(n) \leq -z_{1-\alpha}$;
- (iii) lehne die Hypothese H_0 ab, falls $|T(n)| \geq z_{1-\alpha/2}$;

unter Einhaltung des vorgegebenen asymptotischen Niveau $\alpha \in (0, 1)$ getestet werden.

§04.11 **Beweis** von Korollar §04.10. Die Aussage folgt direkt aus Satz §04.02. \square

§04.12 **Beispiel** (Beispiel §04.08 fortgesetzt). Für jedes $n \in \mathbb{N}$ sei $Y(n) \odot \text{IM}_{\mathbb{R} \times \mathbb{R}_+^n}$ durch ein *Lokations-Skalen-Modell* mit u.i.v. Koordinaten adäquat beschrieben, dann kann (a) die *Hypothese $H_0 : \mu \leq \mu_o$* gegen die Alternative $H_1 : \mu > \mu_o$; (b) $H_0 : \mu \geq \mu_o$ gegen $H_1 : \mu < \mu_o$ sowie (c) $H_0 : \mu = \mu_o$ $H_1 : \mu \neq \mu_o$ mit Hilfe der Entscheidungsregeln

- (a) lehne die Hypothese H_0 ab, falls $\bar{Y}(n) - \mu_o \geq z_{1-\alpha} n^{-1/2} \sqrt{\hat{\sigma}_{(n)}^2}$;
- (b) lehne die Hypothese H_0 ab, falls $\bar{Y}(n) - \mu_o \leq -z_{1-\alpha} n^{-1/2} \sqrt{\hat{\sigma}_{(n)}^2}$;
- (c) lehne die Hypothese H_0 ab, falls $|\bar{Y}(n) - \mu_o| \geq z_{1-\alpha/2} n^{-1/2} \sqrt{\hat{\sigma}_{(n)}^2}$;

unter Einhaltung des vorgegebenen asymptotischen Niveau $\alpha \in (0, 1)$ getestet werden. \square

Residuenanalyse Wir nehmen im Folgenden an, dass der Zusammenhang zwischen der Zielgröße Y und der Designmatrix durch ein gewöhnliches lineares Modell $Y \odot \text{IM}_{(X \mathbb{R}^p \times \mathbb{R}_+^n)}$ adäquat dargestellt ist. Bezeichnen wir mit \bar{Y} das arithmetische Mittel der Beobachtung, so ist die totale Quadratsumme $\|Y - \bar{Y} \mathbf{1}_n\|^2 = \sum_{i \in \llbracket n \rrbracket} (Y_i - \bar{Y})^2$ (SST für *total sum of squares*) ein Maß der Variabilität der Realisierungen der Zielgrößen. Wir wollen nun untersuchen in wie weit diese Variabilität durch die Variabilität der angepassten Schätzwerte $\hat{Y} = X\hat{\beta}$ oder der Residuen $Y - \hat{Y}$ erklärt wird. Unter der Annahme $\mathbf{1}_n \in \text{Bild}(X)$ eine einfache Zerlegung der totalen Quadratsumme in eine Quadratsumme der Regression bzgl. der angepassten Werte (SSR für *regression sum of squares*) und eine Quadratsumme der Residuen (SSE für *error sum of squares*) ergibt

$$SST := \|Y - \bar{Y} \mathbf{1}_n\|^2 = \|\hat{Y} - \bar{Y} \mathbf{1}_n\|^2 + \|Y - \hat{Y}\|^2 =: SSR + SSE.$$

Offensichtlich, spricht ein im Verhältnis zum SSR kleiner Wert des SSE für eine gute Anpassung des linearen Modells. Betrachten wir den standardisierten Quotienten

$$F^* = \frac{\frac{1}{p}SSR}{\frac{1}{n-p}SSE},$$

so sprechen große Werte von F für eine gute Anpassung des linearen Modells. Nehmen wir zusätzlich an, dass die Beobachtung Y normalverteilt ist, so vergleichen wir die Anpassung in dem linearen Modell $Y \odot N_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ mit der in einem Lokations-Skalen-Modell $Y \odot N_{\mathbb{R} \times \mathbb{R}_0^+}^n$. Unter der Annahme, dass $\mathbb{1}_n \in \text{Bild}(X)$ gilt, hat F^* eine $F_{p, n-p}$ -Verteilung mit $(p, n-p)$ Freiheitsgraden. Alternativ können wir des Verhältnis zwischen der totalen Variabilität und der Variabilität der Schätzwerte betrachten:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

Der Wert R^2 wird *Bestimmtheitsmaß* genannt und entspricht im Fall $p = 1$ dem Quadrat des *empirischen Korrelationskoeffizienten*

$$\rho = \frac{\sum_{i \in \llbracket n \rrbracket} (Y_i - \bar{Y})(x_i - \bar{x})}{\sqrt{\sum_{i \in \llbracket n \rrbracket} (Y_i - \bar{Y})^2} \cdot \sqrt{\sum_{i \in \llbracket n \rrbracket} (x_i - \bar{x})^2}}.$$

Bezeichnet $\hat{\beta}_{-i}$ den gewöhnliche Kleinste-Quadrate-Schätzer ohne die i -te Koordinate der Beobachtung Y , dann gilt

$$\hat{\beta}_{-i} - \hat{\beta} = -\frac{\hat{Y}_i - Y_i}{1 - [X(X^t X)^{-1} X]_{ii}} (X^t X)^{-1} x_i.$$

Wir sehen also, dass der Einfluss der i -ten Beobachtung sowohl vom i -ten Residuum als auch vom Diagonalelement $[X(X^t X)^{-1} X]_{ii}$, seinem *Leverage-Score*, abhängt. Um einflussreiche Beobachtungen zu entdecken, plottet man daher oft die Residuen R_i gegen die $[X(X^t X)^{-1} X]_{ii}$. Basierend auf der Differenz der geschätzten Parameter ist die *Cook-Distanz* definiert durch

$$\frac{1}{p\hat{\sigma}^2} \|\hat{\beta}_{-i} - \hat{\beta}\|_{X^t X}^2 = \frac{1}{p\hat{\sigma}^2} \frac{(\hat{Y}_i - Y_i)^2}{1 - [X(X^t X)^{-1} X]_{ii}} \frac{[X(X^t X)^{-1} X]_{ii}}{1 - [X(X^t X)^{-1} X]_{ii}}.$$

Sie ist eine einfache Funktion von $[X(X^t X)^{-1} X]_{ii}$ sowie dem Quadrat des studentisierten Residuums $(\hat{Y}_i - Y_i) / \sqrt{\hat{\sigma}^2(1 - [X(X^t X)^{-1} X]_{ii})}$ welche Student- t_k -verteilt ist unter einer Normalverteilungsannahme. Sie wird häufig als diagnostisches Hilfsmittel verwendet. Diejenigen Beobachtungen, bei denen die Cook-Distanz deutlich größer ist als beim Rest, sollte besonders betrachtet, bzw. in der Analyse weggelassen werden. Analog erhält man als Änderung beim Hinzufügen einer Beobachtung Y_{n+1} zum Effekt x_{n+1} :

$$\frac{1}{1 + x_{n+1}^t (X^t X)^{-1} x_{n+1}} (X^t X)^{-1} x_{n+1} (Y_{n+1} - x_{n+1}^t \hat{\beta}).$$

Durch Hinzufügen einer einzigen Beobachtung kann somit der Kleinste-Quadrate-Schätzer beliebig verändert werden.

2 Entscheidungstheorie

§05 Formalisierung eines statistischen Problem

§05.01 **Erinnerung.** $\mathcal{W}(\mathcal{X})$ bezeichnet die Menge aller Wahrscheinlichkeitsmaße auf einem messbaren Raum $(\mathcal{X}, \mathcal{X})$. Für eine nicht-leere Indexmenge Θ wird eine Familie $\mathbb{P}_\Theta := (\mathbb{P}_\theta)_{\theta \in \Theta}$ von Wahrscheinlichkeitsmaßen auf \mathcal{X} formal durch die Abbildung $\Theta \rightarrow \mathcal{W}(\mathcal{X})$ mit $\theta \mapsto \mathbb{P}_\theta$ definiert. Für jedes $\theta \in \Theta$ bezeichnet im Folgenden \mathbb{E}_θ stets die Erwartung bzgl. \mathbb{P}_θ . Ist X eine Zufallsvariable mit Werten in $(\mathcal{X}, \mathcal{X})$ so schreiben wir abkürzend $X \odot \mathbb{P}_\theta$, falls $X \sim P_\theta$ für ein $\theta \in \Theta$ gilt. Eine Abbildung $\gamma : \Theta \rightarrow \Gamma$ heißt *abgeleiteter* oder *interessierender Parameter* und für jedes $\theta \in \Theta$ wird $\gamma(\theta)$ *abgeleiteter* oder *interessierender Parameterwert* genannt. Ein *abgeleiteter* oder *interessierender* Parameter $\gamma : \Theta \rightarrow \Gamma$ heißt *identifizierbar*, falls für beliebige $\theta_1, \theta_2 \in \Theta$ aus $\gamma(\theta_1) \neq \gamma(\theta_2)$ folgt $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. \square

§05.02 **Definition.** Wir bezeichnen das Tripel $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\Theta)$ als *statistisches Experiment* oder *statistisches Modell*. Die nicht-leere Indexmenge Θ wird *Parametermenge* genannt und \mathcal{X} heißt *Stichprobenraum*. Ein statistisches Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\Theta)$ heißt *adäquat* für eine Zufallsvariable X , falls $X \odot \mathbb{P}_\theta$ gilt. Ist $(\mathcal{S}, \mathcal{S})$ ein messbarer Raum, so wird jede $(\mathcal{S}, \mathcal{S})$ -wertige Zufallsvariable S auf $(\mathcal{X}, \mathcal{X})$, also \mathcal{X} - \mathcal{S} -messbare Funktion $S : \mathcal{X} \rightarrow \mathcal{S}$, *Beobachtung* oder *Statistik* genannt. Wir bezeichnen mit $\mathbb{P}_\Theta^S := (\mathbb{P}_\theta^S)_{\theta \in \Theta}$ die durch S auf $(\mathcal{S}, \mathcal{S})$ induzierte Familie von Wahrscheinlichkeitsmaßen und mit $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\Theta^S)$ das induzierte statistische Modell. Eine Statistik $\hat{\gamma}$ mit Werten in dem messbaren Raum (Γ, \mathcal{G}) heißt *Schätzer* oder *Schätzfunktion* für den abgeleiteten identifizierbaren Parameter γ . Sei $\{\mathcal{H}^0, \mathcal{H}^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ , also $\mathcal{H}^0 \sqcup \mathcal{H}^1 = \Gamma$. Für $X \odot \mathbb{P}_\theta$ sagen wir, die *Nullhypothese* $H_0 : \mathcal{H}^0$ ist erfüllt, wenn das statistische Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\gamma^{-1}(\mathcal{H}^0)})$ für X adäquat ist, das heißt, für ein $\theta \in \Theta$ mit $\gamma(\theta) \in \mathcal{H}^0$ gilt $X \sim \mathbb{P}_\theta$, andernfalls ist die *Alternativhypothese* $H_1 : \mathcal{H}^1$ erfüllt. Eine $(\{0, 1\}, 2^{\{0,1\}})$ -wertige Zufallsvariable (Statistik) φ wird (nicht randomisierter) *Test* für das Testproblem von $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$ genannt. Nimmt φ den Wert eins an, so wird die Hypothese H_0 *abgelehnt*, und andernfalls wird die Hypothese H_0 *nicht abgelehnt*. Eine $([0, 1], \mathcal{B}_{[0,1]})$ -wertige Statistik φ wird *randomisierter Test* genannt. Dabei wird $\varphi(x)$ als bedingte Wahrscheinlichkeit interpretiert, die Nullhypothese $H_0 : \mathcal{H}^0$ abzulehnen, wenn eine Realisierung $X = x$ beobachtet wird. \square

§05.03 **Definition.** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\Theta)$ ein statistisches Experiment. Eine *Entscheidungsregel* (Entscheidungsfunktion) ist eine \mathcal{X} - \mathcal{E} -messbare Abbildung $\delta : \mathcal{X} \rightarrow \mathcal{E}$ mit Werten in einem messbarem Raum $(\mathcal{E}, \mathcal{E})$, der *Entscheidungsraum* genannt wird. Wir bezeichnen mit Δ eine vorgegebene Menge von Entscheidungsfunktionen. Jede Funktion $\nu : \Theta \times \mathcal{E} \rightarrow \mathbb{R}^+$, die messbar im zweiten Argument ist, heißt *Verlustfunktion*. Das *Risiko* (der mittlere Verlust) einer Entscheidungsregel δ bei Vorliegen des Parameters $\theta \in \Theta$ (\mathbb{P}_θ ist die zu Grunde liegende Wahrscheinlichkeitsverteilung und \mathbb{E}_θ die Erwartung bezüglich \mathbb{P}_θ) ist

$$\mathcal{R}_\nu(\theta, \delta) := \mathbb{E}_\theta(\nu(\theta, \delta)) := \int_{\mathcal{X}} \nu(\theta, \delta(x)) \mathbb{P}_\theta(dx).$$

$(\mathcal{E}, \mathcal{E}, \nu)$ wird *statistisches Entscheidungsproblem* genannt. \square

§05.04 **Beispiel.**

- (a) Betrachte ein *gewöhnliches lineares Modell* $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$. Insbesondere, für $Y \odot \text{IM}_{X\beta, \sigma^2 E_n}$ ist die Verteilung \mathbb{P}^ε des zentrierten und standardisierten Fehlers $\varepsilon := \sigma^{-1}(Y - X\beta)$ ein Element von $\text{IM}_{0\mathbb{I}_n, E_n}$. Offensichtlich legen die unbekannt Parameter β und σ^2 sowie die unbekannte Verteilung \mathbb{P}^ε von ε die Verteilung \mathbb{P}^Y von Y fest. Wählen wir nun $\Theta := \mathbb{R}^p \times \mathbb{R}_0^+ \times \text{IM}_{0\mathbb{I}_n, E_n}$ als Parameterraum, so existiert zu jedem $\mathbb{P}^Y \in \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ mindestens ein Parameterwert $\theta = (\beta, \sigma^2, \mathbb{P}^\varepsilon) \in \Theta$ mit $\mathbb{P}^Y = \mathbb{P}^{X\beta + \sigma\varepsilon} = \mathbb{P}_\theta^Y$. Damit ist die Familie \mathbb{P}_θ^Y eine mögliche Parametrisierung von $\text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$. Versieht man den Stichprobenraum $\mathcal{X} = \mathbb{R}^n$ mit seiner Borel- σ -Algebra $\mathcal{X} = \mathcal{B}^n$ so bilden die Verteilungen \mathbb{P}_θ^Y eine Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum und es liegt zusammenfassend das statistische Experiment $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_{\mathbb{R}^p \times \mathbb{R}_0^+ \times \text{IM}_{0\mathbb{I}_n, E_n}})$ vor.

Um den (gewöhnlichen) Kleinste-Quadrate-Schätzer $\hat{\beta}$ als Entscheidungsregel zu interpretieren, sowie seine Güte zu messen, wird der Entscheidungsraum $\mathcal{E} = \mathbb{R}^p$ und beispielsweise die quadratische Verlustfunktion $\nu(\theta, e) = \nu((\beta, \sigma^2, \mathbb{P}^\varepsilon), e) = \|\beta - e\|^2$ betrachtet. Für diese spezielle Wahl der Verlustfunktion sind die Parameter σ^2 und \mathbb{P}^ε irrelevant. Da aber die Verteilung $\mathbb{P}_\theta^Y = \mathbb{P}^{X\beta + \sigma\varepsilon}$ von σ^2 und \mathbb{P}^ε abhängt, werden diese auch *Störparameter* genannt.

Beachte, dass bei der Modellierung $Y \odot \text{IM}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ die Verteilung von ε nicht eindeutig festgelegt wird. Dies gilt offensichtlich in einem gewöhnlichen normalen linearen Modell $Y \odot \text{N}_{X\mathbb{R}^p \times \mathbb{R}_0^+}$ nicht, da in diesem Fall $\text{IM}_{0\mathbb{I}_n, E_n}$ zu der ein-elementigen Menge $\{\text{N}_{(0,1)}^n\}$ reduziert wird. Mit anderen Worten, die multivariate Normalverteilung der Beobachtung Y ist gerade eindeutig durch das erste und zweite Moment festgelegt.

- (b) Für einen Test auf Wirksamkeit eines neuen Medikaments werden 100 Versuchspersonen mit diesem behandelt. Unter der (stark vereinfachenden) Annahme, dass alle Personen identisch und unabhängig auf das Medikament reagieren, wird für jede Person der Erfolg oder Misserfolg der Behandlung notiert, so dass die Anzahl X der erfolgreichen Behandlungen eine Binomial-verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $\pi \in (0, 1)$ ist. Zusammenfassend nehmen wir $X \odot (\text{Bin}_{(100, \pi)})_{\pi \in (0, 1)}$ an. Wählen wir den Stichprobenraum $\mathcal{X} = \llbracket 0, 100 \rrbracket$ versehen mit der Potenzmenge $\mathcal{X} = 2^{\llbracket 0, 100 \rrbracket}$ als σ -Algebra, so liegt das statistische Experiment $(\llbracket 0, 100 \rrbracket, 2^{\llbracket 0, 100 \rrbracket}, (\text{Bin}_{(100, \pi)})_{\pi \in (0, 1)})$ vor. In Abhängigkeit von der Anzahl X der erfolgreichen Behandlungen soll entschieden werden, ob die Erfolgsquote höher ist als diejenige einer klassischen Behandlung mit bekannter Erfolgswahrscheinlichkeit $\pi_o \in (0, 1)$. Die Nullhypothese für den unbekannt Parameter π ist somit $H_0 : \pi \leq \pi_o$. Als Entscheidungsraum dient $\mathcal{E} = \{0, 1\}$ (H_0 nicht ablehnen bzw. ablehnen) versehen mit der σ -Algebra $2^{\{0, 1\}}$, und wir wählen den Verlust $\nu(\pi, e) = \nu_0 e \mathbb{1}_{\{\pi \leq \pi_o\}} + \nu_1 (1 - e) \mathbb{1}_{\{\pi > \pi_o\}}$ mit Konstanten $\nu_0, \nu_1 \in \mathbb{R}^+$. Dies führt auf das Risiko einer Entscheidungsregel (eines Tests) $\delta : 2^{\llbracket 0, 100 \rrbracket} \rightarrow 2^{\{0, 1\}}$

$$\mathcal{R}_\nu(\pi, \delta) = \nu_0 \text{Bin}_{(100, \pi)}(\delta > \pi_o) \mathbb{1}_{\{\pi \leq \pi_o\}} + \nu_1 \text{Bin}_{(100, \pi)}(\delta \leq \pi_o) \mathbb{1}_{\{\pi > \pi_o\}},$$

so dass die Irrtumswahrscheinlichkeit erster Art $\text{Bin}_{(100, \pi)}(\delta > \pi_o)$ mit ν_0 und die zweiter Art $\text{Bin}_{(100, \pi)}(\delta \leq \pi_o)$ mit ν_1 gewichtet wird. \square

§05.05 **Definition.** Seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment und $(\mathcal{E}, \mathcal{E}, \nu)$ ein Entscheidungsproblem. Eine Entscheidungsregel $\delta_o \in \Delta$ heißt *(gleichmäßig) besser* als eine Entscheidungsregel $\delta \in \Delta$, falls $\mathcal{R}_\nu(\theta, \delta_o) \leq \mathcal{R}_\nu(\theta, \delta)$ für alle $\theta \in \Theta$ gilt und falls ein $\theta_o \in \Theta$ mit $\mathcal{R}_\nu(\theta_o, \delta_o) < \mathcal{R}_\nu(\theta_o, \delta)$ existiert. Eine Entscheidungsregel heißt *zulässig in Δ* , wenn es keine (gleichmäßig) bessere Entscheidungsregel in Δ gibt. \square

§05.06 **Bemerkung.** Häufig schränkt die betrachtete Klasse Δ die möglichen Entscheidungsregeln ein. So ist der gKQS im gewöhnlichen linearen Modell nach dem Satz §02.15 von Gauß-Markov

zulässig unter quadratischem Verlust in der Klasse der erwartungstreuen und linearen Schätzern. \square

§05.07 **Beispiel** (*Beispiel §01.06 (a) fortgesetzt*). Wir vergleichen in einem *normalen Lokations-Modell* $Y \odot (\mathbb{N}_{(\mu,1)}^n)_{\mu \in \mathbb{R}}$ die Schätzfunktionen $\hat{\mu}_1 := \bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i$, $\hat{\mu}_2 := \bar{Y} + 0.5$ sowie $\hat{\mu}_3 := 6$ unter Verwendung eines quadratischen Verlustes $\nu(\mu, \delta) = (\mu - \delta)^2$. Da $\mathcal{R}_\nu(\mu, \hat{\mu}_1) = 1/n$, $\mathcal{R}_\nu(\mu, \hat{\mu}_2) = 1/4 + 1/n$ gilt, ist $\hat{\mu}_1$ besser als $\hat{\mu}_2$, allerdings ist weder $\hat{\mu}_1$ besser als $\hat{\mu}_3$ noch umgekehrt. Insbesondere ist $\hat{\mu}_3$ zulässig (in der Klasse aller Schätzer!), da $\mathcal{R}_\nu(6, \hat{\mu}_3) = 0$ gilt und jeder andere Schätzer $\tilde{\mu}$ mit $\mathcal{R}_\nu(6, \tilde{\mu}) \leq \mathcal{R}_\nu(6, \hat{\mu}_3) = 0$ mit $\hat{\mu}_3$ Lebesgue-fast überall übereinstimmt. Später werden wir zeigen dass auch $\hat{\mu}_1$ zulässig ist. \square

§05.08 **Definition**. Zu einem vorgegebenen Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ in einem statistischen Modell $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ heißt eine Entscheidungsregel δ *unverzerrt*, falls

$$\forall \theta, \tilde{\theta} \in \Theta : \mathbb{E}_\theta(\nu(\tilde{\theta}, \delta)) \geq \mathbb{E}_\theta(\nu(\theta, \delta)) = \mathcal{R}_\nu(\theta, \delta). \quad \square$$

§05.09 **Lemma**. Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein statistisches Experiment, $\gamma : \Theta \rightarrow \mathcal{E} \subseteq \mathbb{R}$ ein identifizierbarer Parameter und $(\mathcal{E}, \mathcal{E}, \nu)$ ein statistisches Entscheidungsproblem mit quadratischem Verlust $\nu(\theta, e) := (\gamma(\theta) - e)^2$. Eine Entscheidungsregel $\hat{\gamma} : \mathcal{X} \rightarrow \mathcal{E}$ ist dann ein Schätzer für den abgeleiteten Parameter γ . Gilt für jedes $\theta \in \Theta$ weiterhin $\hat{\gamma} \in \mathcal{L}_1(\mathbb{P})$ und $\mathbb{E}_\theta(\hat{\gamma}) \in \gamma(\Theta) := \{\gamma(\theta_o), \theta_o \in \Theta\}$, dann ist die Entscheidungsregel $\hat{\gamma}$ genau dann unverzerrt, wenn sie *erwartungstreu* ist, also wenn $\mathbb{E}_\theta(\hat{\gamma}) = \gamma(\theta)$ für alle $\theta \in \Theta$ gilt.

§05.10 **Beweis von Lemma §05.09**. In der Vorlesung. \square

§05.11 **Lemma**. Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein statistisches Experiment, $\gamma : \Theta \rightarrow \Gamma = \mathcal{H}^0 \uplus \mathcal{H}^1$ ein identifizierbarer Parameter, und $([0, 1], \mathcal{B}_{[0,1]}, \nu)$ das statistische Entscheidungsproblem mit Verlustfunktion $\nu(\theta, e) = \nu_0 e \mathbb{1}_{\mathcal{H}^0}(\gamma(\theta)) + \nu_1 (1 - e) \mathbb{1}_{\mathcal{H}^1}(\gamma(\theta))$ für $\nu_0, \nu_1 \in \mathbb{R}^+$. Eine Entscheidungsregel $\varphi : \mathcal{X} \rightarrow [0, 1]$ (ein randomisierter Test) für das Testproblem $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$ ist genau dann unverzerrt, wenn sie zum Niveau $\alpha = \nu_1 / (\nu_0 + \nu_1)$ *unverfälscht* ist, also wenn gilt

$$\forall \theta \in \gamma^{-1}(\mathcal{H}^0) : \mathbb{E}_\theta(\varphi) \leq \alpha, \quad \text{und} \quad \forall \theta \in \gamma^{-1}(\mathcal{H}^1) : \mathbb{E}_\theta(\varphi) \geq \alpha.$$

§05.12 **Beweis von Lemma §05.11**. Übungsaufgabe. \square

§05.13 **Erinnerung**. Eine Abbildung $\kappa : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}^+$ heißt *Markovkern* von $(\mathcal{X}, \mathcal{X})$ nach $(\mathcal{S}, \mathcal{S})$ (vgl. *Erinnerung C02.05*), falls sie die zwei Bedingungen: **(Mk1)** $\kappa(x, \bullet) \in \mathcal{W}(\mathcal{S})$ für alle $x \in \mathcal{X}$ und **(Mk2)** $\kappa(\bullet, B) \in \mathcal{X}^+$ für alle $B \in \mathcal{S}$ erfüllt. Zu jedem $\mathbb{P} \in \mathcal{W}(\mathcal{X})$ existiert dann (vgl. *Schreibweise C02.06*) ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\mathbb{P} \odot \kappa \in \mathcal{W}(\mathcal{X} \otimes \mathcal{S})$ mit $\mathbb{P} \odot \kappa(A \times B) = \int_A \kappa(x, B) \mathbb{P}(dx)$ sowie $\kappa \mathbb{P} \in \mathcal{W}(\mathcal{S})$ mit $\kappa \mathbb{P}(B) = \int_{\mathcal{X}} \kappa(x, B) \mathbb{P}(dx)$ für alle $B \in \mathcal{S}$, $A \in \mathcal{X}$. \square

§05.14 **Definition**. Zu einem vorgegebenen Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ in einem statistischen Modell $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ mit $X \odot \mathbb{P}$ heißt ein Markovkern D von $(\mathcal{X}, \mathcal{X})$ nach $(\mathcal{E}, \mathcal{E})$ *Entscheidungskern* oder *randomisierte Entscheidungsregel* mit der Interpretation, dass bei Vorliegen der Beobachtung $X = x$ gemäß $D(x, \bullet)$ eine Entscheidung zufällig ausgewählt wird. Das zugehörige Risiko ist dann definiert durch

$$\mathcal{R}_\nu(\theta, D) := \mathbb{E}_\theta \left(\int_{\mathcal{E}} \nu(\theta, e) D(\bullet, de) \right) = \int_{\mathcal{X}} \int_{\mathcal{E}} \nu(\theta, e) D(x, de) \mathbb{P}_\theta(dx) = \int_{\mathcal{E}} \nu(\theta, \bullet) dD \mathbb{P}_\theta. \quad \square$$

§05.15 **Beispiel**. Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein statistisches Experiment, $X \odot \mathbb{P}$ und $\gamma : \Theta \rightarrow \Gamma$ ein identifizierbarer Parameter und $\Gamma = \mathcal{H}^0 \uplus \mathcal{H}^1$.

- (a) Betrachte $\mathcal{E} = \Gamma$ versehen mit einer σ -Algebra \mathcal{G} . Ein Entscheidungskern D von $(\mathcal{X}, \mathcal{X})$ nach (Γ, \mathcal{G}) ist dann ein „randomisierter“ Schätzer, d.h. bei Vorliegen der Beobachtung $X = x$ ist $D(x, \bullet) \in \mathcal{W}(\mathcal{T})$ eine Verteilung auf dem Parameterraum Γ . Existiert insbesondere eine (messbare) Abbildung $\hat{\gamma} : \mathcal{X} \rightarrow \Gamma$, so dass $D(x, \bullet) = \delta_{\hat{\gamma}(x)}$ für jedes $x \in \mathcal{X}$ ein Punktmaß in $\hat{\gamma}(x) \in \Gamma$ ist, dann ist $\hat{\gamma}$ eine Entscheidungsregel („nicht randomisierter“ Schätzer) und

$$\mathcal{R}_\nu(\theta, D) = \int_{\mathcal{X}} \int_{\mathcal{E}} \nu(\theta, e) D(x, de) \mathbb{P}_\theta(dx) = \int_{\mathcal{X}} \nu(\theta, \hat{\gamma}(x)) \mathbb{P}_\theta(dx) = \mathcal{R}_\nu(\theta, \hat{\gamma}).$$

- (b) Betrachte das Testproblem von $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$ und die statistische Entscheidungsprobleme $([0, 1], \mathcal{B}_{[0,1]}, \nu)$ sowie $(\{0, 1\}, 2^{\{0,1\}}, \nu)$ mit gemeinsamer Verlustfunktion $\nu(\theta, e) := \nu_0 e \mathbb{1}_{\mathcal{H}^0}(\gamma(\theta)) + \nu_1 (1 - e) \mathbb{1}_{\mathcal{H}^1}(\gamma(\theta))$. Jede Entscheidungsregel $\varphi : \mathcal{X} \rightarrow [0, 1]$ (randomisierter Test) zum Entscheidungsproblem $([0, 1], \mathcal{B}_{[0,1]}, \nu)$ definiert mit $D(x, \{1\}) := \varphi(x)$ sowie $D(x, \{0\}) := 1 - \varphi(x)$ eine randomisierte Entscheidungsregel D zum Entscheidungsproblem $(\{0, 1\}, 2^{\{0,1\}}, \nu)$. Auf der anderen Seite definiert jede randomisierte Entscheidungsregel D zum Entscheidungsproblem $(\{0, 1\}, 2^{\{0,1\}}, \nu)$ eine Entscheidungsregel $\varphi(x) := D(x, \{1\})$ zum Entscheidungsproblem $([0, 1], \mathcal{B}_{[0,1]}, \nu)$. Offensichtlich, gilt dann

$$\begin{aligned} \mathcal{R}_\nu(\theta, D) &= \mathbb{E}_\theta \left(\sum_{e \in \{0,1\}} \nu(\theta, e) D(\bullet, \{e\}) \right) = \mathbb{E}_\theta (\nu(\theta, 0)(1 - \varphi) + \nu(\theta, 1)\varphi) \\ &= \mathbb{E}_\theta (\nu_1 (1 - \varphi) \mathbb{1}_{\mathcal{H}^1}(\gamma(\theta)) + \nu_0 \varphi \mathbb{1}_{\mathcal{H}^0}(\gamma(\theta))) = \mathbb{E}_\theta (\nu(\theta, \varphi)) = \mathcal{R}_\nu(\theta, \varphi). \end{aligned}$$

Dies bedeutet also, dass $\varphi(x)$ die Wahrscheinlichkeit angibt, mit der bei Vorliegen der Beobachtung $X = x$ die Hypothese $H_0 : \mathcal{H}^0$ abgelehnt wird. \square

§05.16 **Bemerkung.** Es sei $\mathcal{E} \subseteq \mathbb{R}^d$ konvex sowie $\nu(\theta, e)$ eine im zweiten Argument konvexe Verlustfunktion. Dann gibt es zu jeder randomisierten Entscheidungsregel eine deterministische Entscheidungsregel, deren Risiko nicht größer ist. \square

§06 Minimax- und Bayes-Ansatz

§06.01 **Definition.** Für ein Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ zu einem statistischen Modell $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ heißt eine Entscheidungsregel δ_o Δ -*minimax*, falls

$$\mathcal{R}_\nu^* := \sup_{\theta \in \Theta} \mathcal{R}_\nu(\theta, \delta_o) = \inf_{\delta \in \Delta} \sup_{\theta \in \Theta} \mathcal{R}_\nu(\theta, \delta)$$

gilt, weiterhin wird \mathcal{R}_ν^* Δ -*Minimaxrisiko* genannt. Wir bezeichnen δ_o als *minimax*, falls die Menge Δ alle möglichen Entscheidungsregeln (für die das Risiko definiert ist) enthält. \square

§06.02 **Erinnerung.** Betrachten wir ein statistisches Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$, so gilt definitionsgemäß $\mathbb{P}_\theta \in \mathcal{W}(\mathcal{X})$ für jedes $\theta \in \Theta$. Nehmen wir zusätzlich an, dass (Θ, \mathcal{T}) ein messbarer Raum ist, und dass für jedes $F \in \mathcal{X}$ die Abbildung $\mathbb{P}_\theta(F) : \Theta \rightarrow \mathbb{R}^+$ mit $\theta \mapsto \mathbb{P}_\theta(F)$ (Borel-) messbar ist, also kurz $\mathbb{P}_\theta(F) \in \mathcal{T}^+$, so sind die beiden Bedingungen (Mk1) und (Mk2) eines Markovkerns von (Θ, \mathcal{T}) nach $(\mathcal{X}, \mathcal{X})$ erfüllt, den wir mit \mathbb{P} bezeichnen. \square

§06.03 **Definition.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment, (Θ, \mathcal{T}) ein messbarer Raum und \mathbb{P} mit $(\theta, F) \mapsto \mathbb{P}_\theta(F)$ ein Markovkerns von (Θ, \mathcal{T}) nach $(\mathcal{X}, \mathcal{X})$. Die Verlustfunktion ν des Entscheidungsproblems $(\mathcal{E}, \mathcal{E}, \nu)$ sei zusätzlich $\mathcal{T} \otimes \mathcal{E}$ - \mathcal{B}^+ -messbar. Sei ϑ eine (Θ, \mathcal{T}) -wertige Zufallsvariable, so dass die Parameter $\theta \in \Theta$ als Realisierung von ϑ aufgefasst werden können.

Die Wahrscheinlichkeitsverteilung \mathbb{P}^ϑ von ϑ auf dem messbaren Raum (Θ, \mathcal{T}) wird *a-priori Verteilung* des Parameters θ genannt und wir bezeichnen mit \mathbb{E}^ϑ die Erwartung bezüglich \mathbb{P}^ϑ . Das mit \mathbb{P}^ϑ assoziierte Bayesrisiko einer Entscheidungsregel δ ist (vgl. **Schreibweise** C02.06)

$$\begin{aligned} \mathcal{R}_\nu^\vartheta(\delta) &:= \mathbb{E}^\vartheta [\mathcal{R}_\nu(\bullet, \delta)] = \int_\Theta \mathcal{R}_\nu(\theta, \delta) \mathbb{P}^\vartheta(d\theta) = \int_\Theta \int_{\mathcal{X}} \nu(\theta, \delta(x)) \mathbb{P}_\theta^\vartheta(dx) \mathbb{P}^\vartheta(d\theta) \\ &= \int_{\Theta \times \mathcal{X}} \nu(\bullet, \delta) d[\mathbb{P}^\vartheta \odot \mathbb{P}]. \end{aligned}$$

Eine Entscheidungsregel δ_0 heißt *Δ -Bayesregel* oder *Δ -Bayes-optimal* (bezüglich \mathbb{P}^ϑ) falls

$$\mathcal{R}_\nu^\vartheta(\delta_0) = \inf_{\delta \in \Delta} \mathcal{R}_\nu^\vartheta(\delta)$$

gilt. Erstreckt sich das Infimum über alle möglichen Entscheidungsregeln δ so heißt δ_0 kurz *Bayesregel* oder *Bayes-optimal*. □

§06.04 **Bemerkung.** Während eine Minimaxregel den maximal zu erwartenden Verlust minimiert, kann das Bayesrisiko als ein (mittels \mathbb{P}^ϑ) gewichtetes Mittel des zu erwartenden Verlustes angesehen werden. Alternativ wird \mathbb{P}^ϑ als die subjektive Einschätzung der Verteilung der zu Grunde liegenden Parameter interpretiert. Daher wird das Bayesrisiko auch als insgesamt zu erwartender Verlust verstanden. □

§06.05 **Erinnerung.** Sei $\mathbb{P}^{(S,X)}$ die Verteilung von (S, X) auf $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$. Ein Markovkern $\mathbb{P}^{X|S}$ von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{X}, \mathcal{X})$ heißt *reguläre Festlegung der bedingte Wahrscheinlichkeitsverteilung von X gegeben S* (vgl. **Definition** C02.28 (a)), falls für alle $A \in \mathcal{X}$ und alle $B \in \mathcal{S}$ gilt

$$[\mathbb{P}^S \odot \mathbb{P}^{X|S}](B \times A) = \int_B \mathbb{P}^{X|S=s}(A) \mathbb{P}^S(ds) = \mathbb{P}^{(S,X)}(B \times A).$$

Nach **Eigenschaft** C02.09 (ii) existiert ein solcher Markovkern, wenn $\mathcal{S} \times \mathcal{X} = \mathbb{R}^k$ und $\mathcal{S} \otimes \mathcal{X} = \mathcal{B}^k$ für ein $k \in \mathbb{N}$, und nach **Eigenschaft** C02.02 ist dieser bis auf Gleichheit $\mathbb{P}^{(S,X)}$ -f.ü. eindeutig. Damit legt eine Wahl des Markovkerns $\mathbb{P}^{X|S}$ und der Randverteilung \mathbb{P}^S die gemeinsame Verteilung $\mathbb{P}^{(S,X)}$ fest. □

§06.06 **Definition.** Seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment, (Θ, \mathcal{T}) ein messbarer Raum, (ϑ, X) eine $(\Theta \times \mathcal{X}, \mathcal{T} \otimes \mathcal{X})$ -wertige Zufallsvariable, \mathbb{P}^ϑ eine a-priori Verteilung von ϑ auf (Θ, \mathcal{T}) , \mathbb{P} mit $(\theta, F) \mapsto \mathbb{P}_\theta(F)$ ein Markovkern von (Θ, \mathcal{T}) nach $(\mathcal{X}, \mathcal{X})$. Die gemeinsame Verteilung $\mathbb{P}^{(\vartheta, X)} = \mathbb{P}^\vartheta \odot \mathbb{P}$ von (ϑ, X) ist durch die reguläre bedingte Verteilung $\mathbb{P}^{X|\vartheta} := \mathbb{P}$ von X bei gegebenem ϑ sowie \mathbb{P}^ϑ festgelegt. Eine reguläre Festlegung der bedingten Verteilung $\mathbb{P}^{\vartheta|X}$ von ϑ bei gegebenem X , falls sie existiert, heißt *a-posteriori* Verteilung des zufälligen Parameters ϑ . Entsprechend wird eine reguläre Festlegung der bedingten Erwartung $\mathbb{E}^{\vartheta|X}$ von ϑ gegeben X , falls sie existiert, *a-posteriori* Erwartung des zufälligen Parameters ϑ genannt. Besitzt für ein $x \in \mathcal{X}$ die a-posteriori Verteilung $\mathbb{P}^{\vartheta|X=x}$ ein endliches erstes absolutes Moment, so wird $\mathbb{E}^{\vartheta|X=x}(\vartheta) := \mathbb{E}^{\vartheta|X=x}(\text{id}_\Theta) = \int_\Theta \theta \mathbb{P}^{\vartheta|X=x}(d\theta)$ *a-posteriori* Erwartungswert des zufälligen Parameters ϑ bei gegebenem $X = x$ genannt. Existiert ein endliches erstes absolutes Moment für alle $x \in \mathcal{X}$ so heißt die (messbare) Abbildung $\mathbb{E}^{\vartheta|X}(\vartheta)$ mit $x \mapsto \mathbb{E}^{\vartheta|X=x}(\vartheta)$ *a-posteriori* Erwartungswert des zufälligen Parameters ϑ bei gegebenem X . □

§06.07 **Bemerkung.** Die gemeinsame Verteilung $\mathbb{P}^{(\vartheta, X)}$ des zufälligen Vektors (ϑ, X) erfüllt $\mathbb{P}^{(\vartheta, X)} = \mathbb{P}^\vartheta \odot \mathbb{P}$. Wir bezeichnen mit \mathbb{P}^X die Randverteilung von X bzgl. $\mathbb{P}^{(\vartheta, X)}$, also $\mathbb{P}^X = \mathbb{P} \mathbb{P}^\vartheta =$

$\int_{\Theta} \mathbb{P}_{\theta} \mathbb{P}^{\vartheta} (d\theta)$, und mit \mathbb{E}^X die assoziierte Erwartung. Insbesondere gilt

$$\begin{aligned} \mathcal{R}_{\nu}^{\vartheta}(\delta) &= \mathbb{E}^{(\vartheta, X)} (\nu(\bullet, \delta)) = \mathbb{E}^{\vartheta} (\mathbb{E}^{X|\vartheta} (\nu(\bullet, \delta))) = \int_{\mathcal{X} \times \Theta} \nu(\bullet, \delta) d[\mathbb{P}^{\vartheta} \odot \mathbb{P}^{X|\vartheta}] \\ &= \mathbb{E}^X (\mathbb{E}^{\vartheta|X} (\nu(\bullet, \delta))) = \int_{\Theta \times \mathcal{X}} \nu(\bullet, \delta) d[\mathbb{P}^X \odot \mathbb{P}^{\vartheta|X}]. \quad \square \end{aligned}$$

§06.08 **Schreibweise.** Seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\theta})$ ein statistisches Experiment, \mathbb{P}_{θ} eine bezüglich eines σ -endlichen Maes $\mu \in \mathcal{M}_{\sigma}(\mathcal{X})$ dominierte Verteilungsfamilie ($\mathbb{P}_{\theta} \ll \mu$ fur alle $\theta \in \Theta$) mit entsprechenden μ -Dichten $(f_{\theta})_{\theta \in \Theta}$ und (Θ, \mathcal{T}) ein messbarer Raum. Fordern wir zusatzlich, dass $f_{\bullet} : \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$ mit $(\theta, x) \mapsto f_{\theta}(x)$ eine $(\mathcal{T} \otimes \mathcal{X})$ - \mathcal{B}^+ -messbare Funktion ist, dann ist mit Fubini \mathbb{P} mit $(\theta, F) \mapsto \mathbb{P}_{\theta}(F) = \mu(\mathbb{1}_F f_{\theta}) = \int_F f_{\theta} d\mu$ ein Markovkern von (Θ, \mathcal{T}) nach $(\mathcal{X}, \mathcal{X})$. In dieser Situation bezeichnen wir f_{\bullet} als $(\mu$ -) *Dichte des Markovkernes* \mathbb{P} bezuglich des dominierenden Maes μ . Betrachte eine $(\Theta \times \mathcal{X}, \mathcal{T} \otimes \mathcal{X})$ -wertigen Zufallsvariable (ϑ, X) . Die a-priori Verteilung \mathbb{P}^{ϑ} besitze eine ν -Dichte f^{ϑ} bezuglich eines dominierenden Maes $\nu \in \mathcal{M}_{\sigma}(\mathcal{T})$. Die gemeinsame Verteilung $\mathbb{P}^{(\vartheta, X)} = \mathbb{P}^{\vartheta} \odot \mathbb{P}^{X|\vartheta}$, die durch die regulare bedingte Verteilung $\mathbb{P}^{X|\vartheta} := \mathbb{P}$ und \mathbb{P}^{ϑ} festgelegt sei, ist dann dominiert durch das Produktma $\nu \otimes \mu$, besitzt die $\nu \otimes \mu$ -Dichte $f^{\vartheta} f_{\bullet} : (\theta, x) \mapsto f^{\vartheta}(\theta) f_{\theta}(x)$ und $f^X := \tilde{f}^X \mathbb{1}_{\{\tilde{f}^X \in \mathbb{R}^+\}} \in \mathcal{X}^+$ mit $x \mapsto \tilde{f}^X(x) = \nu(f^{\vartheta} f_{\bullet}(x)) = \int_{\Theta} f_{\bullet}(x) f^{\vartheta}(\theta) \nu(d\theta)$, ist eine μ -Dichte der Randverteilung $\mathbb{P}^X = \mathbb{P}_{\bullet} \mathbb{P}^{\vartheta}$ von X . Die Abbildung

$$f^{\vartheta|X} : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^+ \text{ mit}$$

$$(x, \theta) \mapsto f^{\vartheta|X=x}(\theta) = \frac{f_{\theta}(x) f^{\vartheta}(\theta)}{f^X(x)} \mathbb{1}_{\{f^X(x) \in \mathbb{R}^+\}} + f^{\vartheta}(\theta) \mathbb{1}_{\{f^X(x)=0\}} \quad (06.01)$$

ist dann eine $(\mathcal{X} \otimes \mathcal{T})$ - \mathcal{B}^+ -messbare Funktion und eine ν -Dichte des Markovkernes $\mathbb{P}^{\vartheta|X}$ von $(\mathcal{X}, \mathcal{X})$ nach (Θ, \mathcal{T}) mit $(x, B) \mapsto \mathbb{P}^{\vartheta|X=x}(B) := \nu(\mathbb{1}_B f^{\vartheta|X=x})$ (vgl. **Eigenschaft C03.14 (i)**). \square

§06.09 **Satz.** Seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\theta})$ ein statistisches Experiment, (Θ, \mathcal{T}) ein messbarer Raum und \mathbb{P} ein Markovkern von (Θ, \mathcal{T}) nach $(\mathcal{X}, \mathcal{X})$ mit Dichte f_{\bullet} bezuglich eines dominierenden Maes μ . Betrachte eine $(\Theta \times \mathcal{X}, \mathcal{T} \otimes \mathcal{X})$ -wertigen Zufallsvariable (ϑ, X) . Die a-priori Verteilung \mathbb{P}^{ϑ} besitze eine Dichte f^{ϑ} bezuglich eines dominierenden Maes ν und $\mathbb{P}^{(\vartheta, X)} := \mathbb{P}^{\vartheta} \odot \mathbb{P}$ sei die gemeinsame Verteilung von (ϑ, X) . Der Markovkern $\mathbb{P}^{\vartheta|X}$ von $(\mathcal{X}, \mathcal{X})$ nach (Θ, \mathcal{T}) mit ν -Dichte $f^{\vartheta|X}$ gegeben in (06.01) ist dann eine regulare Festlegung der bedingten Verteilung von ϑ gegeben X , also eine a-posteriori Verteilung.

§06.10 **Beweis** von **Satz** §06.09. *Ubungsaufgabe* \square

§06.11 **Beispiel.** Wir bezeichnen als *Bayestestproblem* (oder Bayes-Klassifikationsproblem) *mit einfachen Hypothesen* ein Entscheidungsproblem $(\{0, 1\}, 2^{\{0,1\}}, \nu)$ mit 0-1-Verlust $\nu(\theta, e) = |\theta - e|$ zu einem statistischen Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\theta})$ mit Parameterraum $\Theta = \{0, 1\}$. Betrachte eine $(\{0, 1\} \times \mathcal{X}, 2^{\{0,1\}} \otimes \mathcal{X})$ -wertige Zufallsvariable (ϑ, X) . Betrachte eine a-priori Verteilung \mathbb{P}^{ϑ} auf $(\{0, 1\}, 2^{\{0,1\}})$ mit Zahldichte p^{ϑ} bezuglich des Zahlmaes $\zeta_{\{0,1\}}$ (vgl. **Beispiel B01.03 (a)**). Die Familie von Wahrscheinlichkeitsmaen $\mathbb{P}_{\theta} = (\mathbb{P}_{\theta})_{\theta \in \{0,1\}}$ ist dominiert bezuglich eines Maes μ (z.B. $\mu = \mathbb{P}_0 + \mathbb{P}_1$) und f_0 und f_1 bezeichne entsprechende μ -Dichten. Dann ist f_{\bullet} eine μ -Dichte des Markovkernes \mathbb{P} und $\mathbb{P}^{(\vartheta, X)} := \mathbb{P}^{\vartheta} \odot \mathbb{P}$ bezeichne die gemeinsame Verteilung von (ϑ, X) . Dann ist $f^X = p^{\vartheta}(0) f_0 + p^{\vartheta}(1) f_1$ eine μ -Dichte der Randverteilung $\mathbb{P}^X = \mathbb{P}_{\bullet} \mathbb{P}^{\vartheta}$ von X bezuglich $\mathbb{P}^{(\vartheta, X)}$. Nach **Satz** §06.09 ist

$$p^{\vartheta|X} : \mathcal{X} \times \{0, 1\} \rightarrow \mathbb{R}^+ \text{ mit}$$

$$(x, \theta) \mapsto p^{\vartheta|X=x}(\theta) = \frac{p^{\vartheta}(\theta) f_{\theta}(x)}{p^{\vartheta}(0) f_0(x) + p^{\vartheta}(1) f_1(x)} \mathbb{1}_{\{f^X(x) \in \mathbb{R}^+\}} + p^{\vartheta}(\theta) \mathbb{1}_{\{f^X(x)=0\}} \quad (06.02)$$

eine $\zeta_{\{0,1\}}$ -Dichte der a-posteriori Verteilung von ϑ gegeben X . □

§06.12 **Satz.** *Es gelten die Annahmen und Notationen von Definition §06.06. Betrachten wir das statistische Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$, so ist jede Entscheidungsregel δ_o mit*

$$\delta_o(x) \in \arg \min_{\delta \in \Delta} \mathbb{E}^{\vartheta|X=x}(\nu(\bullet, \delta(x))) \quad \text{für } \mathbb{P}^X\text{-f.a. } x \in \mathcal{X},$$

Δ -Bayes-optimal bezüglich \mathbb{P}^ϑ .

§06.13 **Beweis** von Satz §06.12. In der Vorlesung. □

§06.14 **Korollar.** *Sei $\Theta \subseteq \mathbb{R}$. Unter den Annahmen von Satz §06.12 gelten die folgenden Aussagen:*

- (i) *Für die quadratische Verlustfunktion $\nu(\theta, e) := (e - \theta)^2$ ist eine Festlegung des bedingten Erwartungswertes $\hat{\theta} := \mathbb{E}^{\vartheta|X}(\vartheta)$ ein Bayes-optimaler Schätzer von θ (Bayes-optimale Entscheidungsregel) bezüglich der a-priori Verteilung $\mathbb{P}^\vartheta \in \mathcal{W}_2(\mathcal{T})$.*
- (ii) *Eine Entscheidungsregel $\hat{\theta}_{med}$, die $\mathbb{P}^{\vartheta|X=x}(\vartheta \leq \hat{\theta}_{med}(x)) \geq 1/2$ und $\mathbb{P}^{\vartheta|X=x}(\vartheta \geq \hat{\theta}_{med}(x)) \geq 1/2$ für \mathbb{P}^X -f.a. $x \in \mathcal{X}$ erfüllt, heißt a-posteriori Median bezüglich der a-priori Verteilung \mathbb{P}^ϑ . Für den Absolutbetrag $\nu(\theta, e) := |e - \theta|$ ist jeder a-posteriori Median bezüglich der a-priori Verteilung \mathbb{P}^ϑ ein Bayes-optimaler Schätzer von θ (Bayes-optimale Entscheidungsregel).*

§06.15 **Beweis** von Korollar §06.14. Übungsaufgabe. □

§06.16 **Beispiel** (§§06.11 fortgesetzt). Nach Satz §06.12 ist ein Bayestest (Bayesklassifizierer) für alle $x \in \{f^X = p^\vartheta(0)f_0 + p^\vartheta(1)f_1 > 0\}$ eine Minimalstelle der Abbildung

$$\{0, 1\} \ni e \mapsto \mathbb{E}^{\vartheta|X=x}[\nu(\bullet, e)] = \frac{p^\vartheta(0)f_0(x)}{f^X(x)}e + \frac{p^\vartheta(1)f_1(x)}{f^X(x)}(1 - e).$$

Eine Lösung des Minimierungsproblems und somit ein Bayestest ist gegeben durch

$$\varphi : \mathcal{X} \rightarrow \{0, 1\} \text{ mit } x \mapsto \varphi(x) := \begin{cases} 0, & p^\vartheta(0)f_0(x) > p^\vartheta(1)f_1(x); \\ 1, & p^\vartheta(0)f_0(x) < p^\vartheta(1)f_1(x); \\ \text{beliebig,} & p^\vartheta(0)f_0(x) = p^\vartheta(1)f_1(x). \end{cases}$$

Damit entscheiden wir uns für dasjenige $\varphi(x) \in \{0, 1\}$, dessen a-posteriori Wahrscheinlichkeit am größten ist (MAPE für maximum a posteriori estimator). Insbesondere sei für später auf die Neymann-Pearson-Struktur des Bayestests φ in Abhängigkeit von $f_1(x)/f_0(x)$ hingewiesen. □

§06.17 **Satz.** *Es seien die Annahmen und Notationen der Definition §06.06 erfüllt. Betrachten wir das statistische Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$, so gelten die folgenden Aussagen*

- (i) *Für jede Entscheidungsregel δ gilt*

$$\sup_{\theta \in \Theta} \mathcal{R}_\nu(\theta, \delta) = \sup_{\mathbb{P}^\vartheta \in \mathcal{W}(\mathcal{T})} \mathcal{R}_\nu^\vartheta(\delta),$$

wobei sich das zweite Supremum über alle a-priori Verteilungen \mathbb{P}^ϑ , also Wahrscheinlichkeitsmaße über (Θ, \mathcal{T}) , erstreckt. Insbesondere ist das Bayes-Risiko einer Δ -Bayesregel stets kleiner oder gleich dem Δ -Minimax-Risiko.

- (ii) *Für eine Δ -Minimaxregel δ_o gilt*

$$\sup_{\mathbb{P}^\vartheta \in \mathcal{W}(\mathcal{T})} \mathcal{R}_\nu^\vartheta(\delta_o) = \inf_{\delta \in \Delta} \sup_{\mathbb{P}^\vartheta \in \mathcal{W}(\mathcal{T})} \mathcal{R}_\nu^\vartheta(\delta).$$

§06.18 **Beweis** von **Satz** §06.17. In der Vorlesung. □

§06.19 **Bemerkung**. Der letzte Satz wird insbesondere dazu verwendet, untere Schranken für das Minimax-Risiko durch das Bayes-Risiko einer Bayesregel herzuleiten. □

§06.20 **Satz**. Es seien die Annahmen und Notationen der **Definition** §06.06 erfüllt. Im statistischen Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ gelten für jede Entscheidungsregel $\delta_\circ (\in \Delta)$ die folgenden Aussagen:

- (a) Ist δ_\circ minimax-optimal und eindeutig (in Δ) in dem Sinne, dass jede andere Minimax-Regel die gleiche Risikofunktion besitzt, so ist δ_\circ zulässig in Δ .
- (b) Ist δ_\circ zulässig mit konstanter Risikofunktion, so ist δ_\circ minimax-optimal.
- (c) Ist δ_\circ eine Bayesregel (bzgl. \mathbb{P}^θ) und eindeutig (in Δ) in dem Sinne, dass jede andere Bayesregel (bzgl. \mathbb{P}^θ) die gleiche Risikofunktion besitzt, so ist δ_\circ zulässig (in Δ).
- (d) Die Parametermenge Θ bilde einen metrischen Raum versehen mit der Borel- σ -Algebra \mathcal{B}_Θ . Ist δ_\circ eine Bayesregel (bzgl. \mathbb{P}^θ) (in Δ), so ist δ_\circ zulässig (in Δ), falls (i) $\mathcal{R}_\nu^\theta(\delta_\circ) \in \mathbb{R}^+$; (ii) für jede nicht leere offene Menge U in Θ gilt $\mathbb{P}^\theta(U) \in \mathbb{R}_0^+$; (iii) für jede Entscheidungsregel $\delta (\in \Delta)$ mit $\mathcal{R}_\nu^\theta(\delta) \leq \mathcal{R}_\nu^\theta(\delta_\circ)$ ist die Abbildung $\theta \mapsto \mathcal{R}_\nu(\theta, \delta)$ stetig.

§06.21 **Beweis** von **Satz** §06.20. Übungsaufgabe. □

§06.22 **Satz**. Für ein $d \in \mathbb{N}$ sei $X \odot \mathbb{N}_{\mathbb{R}^d \times \{E_d\}}^n = (\mathbb{N}_{(\mu, E_d)}^n)_{\mu \in \mathbb{R}^d}$ (vgl. **Beispiel** §01.06 (a)). Bezüglich der quadratischen Verlustfunktion $\nu(\mu, e) = \|\mu - e\|^2$ ist das arithmetische Mittel $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ein minimax-optimaler Schätzer für μ .

§06.23 **Beweis** von **Satz** §06.22. Übungsaufgabe. □

§06.24 **Satz**. Sei $X \odot \mathbb{N}_{\mathbb{R} \times \{1\}}^n = (\mathbb{N}_{(\mu, 1)}^n)_{\mu \in \mathbb{R}}$. Bezüglich der quadratischen Verlustfunktion $\nu(\mu, e) = (\mu - e)^2$ ist das arithmetische Mittel $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ein zulässiger Schätzer für μ .

§06.25 **Beweis** von **Satz** §06.24. In der Vorlesung. □

§06.26 **Bemerkung**. Liegt eine andere Verteilung mit Erwartungswert μ und Varianz eins als die Normalverteilung vor, so ist \bar{X} weder zulässig noch minimax (sofern $n \geq 3$ gilt), vergleiche Lehmann and Casella [1998], Seite 153. Unter der Normalverteilungsannahme ist \bar{X} für $d = 2$ weiterhin zulässig, allerdings gilt dies für $d = 3$ nicht mehr, wie es in Abschnitt §07 das Stein-Phänomen zeigt. □

§06.27 **Definition**. Es gelten die Annahmen und Notationen von **Definition** §06.06. Für ein Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ heißt eine Verteilung $\mathbb{P}^{\theta_\circ}$ auf (Θ, \mathcal{T}) mit

$$\inf_{\delta \in \Delta} \mathcal{R}_\nu^{\theta_\circ}(\delta) = \sup_{\mathbb{P}^\theta \in \mathcal{W}(\mathcal{T})} \inf_{\delta \in \Delta} \mathcal{R}_\nu^\theta(\delta)$$

ungünstigste a-priori Verteilung bzgl. Δ . □

§06.28 **Satz**. Für das Entscheidungsproblem $(\mathcal{E}, \mathcal{E}, \nu)$ sei $\mathbb{P}^{\theta_\circ}$ eine a-priori Verteilung mit zugehöriger Δ -Bayesregel δ_\circ . Dann sind die Eigenschaften

- (i) $\mathcal{R}_\nu^{\theta_\circ}(\delta_\circ) = \sup_{\theta \in \Theta} \mathcal{R}_\nu(\theta, \delta_\circ)$ und
- (ii) die **Sattelpunkteigenschaft**: $\forall \mathbb{P}^\theta \in \mathcal{W}(\mathcal{T}) : \forall \delta \in \Delta : \mathcal{R}_\nu^\theta(\delta_\circ) \leq \mathcal{R}_\nu^{\theta_\circ}(\delta_\circ) \leq \mathcal{R}_\nu^{\theta_\circ}(\delta)$ äquivalent. Aus jeder dieser Eigenschaften folgt, dass δ_\circ minimax-optimal in Δ und $\mathbb{P}^{\theta_\circ}$ ungünstigste a-priori Verteilung bzgl. Δ ist.

§06.29 **Beweis** von **Satz** §06.28. In der Vorlesung. □

§06.30 **Beispiel.** Für ein $n \in \mathbb{N}$ sei $X \odot (\text{Bin}_{(n,\pi)})_{\pi \in (0,1)}$. Wir bestimmen einen minimax-optimalen Schätzer für π in der Klasse $\Delta := (0, 1)^{\llbracket n \rrbracket} = \{\delta : \llbracket n \rrbracket \rightarrow (0, 1)\}$ bezüglich der quadratischen Verlustfunktion $\nu(\pi, e) = (e - \pi)^2$ unter Verwendung von **Satz §06.28**. Dazu betrachten wir auf $((0, 1), \mathcal{B}_{(0,1)})$ die Beta-Verteilung Beta mit Parametern $a, b \in \mathbb{R}_0^+$ als a-priori Verteilung, diese ist definiert durch die Lebesgue-Dichte $f_{\text{Beta}}(x) = \frac{1}{\text{Beta}(a,b)} x^{a-1} (1-x)^{b-1}$, $x \in (0, 1)$, mit Normierungskonstante $\text{Beta}(a, b)$ (Betafunktion). Wir bestimmen zunächst einen zugehörigen Bayesschätzer $\hat{\pi}_{a,b}$ für π . Bezeichne mit $\pi_{a,b}$ einen zufälligen Parameter mit Werten in $(0, 1)$ und a-priori Verteilung Beta. Eine Festlegung der a-posteriori Verteilung $\mathbb{P}^{\pi_{a,b}|X=x}$ ist wieder eine Beta-Verteilung $\text{Beta}[a+x, b+n-x]$ und der zugehörige Bayesschätzer ist $\hat{\pi}_{a,b} := \mathbb{E}^{\pi_{a,b}|X}(\pi_{a,b}) = \frac{a+X}{a+b+n}$. Für sein Risiko gilt $\mathcal{R}_\nu(\pi, \hat{\pi}_{a,b}) = \mathbb{E}_\pi(\hat{\pi}_{a,b} - \pi)^2 = \frac{(a-a\pi-b\pi)^2 + n\pi(1-\pi)}{(a+b+n)^2}$. Im Fall $a^* = b^* = \sqrt{n}/2$ erhält man $\hat{\pi}_{a^*,b^*} := \mathbb{E}^{\pi_{a^*,b^*}|X}(\pi_{a^*,b^*}) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} = \frac{X}{n} - \frac{X-n/2}{n(\sqrt{n}+1)}$ mit zugehörigem Risiko $\mathcal{R}_\nu(\pi, \hat{\pi}_{a^*,b^*}) = (2\sqrt{n} + 2)^{-2}$ welches unabhängig von π ist, woraus die Sattelpunkteigenschaft folgt:

$$\forall \mathbb{P}_\pi^\theta \in \mathcal{W}(\mathcal{B}_{(0,1)}) : \forall \hat{\pi} \in (0, 1)^{\llbracket 1, n \rrbracket} : \mathcal{R}_\nu^\pi(\hat{\pi}_{a^*,b^*}) \leq \mathcal{R}_\nu^{\pi_{a^*,b^*}}(\hat{\pi}_{a^*,b^*}) \leq \mathcal{R}_\nu^{\pi_{a^*,b^*}}(\hat{\pi}).$$

Damit ist $\mathbb{P}_{\pi_{a^*,b^*}}^\theta = \text{Beta}[a^*, b^*]$ ungünstigste a-priori Verteilung und $\hat{\pi}_{a^*,b^*}$ minimax-optimaler Schätzer von π . Insbesondere ist der natürliche Schätzer $\hat{\pi} = X/n$ mit $\mathcal{R}_\nu(\pi, \hat{\pi}) = \pi(1 - \pi)/n$ nicht minimax (er ist jedoch zulässig). □

§06.31 **Bemerkung.** Gehören für ein statistisches Modell die a-posteriori Verteilungen wieder zur der Klasse von a-priori Verteilungen (im Allgemeinen mit geänderten Parametern), so nennt man die entsprechenden Verteilungsklassen *konjugiert*. Zum Beispiel sind Beta-Verteilungen konjugiert zur Binomialverteilung (**Beispiel §06.30**). Konjugierte Verteilungen sind die Ausnahme, nicht die Regel, und für komplexere Modelle werden häufig Rechen-intensive Methoden wie MCMC (Markov Chain Monte Carlo) verwendet, um die a-posteriori Verteilung zu berechnen. □

§07 Das Stein-Phänomen

Für ein $d \in \mathbb{N}$ sei $X \odot N_{\mathbb{R}^d \times \{E_d\}}^n = (N_{(\mu, E_d)}^n)_{\mu \in \mathbb{R}^d}$. Wir betrachten das Entscheidungsproblem, den Parameter μ möglichst gut im Sinne eines quadratischen Verlustes $\nu(\mu, \hat{\mu}) = \|\hat{\mu} - \mu\|^2$ zu schätzen. Auf Grund der Unabhängigkeit der Koordinaten erscheint das (koordinatenweise) arithmetische Mittel \bar{X} , eine natürliche Antwort zu sein. Ein alternativer, sogenannter *empirischer Bayesansatz*, beruht auf der Familie der a-priori Verteilungen $(N_{(0, \sigma^2)}^d)_{\sigma \in \mathbb{R}_0^+}$. Betrachten wir für ein $\sigma \in \mathbb{R}_0^+$ einen zufälligen Parameter $\mu_\sigma \sim N_{(0, \sigma^2)}^d$, so ist der zugehörige Bayesschätzer $\mathbb{E}^{\mu_\sigma|X}(\mu_\sigma) = \frac{n}{n + \sigma^{-2}} \bar{X}$ (vgl. **Beweis §06.23**). Der empirische Bayesansatz beruht nun auf der Ersetzung von σ^2 durch die Schätzung $\hat{\sigma}^2 := \|\bar{X}\|^2/d - n^{-1}$. Da die Randverteilung von \bar{X} bezüglich der gemeinsamen Verteilung $\mathbb{P}^{\mu_\sigma, \bar{X}}$ gerade einer $N_{(0, \sigma^2 + n^{-1})}^d$ -Verteilung entspricht, ist $\hat{\sigma}^2$ ein erwartungstreuer Schätzer von σ^2 . Wir erhalten den Schätzer

$$\hat{\mu} = \frac{n}{n + \hat{\sigma}^{-2}} \bar{X} = \left(1 - \frac{d}{n\|\bar{X}\|^2}\right) \bar{X}.$$

Der Bayessche Ansatz lässt vermuten, dass für kleine Werte von $\|\mu\|$ der Schätzer $\hat{\mu}$ ein kleineres Risiko als \bar{X} hat. Überraschenderweise gilt für Dimension $d \geq 3$ sogar, dass $\hat{\mu}$ besser als \bar{X} ist. Das folgende Steinsche Lemma liefert das zentrale Argument für den Beweis.

§07.01 **Lemma von Stein.** *Es sei $f : \mathbb{R}^d \rightarrow \mathbb{R}$ eine in jeder Koordinate Lebesgue-f.ü. absolut stetige Funktion. Für jedes $\mu \in \mathbb{R}^d$ und $\sigma \in \mathbb{R}_0^+$ mit $Y \sim N_{(\mu, \sigma^2 E_d)}$ gilt dann*

$$\mathbb{E}_{(\mu, \sigma^2 E_d)}((\mu - Y)f(Y)) = -\sigma^2 \mathbb{E}_{(\mu, \sigma^2 E_d)}(\nabla f(Y)),$$

sofern $\mathbb{E}_{(\mu, \sigma^2 E_d)} \left(\left| \frac{\partial f}{\partial y_i}(Y) \right| \right) \in \mathbb{R}^+$ für alle $i \in \llbracket d \rrbracket$ gilt.

§07.02 **Beweis** von **Lemma** §07.01. In der Vorlesung. □

§07.03 **Satz**. Es sei $d \geq 3$ und $X \odot N_{\mathbb{R}^d \times \{E_d\}}^n = (N_{(\mu, E_d)}^n)_{\mu \in \mathbb{R}^d}$. Dann gilt für den *James-Stein-Schätzer*

$$\hat{\mu}_{JS} := \left(1 - \frac{d-2}{n\|\bar{X}\|^2}\right)\bar{X}$$

mit $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$, dass

$$\mathbb{E}_\mu \|\hat{\mu}_{JS} - \mu\|^2 = \frac{d}{n} - \mathbb{E}_\mu \left[\frac{(d-2)^2}{n^2 \|\bar{X}\|^2} \right] < \frac{d}{n} = \mathbb{E}_\mu \|\bar{X} - \mu\|^2.$$

Insbesondere ist \bar{X} für eine quadratische Verlustfunktion kein zulässiger Schätzer von μ im Fall $d \geq 3$.

§07.04 **Beweis** von **Satz** §07.03. In der Vorlesung. □

§07.05 **Bemerkungen**.

(a) Die Abbildung $\mu \mapsto \mathbb{E}_\mu [\|\bar{X}\|^{-2}]$ ist monoton fallend in $\|\mu\|$ und erfüllt $\mathbb{E}_0 [\|\bar{X}\|^{-2}] = n/(d-2)$ und $\mathbb{E}_0 \|\hat{\mu}_{JS}\|^2 = 2/n$. Damit ist $\hat{\mu}_{JS}$ für μ nahe 0, große Dimension d und kleine Stichprobenumfänge n eine deutliche Verbesserung von \bar{X} . Der James-Stein-Schätzer wird auch *Shrinkage-Schätzer* genannt, weil die Koordinaten des ursprünglichen Schätzers \bar{X} gedämpft (zur Null hingezogen) werden.

(b) Der *James-Stein-Schätzer mit positivem Gewicht*

$$\hat{\mu}_{JS+} := \left(1 - \frac{d-2}{n\|\bar{X}\|^2}\right)_+ \bar{X}, \quad (a)_+ := \max(a, 0),$$

ist bei quadratischer Verlustfunktion besser als der James-Stein-Schätzer $\hat{\mu}_{JS}$. Damit ist selbst der James-Stein-Schätzer (sogar mit positivem Gewicht) unzulässig. Die Konstruktion eines zulässigen Minimax-Schätzers ist gelöst für $d \geq 6$ (vgl. Lehmann and Casella [1998], S. 385). □

3 Schätztheorie

§08 Dominierte Modelle

§08.01 **Definition.** Ein statistisches Modell $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ heißt *dominiert*, falls ein σ -endliches Maß μ auf \mathcal{X} , kurz $\mu \in \mathcal{M}_\sigma(\mathcal{X})$, existiert, so dass für jedes $\theta \in \Theta$ das Wahrscheinlichkeitsmaß \mathbb{P}_θ absolut stetig bezüglich μ ist ($\mathbb{P}_\theta \ll \mu$). Eine durch θ paramisierte Radon-Nikodym-Dichte

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

wird auch *Likelihood-Funktion* genannt, wobei diese meist als zufällige Funktion $L : \Theta \rightarrow \mathcal{X}^+$ mit $\theta \mapsto L(\theta) := L(\theta, \bullet)$ aufgefasst wird. □

§08.02 **Beispiel.**

- (a) Ein statistisches Experiment $(\mathbb{R}, \mathcal{B}, \mathbb{P}_\theta)$ ist trivialerweise dominiert wenn jedes $\mathbb{P}_\theta \in \mathbb{P}$ durch eine Lebesgue-dichte f_θ gegeben ist, beispielsweise $N_{\mathbb{R}^p \times \mathbb{R}_0^+}$.
- (b) Jedes statistische Modell mit Stichprobenraum $\mathcal{X} = \mathbb{N}$ und Potenzmenge $\mathcal{X} = 2^{\mathbb{N}}$ oder allgemeiner mit abzählbarem Stichprobenraum \mathcal{X} und Potenzmenge $\mathcal{X} = 2^{\mathcal{X}}$ ist dominiert bezüglich des Zählmaßes $\zeta_{\mathcal{X}}$ (vgl. **Beispiel B01.03 (b)**).
- (c) Ist die Parametermenge $\Theta = \{\theta_i : i \in \mathbb{N}\}$ abzählbar, so ist $\mu = \sum_{i \in \mathbb{N}} c_i \mathbb{P}_{\theta_i}$ mit $c_i > 0$, $\sum_{i \in \mathbb{N}} c_i = 1$ ein dominierendes Maß.
- (d) Sei δ_x das Punktmaß in $x \in \mathbb{R}$. Das statistische Experiment $(\mathbb{R}, \mathcal{B}, (\delta_\theta)_{\theta \in \mathbb{R}})$ ist nicht dominiert. Für ein dominierendes Maß $\mu \in \mathcal{M}_\sigma(\mathcal{B})$ müsste $\mu(\{\theta\}) \in \mathbb{R}_0^+$ für alle $\theta \in \mathbb{R}$ gelten und damit $\mu(A) = \infty$ für jede überabzählbare Borelmenge $A \subseteq \mathbb{R}$ erfüllen (sonst folgt aus $|\{x \in A | \mu(\{x\}) \geq 1/n\}| \leq n\mu(A) \in \mathbb{R}^+$, dass $A = \cup_{n \geq 1} \{x \in A | \mu(\{x\}) \geq 1/n\}$ abzählbar ist). Damit kann μ nicht σ -endlich sein. □

§08.03 **Bemerkung.** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein bezüglich $\mu \in \mathcal{M}_\sigma(\mathcal{X})$ dominiertes Modell. Ist μ endlich, kurz $\mu \in \mathcal{M}_e(\mathcal{X})$, so gilt $\mu \ll \mathbb{P}_\mu := \frac{1}{\mu(\mathcal{X})}\mu \in \mathcal{W}(\mathcal{X})$ und damit ist \mathbb{P}_θ auch durch \mathbb{P}_μ dominiert. Andererseits, ist μ nicht endlich, so existiert eine abzählbare, messbare Partition $\{\mathcal{X}_m, m \in \mathbb{N}\}$ von \mathcal{X} mit $\mu(\mathcal{X}_m) \in \mathbb{R}_0^+$ für jedes $m \in \mathbb{N}$. Für jedes $m \in \mathbb{N}$ definiere $\mathbb{P}_\mu(\bullet | \mathcal{X}_m) \in \mathcal{W}(\mathcal{X})$ mit $A \mapsto \mathbb{P}_\mu(A | \mathcal{X}_m) := \frac{\mu(A \cap \mathcal{X}_m)}{\mu(\mathcal{X}_m)}$. Dann gilt $\mu \ll \mathbb{P}_\mu := \sum_{m \in \mathbb{N}} 2^{-m} \mathbb{P}_\mu(\bullet | \mathcal{X}_m) \in \mathcal{W}(\mathcal{X})$, da aus $\mathbb{P}_\mu(A) = 0$ insbesondere $\mu(A \cap \mathcal{X}_m) = 0$ für alle $m \in \mathbb{N}$ und somit $\mu(A) = 0$ folgt. Damit haben wir gezeigt, dass zu $\mu \in \mathcal{M}_\sigma(\mathcal{X})$ ein $\mathbb{P}_\mu \in \mathcal{W}(\mathcal{X})$ mit $\mu \ll \mathbb{P}_\mu$ existiert, dass dann automatisch \mathbb{P}_θ dominiert. Die nächste Aussage zeigt, dass ein \mathbb{P}_θ dominierendes $\mathbb{P}_\theta \in \mathcal{W}(\mathcal{X})$ existiert mit $\mathbb{P}_\theta \ll \mu$. Im Folgenden können wir also ohne Beschränkung der Allgemeinheit annehmen, dass das dominierende Maß ein Wahrscheinlichkeitsmaß ist. □

§08.04 **Satz.** Es sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein μ -dominiertes statistisches Modell. Dann existiert ein Wahrscheinlichkeitsmaß \mathbb{P}_θ der Form $\sum_{i \in \mathbb{N}} c_i \mathbb{P}_{\theta_i}$ mit $c_i \in \mathbb{R}^+$, $\theta_i \in \Theta$ für alle $i \in \mathbb{N}$ und $\sum_{i \in \mathbb{N}} c_i = 1$, also $\mathbb{P}_\theta \ll \mu$, derart dass $\mathbb{P}_\theta \ll \mathbb{P}_\theta$ für alle $\theta \in \Theta$ gilt. Ein solches Wahrscheinlichkeitsmaß \mathbb{P}_θ wird *privilegiertes dominierendes Maß* genannt.

§08.05 **Beweis** von **Satz** §08.04. In der Vorlesung. □

§09 Erschöpfende Statistik

§09.01 **Beispiel.** Es sei $(X_i)_{i \in \llbracket n \rrbracket} \odot \mathbb{P}_\theta^n$ und jedes \mathbb{P}_θ sei durch eine Lebesguedichte $f_\theta \in \mathcal{B}^+$ gegeben. Allgemeine Informationen über \mathbb{P}_θ und somit θ erhalten wir typischerweise mit Hilfe von Statistiken wie \bar{X} oder $\max\{X_i : i \in \llbracket n \rrbracket\}$. Intuitiv, enthält die **Ordnungsstatistik** $(X_{(i)})_{i \in \llbracket n \rrbracket}$ mit $X_{(1)} = \min\{X_i : i \in \llbracket n \rrbracket\}$, $X_{(k+1)} := \min\{X_i : i \in \llbracket n \rrbracket\} \setminus \{X_{(i)} : i \in \llbracket k \rrbracket\}$, $k \in \llbracket n-1 \rrbracket$, wie jede Statistik nicht mehr Informationen über den Parameter θ . Wir werden im Folgenden zeigen, dass die Ordnungsstatistik keine Information verliert. \square

§09.02 **Definition.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X = (\mathbb{P}_\theta^X)_{\theta \in \Theta})$ und $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S = (\mathbb{P}_\theta^S)_{\theta \in \Theta})$ zwei statistische Experimente zum selben Parameterraum Θ . $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X)$ heißt **informativer** als $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$, falls für alle Entscheidungsprobleme $(\mathcal{E}, \mathcal{E}, \nu)$ mit $\|\nu\|_\infty := \sup_{\theta, e} |\nu(\theta, e)| \in \mathbb{R}^+$ und für alle Entscheidungskerne D_S von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{E}, \mathcal{E})$ ein Entscheidungskern D_X von $(\mathcal{X}, \mathcal{X})$ nach $(\mathcal{E}, \mathcal{E})$ existiert mit (vgl. **Definition** §05.14)

$$\mathcal{R}_\nu(\theta, D_X) = \int_{\mathcal{E}} \nu(\theta, \bullet) d[D_X \mathbb{P}_\theta^X] \leq \int_{\mathcal{E}} \nu(\theta, \bullet) d[D_S \mathbb{P}_\theta^S] = \mathcal{R}_\nu(\theta, D_S) \quad \forall \theta \in \Theta.$$

Ist $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X)$ informativer als $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$ und $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$ informativer als $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X)$, dann heißen $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X)$ und $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$ **äquivalent**. \square

§09.03 **Lemma.** Existiert ein Markovkern κ von $(\mathcal{X}, \mathcal{X})$ nach $(\mathcal{S}, \mathcal{S})$ mit $\kappa \mathbb{P}_\theta^X = \mathbb{P}_\theta^S$ für alle $\theta \in \Theta$ (vgl. **Schreibweise** C02.06), dann ist $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta^X)$ informativer als $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$.

§09.04 **Beweis** von **Lemma** §09.03. In der Vorlesung. \square

§09.05 **Korollar.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment, S eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ und \mathbb{P}_θ^S die induzierte Verteilungsfamilie auf $(\mathcal{S}, \mathcal{S})$. Dann ist $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ informativer als $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$.

§09.06 **Beweis** von **Korollar** §09.05. Übung. \square

§09.07 **Beispiel (Beispiel §01.06 (a) fortgesetzt).** Betrachte $X \odot \mathbb{N}_{\mathbb{R} \times \{1\}}^n$, also das **normale Lokations-Modell** und $S : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $x = (x_i)_{i \in \llbracket n \rrbracket} \mapsto S(x) := \bar{x} = n^{-1} \sum_{i \in \llbracket n \rrbracket} x_i$, dann gilt $S(X) = \bar{X} \odot \mathbb{N}_{\mathbb{R} \times \{n-1\}}$. Insbesondere folgt aus **Korollar** §09.05, dass das normale Lokations-Modell $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{N}_{\mathbb{R} \times \{1\}}^n)$ informativer ist als das statistische Experiment $(\mathbb{R}, \mathcal{B}, \mathbb{N}_{\mathbb{R} \times \{n-1\}})$. Wir werden im nächsten Abschnitt zeigen, dass die statistischen Experimente äquivalent sind. \square

§09.08 **Definition.** Sei \mathbb{P}_θ eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$.

- Eine Teil- σ -Algebra \mathcal{F} von \mathcal{X} heißt **erschöpfend** oder **suffizient** (für \mathbb{P}_θ), wenn es für alle $A \in \mathcal{X}$ eine von $\theta \in \Theta$ unabhängige Festlegung des bedingten Erwartungswertes $\mathbb{E}_\theta(\mathbb{1}_A | \mathcal{F})$ von $\mathbb{1}_A$ gegeben \mathcal{F} gibt, das heißt, wenn für alle $A \in \mathcal{X}$ ein $h_A \in \overline{\mathcal{F}}^+$ mit $h_A = \mathbb{P}_\theta(A | \mathcal{F}) = \mathbb{E}_\theta(\mathbb{1}_A | \mathcal{F})$ \mathbb{P}_θ -f.ü. für alle $\theta \in \Theta$ existiert.
- Eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X})$ heißt **erschöpfend** oder **suffizient** (für \mathbb{P}_θ), wenn die induzierte Teil- σ -Algebra $\sigma(S)$ erschöpfend (für \mathbb{P}_θ) ist. \square

§09.09 **Bemerkung.** Nach **Eigenschaft** A02.06 (iv) ist eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ genau dann erschöpfend (für \mathbb{P}_θ), wenn es für alle $A \in \mathcal{X}$ und alle $\theta \in \Theta$ eine von θ unabhängige Funktion $h_A \in \overline{\mathcal{S}}^+$ gibt mit $h_A = \mathbb{P}_\theta(A | S)$ \mathbb{P}_θ^S -f.ü. für alle $\theta \in \Theta$ (vgl. **Definition** C02.23). \square

§09.10 **Satz.** Sei \mathbb{P}_θ eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$, $\mathcal{F} \subseteq \mathcal{X}$ eine Teil- σ -Algebra bzw. S eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X})$ sowie $Y \in \overline{\mathcal{X}}$ eine weitere numerische Statistik, für die $\mathbb{E}_\theta(Y)$ für alle $\theta \in \Theta$ existiert. Dann gilt:

- (i) Ist \mathcal{F} erschöpfend (für \mathbb{P}_θ), so gibt es eine von θ unabhängige Festlegung $\mathbb{E}_\theta(Y|\mathcal{F})$ des bedingten Erwartungswertes $\mathbb{E}_\theta(Y|\mathcal{F})$ von Y gegeben \mathcal{F} , das heißt, $\mathbb{E}_\theta(Y|\mathcal{F}) \in \overline{\mathcal{F}}$ und $\mathbb{E}_\theta(Y|\mathcal{F}) = \mathbb{E}_\theta(Y|\mathcal{F})$ \mathbb{P}_θ -f.ü. für alle $\theta \in \Theta$.
- (ii) Ist S erschöpfend, so gibt es eine von θ unabhängige Festlegung $\mathbb{E}_\theta(Y|S)$ des bedingten Erwartungswertes $\mathbb{E}_\theta(Y|S)$ von Y gegeben S , das heißt, $\mathbb{E}_\theta(Y|S) \in \overline{\mathcal{S}}$ und $\mathbb{E}_\theta(Y|S) = \mathbb{E}_\theta(Y|S)$ \mathbb{P}_θ^S -f.ü. für alle $\theta \in \Theta$.
- (iii) Für die Festlegung $\mathbb{E}_\theta(Y|\mathcal{F})$ bzw. $\mathbb{E}_\theta(Y|S)$ des bedingten Erwartungswertes von Y gilt $\mathbb{E}_\theta(\mathbb{E}_\theta(Y|\mathcal{F})) = \mathbb{E}_\theta^S(\mathbb{E}_\theta(Y|S)) = \mathbb{E}_\theta(Y)$ für alle $\theta \in \Theta$.

§09.11 **Beweis** von Satz §09.10. In der Vorlesung. □

§09.12 **Bemerkung.** Eine von θ unabhängige Festlegung $\mathbb{E}_\theta(Y|\mathcal{F})$ des bedingten Erwartungswertes $\mathbb{E}_\theta(Y|\mathcal{F})$ ist bis auf eine \mathbb{P}_θ -Nullmenge in \mathcal{F} (kurz $\mathbb{P}_\theta^{\mathcal{F}}$ -f.ü.) eindeutig bestimmt, also für zwei Festlegungen $\mathbb{E}_\theta^1(Y|\mathcal{F})$ und $\mathbb{E}_\theta^2(Y|\mathcal{F})$ existiert $N \in \mathcal{F}$ mit $\mathbb{E}_\theta^1(Y|\mathcal{F})(x) = \mathbb{E}_\theta^2(Y|\mathcal{F})(x)$ für alle $x \in N^c$ und $\mathbb{P}_\theta(N) = 0$ für alle $\theta \in \Theta$. Entsprechend ist eine von θ unabhängige Festlegung $\mathbb{E}_\theta(Y|S)$ des bedingten Erwartungswertes $\mathbb{E}_\theta(Y|S)$ \mathbb{P}_θ^S -f.ü., also bis auf eine \mathbb{P}_θ^S -Nullmenge in \mathcal{S} , eindeutig bestimmt. □

§09.13 **Satz.** Sei \mathbb{P} eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ und $\mathcal{F} \subseteq \mathcal{X}$ eine für \mathbb{P} erschöpfende Teil- σ -Algebra über \mathcal{X} . Weiter sei $Y \in \mathcal{X}^q$ ein Zufallsvektor. Dann gilt:

- (i) Es gibt stets eine von θ unabhängige reguläre Festlegung $\mathbb{P}^{Y|\mathcal{F}}$ der bedingten Verteilung $\mathbb{P}_\theta^{Y|\mathcal{F}}$ von Y gegeben \mathcal{F} , das heißt, $\mathbb{P}^{Y|\mathcal{F}}$ ist ein Markovkern von $(\mathcal{X}, \mathcal{F})$ nach $(\mathbb{R}^q, \mathcal{B}^q)$.
- (ii) Ist $h \in \mathcal{B}^q$ eine Funktion, für die $\mathbb{E}_\theta^Y(h) = \int h(y)\mathbb{P}_\theta^Y(dy)$ für alle $\theta \in \Theta$ existiert, so gilt $\mathbb{E}_\theta(h(Y)|\mathcal{F}) = \mathbb{E}_\theta^{Y|\mathcal{F}}(h) = \int h(y)\mathbb{P}^{Y|\mathcal{F}}(dy)$ $\mathbb{P}_\theta^{\mathcal{F}}$ -f.ü..

§09.14 **Beweis** von Satz §09.13. Witting [1985], Satz 3.15, S. 341 □

§09.15 **Lemma.** Ist eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ mit induzierter Verteilungsfamilie \mathbb{P}_θ^S auf $(\mathcal{S}, \mathcal{S})$ erschöpfend für \mathbb{P} mit regulärer Festlegung $\mathbb{P}(\cdot|S)$, dann sind die statistischen Experimente $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ und $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$ äquivalent.

§09.16 **Beweis** von Lemma §09.15. In der Vorlesung. □

§09.17 **Bemerkung.** Seien \mathcal{X} polnisch, \mathcal{X} die Borel- σ -Algebra über \mathcal{X} und $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein bzgl. eines σ -endlichen Maßes dominiertes statistisches Experiment. Dann ist eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ mit induzierter Verteilungsfamilie \mathbb{P}_θ^S auf $(\mathcal{S}, \mathcal{S})$ genau dann erschöpfend, wenn die statistischen Experimente $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ und $(\mathcal{S}, \mathcal{S}, \mathbb{P}_\theta^S)$ äquivalent sind. □

§09.18 **Vorbemerkung.** Für die späteren Überlegungen interessieren besonders von θ unabhängige Festlegungen der bedingten Erwartung einer $(\mathcal{T}, \mathcal{T})$ -wertigen Statistik T auf $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ bei gegebener (für \mathbb{P}) suffizienter $(\mathcal{S}, \mathcal{S})$ -wertigen Statistik S auf $(\mathcal{X}, \mathcal{X})$ und demgemäß von θ unabhängige reguläre Festlegungen $\mathbb{P}^{T|S}$ und $\mathbb{P}^{(S,T)|S}$. Diese sind wieder als von θ unabhängige Markovkerne von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{T}, \mathcal{T})$ bzw. nach $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{T})$ erklärt, die bei festem $A \in \mathcal{T}$ bzw. bei festem $D \in \mathcal{S} \otimes \mathcal{T}$ die folgenden Gleichungen erfüllen (vgl. Erinnerung C03.07 und Eigenschaft C03.08)

$$\begin{aligned} \mathbb{P}_\theta^{(S,T)}(B \times A) &= \mathbb{E}_\theta^S(\mathbb{1}_B \mathbb{P}^{T|S}(A)) \quad \forall B \in \mathcal{S} \quad \forall \theta \in \Theta, \\ \mathbb{P}_\theta^{(S,T)}(D \cap (B \times \mathcal{T})) &= \mathbb{E}_\theta^S(\mathbb{1}_B \mathbb{P}^{(S,T)|S}(D)) \quad \forall B \in \mathcal{S} \quad \forall \theta \in \Theta. \end{aligned}$$

Dabei ergibt sich $\mathbb{P}^{T|S}$ aus $\mathbb{P}^{(S,T)|S}$ gemäß $\mathbb{P}^{T|S}(A) = \mathbb{P}^{(S,T)|S}(\mathcal{S} \times A)$, $A \in \mathcal{T}$, also $\mathbb{P}^{T|S}$ ist gerade die Randverteilung bezüglich $\mathbb{P}^{(S,T)|S}$. □

§09.19 **Satz.** Seien \mathbb{P}_θ eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$, S eine (für \mathbb{P}_θ) suffiziente $(\mathcal{S}, \mathcal{S})$ -wertige Statistik und T eine weitere $(\mathcal{T}, \mathcal{T})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X})$.

- (i) Ist $(\mathcal{T}, \mathcal{T}) = (\mathbb{R}^k, \mathcal{B}^k)$, so existiert stets ein von θ unabhängige reguläre Festlegung $\mathbb{P}_\theta^{T|S}$ der bedingten Verteilung von T gegeben S .
- (ii) Es existiere eine von θ unabhängige reguläre Festlegung $\mathbb{P}_\theta^{T|S}$ der bedingten Verteilung von T gegeben S . Dann ist $\mathbb{P}_\theta^{(S,T)|S}$ mit

$$(s, D) \mapsto \mathbb{P}_\theta^{(S,T)|S=s}(D) := \mathbb{P}_\theta^{T|S=s}(D_s) \quad \text{mit } D_s := \{t \in \mathcal{T} : (s, t) \in D\}$$

eine von θ unabhängige reguläre Festlegung der bedingten Verteilung von (S, T) gegeben S . Ist $h \in \overline{\mathcal{T} \otimes \mathcal{S}}$ eine Funktion, für die $\mathbb{E}_\theta^{(S,T)}(h)$ existiert für alle $\theta \in \Theta$, so gilt

$$\begin{aligned} \mathbb{E}_\theta^{(S,T)|S=s}(h) &= \mathbb{E}_\theta^{T|S=s}(h_s) = \int_{\mathcal{T}} h_s(t) \mathbb{P}_\theta^{T|S=s}(dt) \quad \mathbb{P}_\theta^S\text{-f.ü. mit } h_s := h(s, \bullet) \in \overline{\mathcal{T}}, \\ \mathbb{E}_\theta^{(S,T)}(h \mathbb{1}_{B \times A}) &= \int_B \int_A h_s(t) \mathbb{P}_\theta^{T|S=s}(dt) \mathbb{P}_\theta^S(ds) \quad \forall A \in \mathcal{T}, \forall B \in \mathcal{S}, \forall \theta \in \Theta. \end{aligned}$$

§09.20 **Beweis** von **Satz** §09.19. Witting [1985], Satz 3.16, S. 341 (vgl. **Eigenschaft** C03.08) □

§09.21 **Bemerkung.** Eine Statistik T auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ wird *unwesentlich* (ancillary) genannt, wenn ihre Verteilung $\mathbb{P}_\theta^T := \mathbb{P}_\theta^T$ nicht vom Parameter θ abhängt. □

§09.22 **Satz.** Seien \mathbb{P}_θ eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ und (S, T) eine $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{T})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X})$. S sei erschöpfend für \mathbb{P}_θ und es geben eine von θ unabhängige reguläre Festlegung $\mathbb{P}_\theta^{T|S}$ der bedingten Verteilung von T gegeben S . Dann sind äquivalent:

- (i) T und S sind stochastisch unabhängig unter jedem \mathbb{P}_θ , $\theta \in \Theta$, und T ist unwesentlich.
- (ii) $\mathbb{P}_\theta^{T|S=s}$ ist unabhängig von s wählbar.
- (iii) T ist unwesentlich, und es gilt $\mathbb{P}_\theta^{T|S} = \mathbb{P}_\theta^T \mathbb{P}_\theta^S$ -f.ü..

§09.23 **Beweis** von **Satz** §09.22. Witting [1985], Satz 3.17, S. 342 □

§09.24 **Satz (Halmos-Savage).** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein dominiertes statistisches Experiment und \mathbb{P}_θ ein privilegiertes dominierendes Wahrscheinlichkeitsmaß. Dann gilt:

- (i) Eine Teil- σ -Algebra \mathcal{F} von \mathcal{X} ist genau dann erschöpfend (für \mathbb{P}_θ), wenn es für jedes $\theta \in \Theta$ eine \mathcal{F} -messbare \mathbb{P}_θ -Dichte von \mathbb{P}_θ gibt, das heißt, für alle $\theta \in \Theta$ existiert $f_\theta \in \overline{\mathcal{F}}^+$ mit $d\mathbb{P}_\theta/d\mathbb{P}_\theta = f_\theta \mathbb{P}_\theta$ -f.ü..
- (ii) Eine $(\mathcal{S}, \mathcal{S})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X})$ ist genau dann erschöpfend (für \mathbb{P}_θ), wenn es für jedes $\theta \in \Theta$ ein $f_\theta \in \overline{\mathcal{S}}^+$ mit $d\mathbb{P}_\theta/d\mathbb{P}_\theta = f_\theta(S) \mathbb{P}_\theta$ -f.ü. gibt.

§09.25 **Beweis** von **Satz** §09.24. In der Vorlesung. □

§09.26 **Vorbetrachtung.** Sei $\mathbb{P}_\theta = (\mathbb{P}_{\theta,\tau})_{(\theta,\tau) \in \Theta}$ eine dominierte Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ und (S, T) eine für \mathbb{P}_θ erschöpfende, $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{T})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X})$, derart dass für ein privilegiertes dominierendes Wahrscheinlichkeitsmaß \mathbb{P}_θ es für jedes $\mathbb{P}_{\theta,\tau}$ mit $(\theta, \tau) \in \Theta$ eine Normierungskonstante $c_{\theta,\tau} \in \mathbb{R}_0^+$ sowie Funktionen $f_\theta \in \overline{\mathcal{S}}^+$ und $g_\tau \in \overline{\mathcal{T}}^+$ mit

$$\frac{d\mathbb{P}_{\theta,\tau}}{d\mathbb{P}_\theta} = c_{\theta,\tau} f_\theta(S) g_\tau(T) \quad \mathbb{P}_\theta\text{-f.ü.} \tag{09.01}$$

gibt. Weiterhin existiere eine reguläre Festlegung $\mathbb{E}_o^{S|T}$ der bedingten Erwartung von S gegeben T bzgl. \mathbb{P}_o . Für jedes $(\theta, \tau) \in \Theta$ ist das Bildmaß $\mathbb{P}_{\theta, \tau}^T$ absolut stetig bzgl. des Bildmaßes \mathbb{P}_o^T , wobei für $f_\theta^T := \mathbb{E}_o^{S|T}(f_\theta) \in \overline{\mathcal{T}}^+$ und für jedes $A \in \mathcal{T}$ gilt

$$\begin{aligned} \mathbb{P}_{\theta, \tau}^T(A) &= \mathbb{P}_{\theta, \tau}^{(S, T)}(\mathcal{S} \times A) = \mathfrak{c}_{\theta, \tau} \int_{\mathcal{S} \times A} f_\theta g_\tau d\mathbb{P}_o^{(S, T)} \\ &= \mathfrak{c}_{\theta, \tau} \int_A g_\tau(t) \mathbb{E}_o^{S|T=t}(f_\theta) \mathbb{P}_o^T(dt) = \mathfrak{c}_{\theta, \tau} \int_A g_\tau f_\theta^T d\mathbb{P}_o^T. \end{aligned}$$

$\mathfrak{c}_{\theta, \tau} g_\tau f_\theta^T$ ist somit eine Festlegung der \mathbb{P}_o^T -Dichte von $\mathbb{P}_{\theta, \tau}^T$ und \mathbb{P}_o^T -f.ü. gilt $f_\theta^T \in \mathbb{R}_o^+$. Für jedes $A \in \mathcal{S}$ ist definitionsgemäß $\mathbb{E}_o^{S|T}(\mathbb{1}_A f_\theta) \in \overline{\mathcal{T}}^+$ und für jedes $t \in \mathcal{T}$ mit $f_\theta^T(t) \in \mathbb{R}_o^+$ ist $A \mapsto (f_\theta^T(t))^{-1} \mathbb{E}_o^{S|T=t}(\mathbb{1}_A f_\theta)$ ein Wahrscheinlichkeitsmaß über $(\mathcal{T}, \mathcal{S})$. Damit ist die Abbildung

$$\begin{aligned} \mathbb{P}_{\theta, \bullet}^{S|T} : \mathcal{T} \times \mathcal{S} &\rightarrow \mathbb{R}^+ \quad \text{mit} \\ (t, B) &\mapsto \mathbb{P}_{\theta, \bullet}^{S|T=t}(B) := \mathbb{E}_o^{S|T=t}(\mathbb{1}_B \frac{f_\theta}{f_\theta^T(t)}) \mathbb{1}_{\{f_\theta^T(t) \in \mathbb{R}_o^+\}} + \mathbb{P}_o^S(B) \mathbb{1}_{\{f_\theta^T(t) \notin \mathbb{R}_o^+\}} \quad (09.02) \end{aligned}$$

ein Markovkern von $(\mathcal{T}, \mathcal{S})$ nach $(\mathcal{S}, \mathcal{S})$, für den weiterhin gilt

$$\begin{aligned} [\mathbb{P}_{\theta, \tau}^T \odot \mathbb{P}_{\theta, \bullet}^{S|T}](A \times B) &= \int_A \mathbb{P}_{\theta, \bullet}^{S|T=t}(B) \mathbb{P}_{\theta, \tau}^T(dt) = \mathfrak{c}_{\theta, \tau} \int_A \mathbb{P}_{\theta, \bullet}^{S|T=t}(B) g_\tau(t) f_\theta^T(t) \mathbb{P}_o^T(dt) \\ &= \mathfrak{c}_{\theta, \tau} \int_A \mathbb{E}_o^{S|T=t}(\mathbb{1}_B \frac{f_\theta}{f_\theta^T(t)}) \mathbb{1}_{\{f_\theta^T(t) \in \mathbb{R}_o^+\}} g_\tau(t) f_\theta^T(t) \mathbb{P}_o^T(dt) \\ &= \mathfrak{c}_{\theta, \tau} \int_A \mathbb{E}_o^{S|T=t}(\mathbb{1}_B f_\theta) g_\tau(t) \mathbb{P}_o^T(dt) = \int_{A \times B} \mathfrak{c}_{\theta, \tau} f_\theta g_\tau d\mathbb{P}_o^{(T, S)} = \mathbb{P}_{\theta, \tau}^{(T, S)}(A \times B). \end{aligned}$$

Für jedes $(\theta, \tau) \in \Theta$ ist somit $\mathbb{P}_{\theta, \bullet}^{S|T}$ eine von τ unabhängige reguläre Festlegung der bedingten Verteilung $\mathbb{P}_{\theta, \tau}^{S|T}$ und für \mathbb{P}_o^T -f.a. $t \in \mathcal{T}$ ist $\frac{1}{f_\theta^T(t)} f_\theta$ eine $\mathbb{P}_o^{S|T=t}$ -Dichte der Verteilung $\mathbb{P}_{\theta, \bullet}^{S|T=t}$. Für jedes θ mit $\Theta^\theta := \{\tau : (\theta, \tau) \in \Theta\}$ und $\Theta^\bullet := \{\theta : \Theta^\theta \neq \emptyset\}$ ist T somit erschöpfend für $\mathbb{P}_{\{\theta\} \times \Theta^\theta} := (\mathbb{P}_{\theta, \tau})_{\tau \in \Theta^\theta}$ und nach Satz §09.24 ist \mathbb{P}_o^T -f.ü. die Statistik S erschöpfend für die $\mathbb{P}_o^{S|T=t}$ -dominierte Familie $(\mathbb{P}_{\theta, \bullet}(\bullet | T = t))_{\theta \in \Theta^\bullet}$. \square

§09.27 **Lemma.** Sei \mathbb{P}_o eine dominierte Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ und (S, T) eine für \mathbb{P}_o erschöpfende, $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{T})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X})$, derart dass für ein privilegiertes dominierendes Wahrscheinlichkeitsmaß \mathbb{P}_o jedes $\mathbb{P}_{\theta, \tau}^{(S, T)}$ mit $(\theta, \tau) \in \Theta$ eine $\mathbb{P}_o^{(S, T)}$ -Dichte der Form (09.01) mit $\mathfrak{c}_{\theta, \tau} \in \mathbb{R}_o^+$, $f_\theta \in \overline{\mathcal{T}}^+$ und $g_\tau \in \overline{\mathcal{T}}^+$ besitzt. Weiterhin sei $\mathbb{E}_o^{S|T}$ eine reguläre Festlegung der bedingten Erwartung von S gegeben T bzgl. \mathbb{P}_o . Dann gilt:

- (i) Für jedes $(\theta, \tau) \in \Theta$ ist $\mathfrak{c}_{\theta, \tau} f_\theta^T g_\tau$ eine \mathbb{P}_o^T -Dichte von $\mathbb{P}_{\theta, \tau}^T$ mit $f_\theta^T := \mathbb{E}_o^{S|T}(f_\theta) \in \overline{\mathcal{T}}^+$.
- (ii) Für jedes $(\theta, \tau) \in \Theta$ ist $\mathbb{P}_{\theta, \bullet}^{S|T}$ gemäß (09.02) eine von τ unabhängige reguläre Festlegung der bedingten Verteilung $\mathbb{P}_{\theta, \tau}^{S|T}$ mit der Eigenschaft, dass für jedes $t \in \mathcal{T}$ mit $f_\theta^T(t) \in \mathbb{R}_o^+$ die Funktion $\frac{1}{f_\theta^T(t)} f_\theta$ eine $\mathbb{P}_o^{S|T=t}$ -Dichte von $\mathbb{P}_{\theta, \bullet}^{S|T=t}$ ist.

Für jedes $\theta \in \Theta^\bullet := \{\theta : \Theta^\theta \neq \emptyset\}$ mit $\Theta^\theta := \{\tau : (\theta, \tau) \in \Theta\}$ ist T somit erschöpfend für $\mathbb{P}_{\{\theta\} \times \Theta^\theta} := (\mathbb{P}_{\theta, \tau})_{\tau \in \Theta^\theta}$ und für \mathbb{P}_o^T -f.a. $t \in \mathcal{T}$ ist S erschöpfend für $(\mathbb{P}_{\theta, \bullet}(\bullet | T = t))_{\theta \in \Theta^\bullet}$.

§09.28 **Beweis** von Lemma §09.27. Die Aussage folgt direkt aus der Bemerkung §09.26. \square

§09.29 **Faktorisierungskriterium von Neyman.** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_o)$ ein dominiertes statistisches Experiment bzgl. $\mu \in \mathcal{M}_o(\mathcal{X})$. Dann gilt:

- (i) Eine Teil- σ -Algebra \mathcal{F} von \mathcal{X} ist genau dann erschöpfend (für \mathbb{P}_θ), wenn es $h \in \overline{\mathcal{X}}^+$ und $f_\theta \in \overline{\mathcal{F}}^+$ für jedes $\theta \in \Theta$ mit $[d\mathbb{P}_\theta/d\mu](x) = f_\theta(x)h(x)$ für μ -f.a. $x \in \mathcal{X}$ gibt.
- (ii) Eine $(\mathcal{S}, \mathcal{I})$ -wertige Statistik S auf $(\mathcal{X}, \mathcal{X})$ ist genau dann erschöpfend (für \mathbb{P}_θ), wenn es $h \in \overline{\mathcal{X}}^+$ und $f_\theta \in \overline{\mathcal{I}}^+$ für jedes $\theta \in \Theta$ mit $d\mathbb{P}_\theta/d\mu = f_\theta(S)h$ μ -f.ü. gibt.

§09.30 **Beweis** von Satz §09.29. In der Vorlesung. □

§09.31 **Beispiel.**

- (a) Die Identität $\text{id}_x(x) = x$ und allgemein jede bijektive, bi-messbare Transformation S ist stets erschöpfend.
- (b) Sind $(X_i)_{i \in [n]} \odot \mathbb{P}_\theta^n$ und für jedes $\theta \in \Theta$ ist \mathbb{P}_θ durch eine Lebesguedichte $f_\theta \in \overline{\mathcal{B}}^+$ gegeben, so ist die Ordnungsstatistik $(X_{(i)})_{i \in [n]}$ erschöpfend, da die Likelihood-Funktion sich in der Form $L(\theta, x) = \prod_{i \in [n]} f_\theta(x_{(i)})$ schreiben lässt. □

§09.32 **Vorbetrachtung.** Sei (S, T) eine $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{I})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ und die induzierte Verteilungsfamilie $\mathbb{P}_\theta^{(S, T)} = (\mathbb{P}_\theta^{(S, T)})_{\theta \in \Theta}$ sei dominiert mit $\mu \otimes \nu$ -Dichten $(f_\theta^{(S, T)})_{\theta \in \Theta}$ bei σ -endlichen Maßen $\mu \in \mathcal{M}_\sigma(\mathcal{S})$ und $\nu \in \mathcal{M}_\sigma(\mathcal{I})$. Wir bezeichnen weiterhin mit $\mathbb{P}_\theta^S = (\mathbb{P}_\theta^S)_{\theta \in \Theta}$ und $\mathbb{P}_\theta^T = (\mathbb{P}_\theta^T)_{\theta \in \Theta}$ die entsprechenden Familien von Randverteilungen über $(\mathcal{S}, \mathcal{S})$ bzw. $(\mathcal{T}, \mathcal{I})$ sowie mit $(f_\theta^S)_{\theta \in \Theta}$ und $(f_\theta^T)_{\theta \in \Theta}$ ihre μ bzw. ν -Dichten. In dieser Situation erlaubt das Neyman-Kriterium Satz §09.29 den konstruktiven Nachweis der Existenz von Markovkernen $\mathbb{P}^{T|S}$ und $\mathbb{P}^{(S, T)|S}$ bei Suffizienz von S . Dieser beruht auf der Existenz einer von θ unabhängigen Festlegung $f_\bullet^{T|S}$ der ν -Dichte des Markovkerns $\mathbb{P}^{T|S}$. Der Konstruktion in Schreibweise §06.08 folgend ist analog zu (06.01) für jedes $\theta \in \Theta$ die Abbildung

$$f_\theta^{T|S} : \mathcal{S} \times \mathcal{T} \rightarrow \overline{\mathbb{R}}^+ \text{ mit}$$

$$(s, t) \mapsto f_\theta^{T|S=s}(t) = \frac{f_\theta^{(S, T)}(s, t)}{f_\theta^S(s)} \mathbb{1}_{\{f_\theta^S(s) \in \mathbb{R}_0^+\}} + f_\theta^T(t) \mathbb{1}_{\{f_\theta^S(s) \notin \mathbb{R}_0^+\}} \quad (09.03)$$

dann eine $(\mathcal{S} \otimes \mathcal{I}) - \overline{\mathcal{B}}^+$ -messbare Funktion und eine ν -Dichte des Markovkerns $\mathbb{P}^{T|S}$ von $(\mathcal{S}, \mathcal{I})$ nach $(\mathcal{T}, \mathcal{I})$ mit $(s, A) \mapsto \mathbb{P}^{T|S=s}(A) := \int_A f_\theta^{T|S=s}(t) \nu(dt)$. Wegen der Suffizienz (für $\mathbb{P}_\theta^{(S, T)}$) der Statistik $\Pi_S : \mathcal{S} \times \mathcal{T} \rightarrow \mathcal{S}$ mit $(s, t) \mapsto \Pi_S(s, t) := s$ existieren nach Satz §09.29 Funktionen $h \in \overline{\mathcal{I} \otimes \mathcal{I}}^+$ und $g_\theta \in \overline{\mathcal{I}}^+$ für jedes $\theta \in \Theta$ mit $f_\theta^{(S, T)}(s, t) = g_\theta(s)h(s, t)$ für $[\mu \otimes \nu]$ -f.a. $(s, t) \in \mathcal{S} \times \mathcal{T}$. Für $h^S := \int_{\mathcal{T}} h(\bullet, t) \nu(dt) \in \overline{\mathcal{I}}^+$ mit $h^S \in \mathbb{R}$ μ -f.ü. gilt dann $f_\theta^S = g_\theta h^S$ μ -f.ü. und wegen (09.03) ist die Abbildung

$$f_\bullet^{T|S} : \mathcal{S} \times \mathcal{T} \rightarrow \overline{\mathbb{R}}^+ \text{ mit}$$

$$(s, t) \mapsto f_\bullet^{T|S=s}(t) = \frac{h(s, t)}{h^S(s)} \mathbb{1}_{\{h^S(s) \in \mathbb{R}_0^+\}} + f(t) \mathbb{1}_{\{h^S(s) \notin \mathbb{R}_0^+\}} \quad (09.04)$$

für eine beliebige ν -Dichte f dann eine ν -Dichte einer von θ unabhängigen regulären Festlegung $\mathbb{P}^{T|S}$ der bedingten Verteilung von T gegeben S . □

§09.33 **Korollar.** Sei (S, T) eine $(\mathcal{S} \times \mathcal{T}, \mathcal{S} \otimes \mathcal{I})$ -wertige Statistik auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$, S sei erschöpfend für \mathbb{P}_θ und $\mathbb{P}_\theta^{(S, T)}$ sei $\mu \otimes \nu$ -dominiert für $\mu \in \mathcal{M}_\sigma(\mathcal{S})$ und $\nu \in \mathcal{M}_\sigma(\mathcal{I})$, so gilt

- (i) Mit $f_\bullet^{T|S}$ gemäß (09.04) sind

$$(s, A) \mapsto \mathbb{P}^{T|S=s}(A) = \int_A f_\bullet^{T|S=s}(t) \nu(dt) \quad \text{sowie}$$

$$(s, D) \mapsto \mathbb{P}^{(S, T)|S=s}(D) = \int_{D_s} f_\bullet^{T|S=s}(t) \nu(dt) \quad \text{mit } D_s := \{t \in \mathcal{T} : (s, t) \in D\}.$$

Markovkerne und von θ unabhängige reguläre Festlegungen $\mathbb{P}_\theta^{T|S}$ und $\mathbb{P}_\theta^{(S,T)|S}$ der bedingten Verteilung von T bzw. (S, T) gegeben S .

(ii) Für $h \in \overline{\mathcal{S} \otimes \mathcal{T}^+}$, so dass $\mathbb{E}_\theta(h(S, T)) = \mathbb{E}_\theta^{(S,T)}(h)$ existiert für alle $\theta \in \Theta$, gilt

$$\mathbb{E}_\bullet(h(S, T)|S = s) = \int h_s \mathbb{f}_\bullet^{T|S=s} d\nu \mathbb{P}_\theta^S\text{-f.ü.} \quad \text{mit } h_s := h(s, \bullet)$$

$$\mathbb{E}_\theta^{(S,T)}(h \mathbb{1}_{B \times A}) = \int_{B \times A} h \mathbb{f}_\theta^{(S,T)} d[\mu \otimes \nu] = \int_B \left(\int_A h_s \mathbb{f}_\bullet^{T|S=s} d\nu \right) \mathbb{f}_\theta^S d\mu$$

für alle $B \in \mathcal{S}$, $A \in \mathcal{T}$ und $\theta \in \Theta$.

§09.34 **Beweis** von **Korollar** §09.33. Mit Hilfe der **Bemerkung** §09.32 folgt die Aussage direkt aus dem Neyman-Kriterium **Satz** §09.29. □

§09.35 **Erinnerung**. Für die nächste Aussage sei an die **Ungleichung von Jensen** (**Eigenschaft** A04.05 (v)) und ihre bedingte Version (**Eigenschaft** C03.04 (v)) erinnert. □

§09.36 **Rao-Blackwell Verbesserung**. Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment, $(\mathcal{E}, \mathcal{E}, \nu)$ ein Entscheidungsproblem mit konvexem Entscheidungsraum $\mathcal{E} \subseteq \mathbb{R}^k$ und im zweiten Argument konvexer Verlustfunktion $\nu(\theta, e)$. Ist S eine erschöpfende Statistik für \mathbb{P}_θ , so gilt für jede Entscheidungsregel δ mit $\|\delta\|_1, \nu(\theta, \delta) \in \mathcal{L}_1(\mathbb{P}_\theta)$ für alle $\theta \in \Theta$, und für jede (von θ unabhängige) Festlegung $\delta_\bullet = \mathbb{E}_\bullet(\delta|\sigma(S))$ die Risikoabschätzung

$$\forall \theta \in \Theta : \mathcal{R}_\nu(\theta, \delta_\bullet) \leq \mathcal{R}_\nu(\theta, \delta).$$

§09.37 **Beweis** von **Satz** §09.36. In der Vorlesung. □

§09.38 **Bemerkung**. Ist die Verlustfunktion strikt konvex im zweiten Argument sowie $\mathbb{P}_\theta(\delta = \delta_\bullet) < 1$, so ist δ_\bullet besser als δ . In dieser Situation gilt im **Satz** §09.36 also Gleichheit für die Risiken von δ_\bullet und δ genau dann, wenn $\delta_\bullet = \delta$ \mathbb{P}_θ -f.s. (vgl. Witting [1985], Satz 3.27, S.349). □

§09.39 **Korollar**. Es sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment und S eine erschöpfende Statistik für \mathbb{P}_θ . Zu jedem randomisierten Test φ gibt es einen randomisierten Test φ_\bullet , der nur von S abhängt und dieselben Irrtumswahrscheinlichkeiten erster und zweiter Art besitzt, nämlich jede (von θ unabhängige) Festlegung $\varphi_\bullet = \mathbb{E}_\bullet(\varphi|S)$.

§09.40 **Beweis** von **Korollar** §09.39. In der Vorlesung. □

§09.41 **Beispiel**. Für $\theta \in \mathbb{R}_0^+ =: \Theta$ bezeichne $\mathbb{P}_\theta := U_{[0,\theta]}$ eine Gleichverteilung auf dem Intervall $[0, \theta]$ mit Lebesguedichte $\mathbb{f}_\theta := \theta^{-1} \mathbb{1}_{[0,\theta]} \in \mathcal{B}^+$. Es sei $X = (X_i)_{i \in \llbracket n \rrbracket} \odot (U_{[0,\theta]}^n)_{\theta \in \mathbb{R}_0^+}$. Ein erwartungstreuer Schätzer des Erwartungswertes $\mathbb{E}_\theta(X_1) = \theta/2$ ist das arithmetische Mittel \bar{X} , so dass $\hat{\theta} := 2\bar{X}$ ein natürlicher Schätzer für θ ist. Sein Risiko bzgl. der quadratischen Verlustfunktion ist $\mathcal{R}_\nu(\theta, \hat{\theta}) = 4 \text{Var}_\theta(\bar{X}) = \frac{4\theta^2}{12n}$. Andererseits die Likelihood-Funktion bezüglich des Lebesguemaßes auf \mathbb{R}^n ist

$$L(\theta, x) = \prod_{i \in \llbracket n \rrbracket} \mathbb{f}_\theta(x_i) = \prod_{i \in \llbracket n \rrbracket} (\theta^{-1} \mathbb{1}_{[0,\theta]}(x_i)) = \theta^{-n} \mathbb{1}_{\mathbb{R}^+}(\min_{i \in \llbracket n \rrbracket} x_i) \mathbb{1}_{[0,\theta]}(\max_{i \in \llbracket n \rrbracket} x_i).$$

Das Faktorisierungskriterium von Neyman (**Satz** §09.29) anwendend ist $S(X) := X_{(n)} := \max\{X_i : i \in \llbracket n \rrbracket\}$ eine erschöpfende Statistik für $(U_{[0,\theta]}^n)_{\theta \in \mathbb{R}_0^+}$. Für jedes $\theta \in \mathbb{R}_0^+$ besitzt die Verteilung $\mathbb{P}_\theta^{X_{(n)}}$ die Lebesguedichte $\mathbb{f}_\theta^{X_{(n)}}(s) := ns^{n-1} \theta^{-n} \mathbb{1}_{[0,\theta]}(s)$. Wir betrachten

$$\hat{\theta}_\bullet := \mathbb{E}_\bullet(\hat{\theta}|\sigma(X_{(n)})) = \frac{2}{n} \sum_{i \in \llbracket n \rrbracket} \mathbb{E}_\bullet(X_i|\sigma(X_{(n)})). \tag{09.05}$$

Aus Symmetriegründen genügt es, $\mathbb{E}_\theta(X_1 | \sigma(X_{(n)}))$ zu bestimmen. Für $x \in [0, \theta]$ gilt:

$$\begin{aligned} \mathbb{E}_\theta(\mathbb{1}_{[0,x]}) &= \mathbb{P}_\theta([0, x]) = (x/\theta) \\ &= \frac{1}{n}(x/\theta)^n + \frac{n-1}{n} \left((x/\theta)^n + \frac{nx(\theta^{n-1} - x^{n-1})}{(n-1)\theta^n} \right) \\ &= \int_{[0,\theta]} \left(\frac{1}{n} \mathbb{1}_{[0,x]}(s) + \frac{n-1}{n} \frac{x \wedge s}{s} \right) n s^{n-1} \theta^{-n} ds \\ &= \int_{\mathbb{R}} \left(\frac{1}{n} \delta_s([0, x]) + \frac{n-1}{n} \mathbb{P}_s([0, x]) \right) \mathbb{f}_\theta^{X_{(n)}}(s) ds \\ &= \mathbb{E}_\theta^{X_{(n)}} \left(\frac{1}{n} \delta_\bullet([0, x]) + \frac{n-1}{n} \mathbb{P}_\bullet([0, x]) \right) = \mathbb{E}_\theta^{X_{(n)}}(\mathbb{P}([0, x] | X_{(n)})). \end{aligned}$$

Damit ist $\mathbb{P}(\bullet | X_{(n)})$ mit $(s, A) \mapsto \frac{1}{n} \delta_s(A) + \frac{n-1}{n} \mathbb{P}_s(A) =: \mathbb{P}(A | X_{(n)} = s)$ eine Festlegung der bedingten Verteilung von \mathbb{P}_θ gegeben $X_{(n)}$, die nicht von θ abhängt und regulär ist. Für die assoziierte reguläre Festlegung der bedingten Erwartung folgt $\mathbb{E}_\theta(X_1 | X_{(n)} = s) = \frac{1}{n}s + \frac{n-1}{2n}s = \frac{n+1}{2n}s$ und somit $\mathbb{E}_\theta(X_1 | \sigma(X_{(n)})) = \frac{n+1}{2n}X_{(n)}$, so dass wir mit (09.05) $\hat{\theta}_o = \frac{n+1}{n}X_{(n)}$ erhalten. Der Schätzer $\hat{\theta}_o$ ist erwartungstreu und sein quadratisches Risiko ist $\mathcal{R}_\nu(\theta, \hat{\theta}_o) = \frac{\theta^2}{n^2+2n}$. Für $n > 1$ ist der Schätzer $\hat{\theta}_o$ offensichtlich besser als $\hat{\theta}$, für $n \rightarrow \infty$ erhalten wir sogar die Ordnung $O(n^{-2})$ gegenüber $O(n^{-1})$. \square

§10 Exponentialfamilien

§10.01 **Definition.** Es sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein bzgl. eines σ -endlichen Maßes μ dominiertes statistisches Experiment. \mathbb{P}_θ wird (*k-parametrische*) *Exponentialfamilie* (in η und S) genannt, wenn $k \in \mathbb{N}$, $\eta : \Theta \rightarrow \mathbb{R}^k$, $C : \Theta \rightarrow \mathbb{R}_0^+$, $S \in \mathcal{X}^k$ und $h \in \mathcal{X}^+$ existieren mit

$$\frac{d\mathbb{P}_\theta}{d\mu}(x) = C(\theta)h(x) \exp(\langle \eta(\theta), S(x) \rangle), \quad \text{für } \mu\text{-f.a. } x \in \mathcal{X} \text{ und für alle } \theta \in \Theta.$$

Die Statistik S wird *natürlich erschöpfend* für \mathbb{P}_θ genannt. Sind die Koordinatenfunktionen von $\eta = (\eta_i)_{i \in [k]}$ linear unabhängige Funktionen und für die Koordinatenfunktionen von $S = (S_i)_{i \in [k]}$ gilt

$$\lambda_0 + \lambda_1 S_1 + \dots + \lambda_k S_k = 0 \quad \mathbb{P}_\theta\text{-f.s.} \quad \Rightarrow \quad \lambda_0 = \lambda_1 = \dots = \lambda_k = 0$$

d.h. $\mathbb{1}, S_1, \dots, S_k$ sind \mathbb{P}_θ -f.s. linear unabhängig. Dann wird die Exponentialfamilie *strikt k-parametrisch* genannt. \square

§10.02 **Bemerkung.**

- (i) $C(\theta) = (\int_{\mathcal{X}} h(x) \exp(\langle \eta(\theta), S(x) \rangle) \mu(dx))^{-1}$ ist gerade die Normierungskonstante. Die Wahrscheinlichkeitsmaße \mathbb{P}_θ , $\theta \in \Theta$, sind paarweise wie auch zu dem Maß $\nu := h\mu \in \mathcal{M}_\sigma(\mathcal{X})$ (da $\mu \in \mathcal{M}_\sigma(\mathcal{X})$) äquivalent. Eine \mathbb{P}_θ -Nullmenge ist somit Nullmenge für alle \mathbb{P}_θ , $\theta \in \Theta$, also für \mathbb{P}_θ , sowie für ν .
- (ii) Die Darstellung ist nicht eindeutig, mit einer invertierbaren Matrix $A \in \mathbb{R}^{k \times k}$ erhält man beispielsweise eine Exponentialfamilie in $\tilde{\eta}(\theta) = A\eta(\theta)$ und $\tilde{S}(x) = (A^t)^{-1}S(x)$. Außerdem kann die Funktion h in das dominierende Maß $\nu = h\mu$ absorbiert werden.
- (iii) Aus der Identifizierbarkeit des Parameters, d.h. $\mathbb{P}_\theta \neq \mathbb{P}_{\theta_0}$ für alle $\theta \neq \theta_0$, folgt die Injektivität von η . Andererseits impliziert die Injektivität von η bei einer strikt k -parametrischen Exponentialfamilie die Identifizierbarkeit des Parameters.

- (iv) Ist \mathbb{P} eine Exponentialfamilie in η und $S \in \mathcal{X}^k$, so ist $\mathbb{P}_\theta^S = (\mathbb{P}_\theta^S)_{\theta \in \Theta}$ eine Exponentialfamilie in η und der Identität $\text{id}_{\mathbb{R}^k}$.
- (v) Das Faktorisierungskriterium von Neyman (Satz §09.29) anwendend ist die natürliche erschöpfende Statistik S einer Exponentialfamilie \mathbb{P} in der Tat erschöpfend für \mathbb{P} im Sinne der Definition §09.08. \square

§10.03 **Definition.** Unter den Annahmen und Notationen der Definition §10.01 bezeichnet

$$\Theta_{\text{nat}} := \left\{ u \in \mathbb{R}^k : \int \exp(\langle u, S(x) \rangle) h(x) \mu(dx) \in \mathbb{R}_0^+ \right\}$$

den *natürlichen Parameterraum* einer Exponentialfamilie \mathbb{P} . Die entsprechend mit $u \in \Theta_{\text{nat}}$ reparametrisierte Familie $\mathbb{P}_{\Theta_{\text{nat}}} = (\mathbb{P}_u)_{u \in \Theta_{\text{nat}}}$ wird *natürliche Exponentialfamilie* in S genannt. \square

§10.04 **Beispiel.**

- (a) Die Normalverteilungsfamilie $N_{\mathbb{R} \times \mathbb{R}_0^+}$ ist eine zweiparametrische Exponentialfamilie in $\eta(\mu, \sigma) = (\mu/\sigma^2, 1/(2\sigma^2))$ und $S(x) = (x, -x^2)$ bzgl. des Lebesguemaßes λ als dominierendes Maß. Jedes u der Form $u = (\mu/\sigma^2, 1/(2\sigma^2))$ ist ein natürlicher Parameter, und der natürliche Parameterraum ist gegeben durch $\Theta_{\text{nat}} = \mathbb{R} \times \mathbb{R}_0^+$. Ist entweder $\sigma \in \mathbb{R}_0^+$ oder $\mu \in \mathbb{R}$ bekannt so liegt eine einparametrische Exponentialfamilie in $\eta(\mu) = \mu/\sigma^2$ bzw. $\eta(\sigma) = 1/(2\sigma^2)$ und $S(x) = x$ bzw. $S(x) = -x^2$ vor.
- (b) Die Binomialverteilungsfamilie $(\text{Bin}_{(n,p)})_{p \in (0,1)}$ bildet eine Exponentialfamilie in $\eta(p) = \log(p/(1-p))$ (Logitfunktion vgl. Bemerkung §01.15) und $S(x) = x$ bezüglich des Zählmaßes $\zeta_{[0,n]}$ auf $[0, n]$. Der natürliche Parameterraum ist \mathbb{R} , insbesondere liegt für den Parameterbereich $[0, 1]$ keine Exponentialfamilie vor. \square

§10.05 **Lemma.** Der natürlichen Parameterraum Θ_{nat} einer Exponentialfamilie ist konvex.

§10.06 **Beweis** von Lemma §10.05. Dies folgt aus der Höldersche Ungleichung (B03.08 (i)). \square

§10.07 **Lemma.** Bildet \mathbb{P} eine k -parametrische Exponentialfamilie in η und S , so bildet auch die Familie der Produktmaße \mathbb{P}^n eine k -parametrische Exponentialfamilie in η und $S_n \in \mathcal{X}^n$ mit $S_n(x) = \sum_{i \in [n]} S(x_i)$ und der Darstellung

$$\frac{d\mathbb{P}_\theta^n}{d\mu^n}(x) = C(\theta)^n \left(\prod_{i \in [n]} h(x_i) \right) \exp \left(\langle \eta(\theta), \sum_{i \in [n]} S(x_i) \rangle \right), \quad x \in \mathcal{X}^n, \theta \in \Theta.$$

§10.08 **Beweis** von Lemma §10.07. Dies folgt aus der Produktformel $\frac{d\mathbb{P}_\theta^n}{d\mu^n}(x) = \prod_{i \in [n]} \frac{d\mathbb{P}_\theta}{d\mu}(x_i)$. \square

§10.09 **Beispiel.**

- (a) (Fortsetzung von Beispiel §01.06 (a)) Betrachte $(\mathbb{R}^n, \mathcal{B}^n, N_{\mathbb{R} \times \mathbb{R}_0^+}^n)$, also das *normale Lokations-Skalen-Modell*. Die natürliche erschöpfende Statistik ist $\left(\sum_{i \in [n]} x_i, -\sum_{i \in [n]} x_i^2 \right)$. Durch Transformation sind damit auch (\bar{x}, \bar{x}^2) und (\bar{x}, S^2) mit $S^2 = \frac{1}{n-1} \sum_{i \in [n]} (x_i - \bar{x})^2$ erschöpfende Statistiken.
- (b) Sei $(B_p^n)_{p \in (0,1)}$ die Verteilungsfamilie einer Bernoullikette, dann ist die Anzahl der Erfolge $\sum_{i \in [n]} x_i$ erschöpfend. \square

§10.10 **Satz.** Sei $\mathbb{P}_{\Theta_{\text{nat}}}$ eine Exponentialfamilie mit natürlichem Parameterraum $\Theta_{\text{nat}} \subseteq \mathbb{R}^k$ und für alle $\theta \in \Theta_{\text{nat}}$ der Darstellung

$$\frac{d\mathbb{P}_\theta}{d\mu} = C(\theta) h \exp(\langle \theta, S \rangle) = h \exp(\langle \theta, S \rangle - A(\theta)) \quad \mu\text{-f.ü.}, \quad (10.01)$$

mit $A(\theta) = \log \left(\int h \exp(\langle \theta, S \rangle) d\mu \right)$. Ist θ_o ein innerer Punkt von Θ_{nat} , so ist die erzeugende Funktion $\psi_{\theta_o}(s) = \mathbb{E}_{\theta_o}[\exp(\langle S, s \rangle)]$ in einer Umgebung der Null wohldefiniert und beliebig oft differenzierbar. Es gilt weiterhin $\psi_{\theta_o}(s) = \exp(A(\theta_o + s) - A(\theta_o))$ für alle s mit $\theta_o + s \in \Theta_{\text{nat}}$. Für $i, j \in \llbracket k \rrbracket$ folgt außerdem $\mathbb{E}_{\theta_o}(S_i) = (\dot{A}_{\theta_o})_i = \frac{\partial A}{\partial \theta_i}(\theta_o)$ und $\text{Cov}_{\theta_o}(S_i, S_j) = (\ddot{A}_{\theta_o})_{ij} = \frac{\partial^2 A}{\partial \theta_i \partial \theta_j}(\theta_o)$.

§10.11 **Beweis von Satz §10.10.** Übungsaufgabe. □

§10.12 **Vorbetrachtung.** Im Folgenden sei $\mathbb{P}_{\Theta_{\text{nat}}} = (\mathbb{P}_{\theta, \tau})_{(\theta, \tau) \in \Theta_{\text{nat}}}$ mit $\Theta_{\text{nat}} \subseteq \mathbb{R}^{k+m}$ eine natürliche Exponentialfamilie in $(S, T) \in \mathcal{X}^{k+m}$ mit der Darstellung

$$\frac{d\mathbb{P}_{\theta, \tau}}{d\nu}(x) = C(\theta, \tau) \exp(\langle \theta, S(x) \rangle + \langle \tau, T(x) \rangle) \quad \nu\text{-f.ü.} \tag{10.02}$$

Für jedes $(\theta_o, \tau_o) \in \Theta_{\text{nat}}$ ist $\mathbb{P}_o := \mathbb{P}_{\theta_o, \tau_o}$ ein privilegiertes dominierendes Wahrscheinlichkeitsmaß für $\mathbb{P}_{\Theta_{\text{nat}}}$ (vgl. **Bemerkung §10.02 (i)**) und für jedes $(\theta, \tau) \in \Theta_{\text{nat}}$ ist

$$\frac{d\mathbb{P}_{\theta, \tau}}{d\mathbb{P}_o}(x) = \frac{C(\theta, \tau)}{C(\theta_o, \tau_o)} \exp(\langle \theta - \theta_o, S(x) \rangle) \exp(\langle \tau - \tau_o, T(x) \rangle) \quad \mathbb{P}_o\text{-f.ü.}$$

eine \mathbb{P}_o -Dichte von $\mathbb{P}_{\theta, \tau}$. Setzen wir $c_{\theta, \tau} := \frac{C(\theta, \tau)}{C(\theta_o, \tau_o)} \in \mathbb{R}_{>0}^+$, $f_{\theta} := \exp(\langle \theta - \theta_o, \bullet \rangle) \in (\mathcal{B}^m)^+$ sowie $g_{\tau} := \exp(\langle \tau - \tau_o, \bullet \rangle) \in (\mathcal{B}^k)^+$ so besitzt $\mathbb{P}_{\theta, \tau}$ eine \mathbb{P}_o -Dichte der Form (09.01). Weiterhin existiert nach **Eigenschaft C02.09 (ii)** eine reguläre Festlegung $\mathbb{P}_o^{S|T}$ der bedingten Verteilung von S gegeben T bzgl. \mathbb{P}_o . Somit können wir **Lemma §09.27** anwenden, wobei wir für jedes θ , $\Theta_{\text{nat}}^{\theta} := \{\tau : (\theta, \tau) \in \Theta_{\text{nat}}\}$ sowie $\Theta_{\text{nat}}^{\bullet} := \{\theta : \Theta_{\text{nat}}^{\theta} \neq \emptyset\}$ definieren. □

§10.13 **Korollar.** Gegeben sei eine natürliche Exponentialfamilie $\mathbb{P}_{\Theta_{\text{nat}}}$ in (θ, τ) und (S, T) mit der Darstellung (10.02). \mathbb{P}_o sei eine spezielle Verteilung $\mathbb{P}_{\theta_o, \tau_o}$, $(\theta_o, \tau_o) \in \Theta_{\text{nat}}$. Bei gegebener regulärer Festlegung $\mathbb{P}_o^{S|T}$ der bedingten Verteilung von S gegeben T bzgl. \mathbb{P}_o gilt dann:

(i) Für jedes $\theta \in \Theta_{\text{nat}}^{\bullet}$ bilden die von T induzierten Verteilungen $\mathbb{P}_{\{\theta\} \times \Theta_{\text{nat}}^{\theta}}^T := (\mathbb{P}_{\theta, \tau}^T)_{\tau \in \Theta_{\text{nat}}^{\theta}}$ eine Exponentialfamilie in τ und der Identität $\text{id}_{\mathbb{R}^m}$, das heißt für alle $\tau \in \Theta_{\text{nat}}^{\theta}$ gilt

$$\frac{d\mathbb{P}_{\theta, \tau}^T}{d\mathbb{P}_o^T}(t) = C(\theta, \tau) \exp(\langle \tau, t \rangle) f_{\theta}^T(t) \quad \mathbb{P}_o^T\text{-f.a. } t \in \mathbb{R}^m$$

mit $f_{\theta}^T(t) := \frac{1}{C(\theta_o, \tau_o)} \exp(\langle -\tau_o, t \rangle) \int \exp(\langle \theta - \theta_o, s \rangle) \mathbb{P}_o^{S|T=t}(ds)$.

(ii) Für jedes $\theta \in \Theta_{\text{nat}}^{\bullet}$ gibt es eine von τ unabhängige reguläre Festlegung $\mathbb{P}_{\theta, \bullet}^{S|T}$ der bedingten Verteilung $\mathbb{P}_{\theta, \tau}^{S|T}$, die für \mathbb{P}_o^T -f.a. $t \in \mathbb{R}^m$ eine Exponentialfamilie $(\mathbb{P}_{\theta, \bullet}^{S|T=t})_{\theta \in \Theta_{\text{nat}}^{\bullet}}$ in θ und $\text{id}_{\mathbb{R}^k}$ bilden, das heißt für alle $\theta \in \Theta_{\text{nat}}^{\bullet}$ gilt

$$\frac{d\mathbb{P}_{\theta, \bullet}^{S|T=t}}{d\mathbb{P}_o^{S|T=t}}(s) = \frac{1}{f_{\theta}^T(t)} \exp(\langle \theta, s \rangle) \exp(\langle -\theta_o, s \rangle) \quad \mathbb{P}_o^{S|T=t}\text{-f.a. } s \in \mathbb{R}^k.$$

§10.14 **Beweis von Korollar §10.13.** Mit **Bemerkung §10.12** folgt die Behauptung direkt aus **Lemma §09.27**. □

§11 Vollständige Statistik

§11.01 **Definition.** Eine (S, \mathcal{S}) -wertige Statistik S auf $(\mathcal{X}, \mathcal{A}, \mathbb{P}_o)$ heißt *vollständig*, falls

$$\forall f \in \mathcal{S} : (\forall \theta \in \Theta : \mathbb{E}_{\theta}^S(f) = 0 \Rightarrow f = 0 \quad \mathbb{P}_o^S\text{-f.ü.}).$$

□

§11.02 **Bemerkung.** Eine Statistik V auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ wird *unwesentlich* (ancillary) genannt, wenn ihre Verteilung $\mathbb{P}_\theta^V := \mathbb{P}_\theta^V$ nicht vom Parameter θ abhängt. Sie heißt *unwesentlich erster Ordnung*, falls $\mathbb{E}_\theta[V] := \mathbb{E}_\theta[V]$ unabhängig von θ ist. Falls jede Statistik der Form $V = f(S)$, die unwesentlich erster Ordnung ist, auch \mathbb{P}_θ -f.ü. konstant ist, so ist keine redundante Information mehr in S enthalten, und S ist vollständig. \square

§11.03 **Lemma von Basu.** Es seien S und V Statistiken auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$. Ist S erschöpfend und vollständig sowie V unwesentlich (ancillary), d.h. \mathbb{P}_θ^V ist unabhängig von $\theta \in \Theta$, so sind S und V unabhängig.

§11.04 **Beweis** von Satz §11.03. In der Vorlesung. \square

§11.05 **Satz von Koopman.** Es sei $\mathbb{P}_{\Theta_{\text{nat}}}$ eine (strikt) k -parametrische Exponentialfamilie in S mit natürlichem Parameter $\Theta_{\text{nat}} \subseteq \mathbb{R}^k$. Besitzt Θ_{nat} ein nichtleeres Inneres $\text{inn}(\Theta_{\text{nat}})$, so ist S erschöpfend und vollständig.

§11.06 **Beweis** von Satz §11.05. In der Vorlesung. \square

§11.07 **Bemerkung.** Der natürliche Parameterraum Θ_{nat} einer (strikt) k -parametrischen Exponentialfamilie ist konvex und enthält ein nicht entartetes k -dimensionales Rechteck. \square

§11.08 **Korollar.** Gegeben sei eine natürliche Exponentialfamilie $\mathbb{P}_{\Theta_{\text{nat}}}$ in (θ, τ) und (S, T) mit der Darstellung (10.02). Hängt die Verteilung einer weiteren Statistik $V(S, T)$ auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\Theta_{\text{nat}}})$ nicht von τ ab, so sind die Statistiken V und T unabhängig unter jedem $\mathbb{P}_{\theta, \tau}$ für alle $\tau \in \Theta_{\text{nat}}^\theta$ mit $\text{inn}(\Theta_{\text{nat}}^\theta) \neq \emptyset$ und $\theta \in \Theta_{\text{nat}}^\theta$.

§11.09 **Beweis** von Korollar §11.08. In der Vorlesung. \square

§11.10 **Beispiel.**

(a) Besitzt in einem *gewöhnlichen normalen linearen Modell* $Y \odot N_{X\mathbb{R}^k \times \mathbb{R}_0^+}$ (siehe §03) die Designmatrix X einen vollen Spaltenrang ($\text{rg}(X) = k$), so liegt eine (strikt) $(k+1)$ -parametrische Exponentialfamilie in $\eta(\beta, \sigma) = \sigma^{-2}(\beta, 1/2)^t \in \mathbb{R}^k \times \mathbb{R}_0^+$ und $(S(Y), T(Y)) = (X^t Y, -\|Y\|^2)^t \in \mathbb{R}^{k+1}$ vor. Der natürliche Parameterraum $\Theta_{\text{nat}} = \mathbb{R}^k \times \mathbb{R}_0^+$ besitzt ein nichtleeres Inneres in \mathbb{R}^{k+1} , so dass (S, T) erschöpfend und vollständig ist. Mittels einer bijektiven Transformation ergibt sich, dass für den gewöhnlichen Kleinste-Quadrate-Schätzer (gKQS) $\hat{\beta} = (X^t X)^{-1} X^t Y$ (Korollar §02.09) und $\hat{\sigma}^2 = (n-k)^{-1} \|Y - X\hat{\beta}\|^2$ (Satz §03.02) auch die Statistik $(\hat{\beta}, \|Y\|^2) = (\hat{\beta}, \|\Pi_{X\mathbb{R}^k} Y\|^2 + (n-k)\hat{\sigma}^2)$ erschöpfend und vollständig ist. Da weiterhin gilt $\Pi_{X\mathbb{R}^k} Y = X\hat{\beta}$ (Lemma §02.07), ist auch $(\hat{\beta}, \hat{\sigma}^2)$ erschöpfend und vollständig. Insbesondere hängt die Verteilung von $\hat{\sigma}^2$ nicht von β ab, so dass nach Korollar §11.08 $\hat{\beta}$ und $\hat{\sigma}^2$ unabhängig für alle $(\beta, \sigma) \in \mathbb{R}^k \times \mathbb{R}_0^+$ sind.

(b) Sei $X = (X_i)_{i \in \llbracket n \rrbracket} \odot (U_{[0, \theta]}^n)_{\theta \in \mathbb{R}_0^+}$ (siehe Beispiel §09.41). Dann folgt aus der Form $L(\theta, x) = \theta^{-n} \mathbf{1}_{\mathbb{R}^+}(x_{(1)}) \mathbf{1}_{[0, \theta]}(x_{(n)})$ mit $x_{(1)} := \min\{x_i : i \in \llbracket n \rrbracket\}$ und $x_{(n)} := \max\{x_i : i \in \llbracket n \rrbracket\}$ der Likelihood-Funktion, dass das Maximum $S(X) := X_{(n)}$ der Beobachtungen eine erschöpfende Statistik ist. Da \mathbb{P}_θ^S die Lebesgue-dichte $f_\theta^S(s) := n s^{n-1} \theta^{-n} \mathbf{1}_{[0, \theta]}(s)$ besitzt, folgt aus

$$\mathbb{E}_\theta^S(f) = \int_{[0, \theta]} f(s) n \theta^{-n} s^{n-1} dt = 0,$$

für alle $\theta \in \mathbb{R}_0^+$, dass $f = 0$ Lebesgue-f.ü. gelten muss, woraus die Vollständigkeit für $S(X) = X_{(n)}$ folgt. \square

§12 Erwartungstreue Schätzer

§12.01 **Satz von Lehmann-Scheffé.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein statistisches Experiment, $\hat{\gamma}$ ein erwartungstreuer Schätzer des identifizierbaren Parameters $\gamma : \Theta \rightarrow \mathbb{R}$ und S eine erschöpfende und vollständige Statistik für \mathbb{P} . Dann ist $\hat{\gamma}_o = \mathbb{E}_\bullet(\hat{\gamma} | \sigma(S))$ der eindeutig bestimmte Schätzer, der in der Klasse aller erwartungstreuen Schätzer gleichmäßig die kleinste Varianz besitzt (KVS für Kleinste-Varianz-Schätzer oder UMVU für uniformly minimum variance unbiased oder BUE für best unbiased estimator). Insbesondere gilt damit:

- (i) **(Existenz)** Es gibt einen KVS der Form $\hat{\gamma}_o(x) = g(S(x))$ für alle $x \in \mathcal{X}$.
- (ii) **(Eindeutigkeit)** Ist $\tilde{\gamma}$ ein KVS, dann gilt $\mathbb{P}(\tilde{\gamma} = \hat{\gamma}_o) = 1$ für alle $\theta \in \Theta$.
- (iii) Ist $\tilde{\gamma} = h(S)$ erwartungstreu für γ , dann gilt $\mathbb{P}(\tilde{\gamma} = \hat{\gamma}_o) = 1$ für alle $\theta \in \Theta$.

§12.02 **Beweis** von Satz §12.01. In der Vorlesung. □

§12.03 **Bemerkung.** Aus dem Satz §09.36 (Rao-Blackwell Verbesserung) folgt die Aussage des Satzes von Lehmann-Scheffé sogar für das Risiko bzgl. einer beliebigen im zweiten Argument strikt konvexen Verlustfunktion. □

§12.04 **Beispiel** (Beispiel §11.10 fortgesetzt).

- (a) Sei $Y \odot N_{\mathbb{R}^k \times \mathbb{R}_0^+}$ mit $\text{rg}(X) = k$. Da $\hat{\beta}$ und $\hat{\sigma}$ erwartungstreue Schätzer für β und σ^2 sind (Satz §02.15), die insbesondere erschöpfend und vollständig sind, besitzen beide Schätzer jeweils minimale Varianz in der Klasse aller erwartungstreuen Schätzer von β und σ^2 . Für diese Aussage ist die Normalverteilungsannahme essentiell.
- (b) $X = (X_i)_{i \in [1, n]} \odot (U_{[0, \theta]}^n)_{\theta \in \mathbb{R}_0^+}$. Da $X_{(n)}$ eine erschöpfende und vollständige Statistik mit $\mathbb{E}_\theta(X_{(n)}) = \frac{n}{n+1}\theta$ für alle $\theta \in \Theta$ ist, besitzt der erwartungstreue Schätzer $\hat{\theta}_o = \frac{n+1}{n}X_{(n)}$ minimale Varianz in der Klasse aller erwartungstreuen Schätzer von θ . □

§12.05 **Berechnung des Kleinste-Varianz-Schätzer.** Sei S vollständig und erschöpfend für \mathbb{P} . Möchte man den Kleinste-Varianz-Schätzer für γ bestimmen so gibt es zwei Möglichkeiten, ihn zu berechnen (die Existenz vorausgesetzt):

- (a) **(Direkte Methode, geeigneter für diskrete Verteilungen)** Man sucht einen erwartungstreuen Schätzer der Form $\hat{\gamma}_o = h(S)$ für γ , dieser ist dann der Kleinste-Varianz-Schätzer. Dies führt zu folgendem Gleichungssystem für die unbekannte Funktion h

$$\forall \theta \in \Theta : \quad \gamma(\theta) = \mathbb{E}_\theta(\hat{\gamma}_o) = \mathbb{E}_\theta^S(h) = \int h(s) \mathbb{P}_\theta^S(ds).$$

Als Übungsaufgabe benutze diese Methode im Fall von **Beispiel §10.09(b)** für den abgeleiteten Parameter $\gamma(\pi) = \pi(1 - \pi)$.

- (b) **(Benutze Rao-Blackwell Verbesserung)** Für einen beliebigen erwartungstreuen Schätzer $\tilde{\gamma}$ ist die Rao-Blackwell Verbesserung $\hat{\gamma}_o = h(S) = \mathbb{E}_\bullet(\tilde{\gamma} | \sigma(S))$ dann der Kleinste-Varianz-Schätzer. Die Berechnung kann entweder direkt mit bedingten Dichten durchgeführt werden (häufig aufwendig), oder man nutzt die Charakterisierung der bedingten Erwartung

$$\forall \theta \in \Theta : \forall A \in \sigma(S) : \quad \mathbb{E}_\theta(\mathbb{1}_A h(S)) = \mathbb{E}_\theta(\mathbb{1}_A \tilde{\gamma})$$

was erneut zu einem Gleichungssystem für h führt. Wir haben dieses Verfahren in **Beispiel §09.41** benutzt. □

§12.06 **Kritik.** Kleinste-Varianz-Schätzer werden häufig kritisiert, da

- (a) Eine Einschränkung auf erwartungstreue Schätzer zu viele Schätzer ausschließt. Aus dem **Satz von Lehmann-Scheffé** §12.01 wird deutlich, dass nur ein von einer erschöpfenden und vollständigen Statistik abhängenden Schätzer existiert.
- (b) Die Einschränkung auf erwartungstreue Schätzer schließt häufig interessante Schätzer mit geringerem Risiko aus, da eventuell ein Schätzer mit einer kleinen Verfälschung eine deutlich geringe Varianz besitzen kann (siehe nachfolgendes **Beispiel** §12.07(a)).
- (c) Es gibt Situationen, in denen erwartungstreue Schätzer und Kleinste-Varianz-Schätzer völlig unsinnig sind (siehe nachfolgendes **Beispiel** §12.07(b)).
- (d) Es gibt Situationen, in denen erwartungstreue Schätzer und Kleinste-Varianz-Schätzer nicht existieren (Übungsaufgabe). \square

§12.07 **Beispiel.**

- (a) (*Fortsetzung Beispiel §10.09 (a)*). Im normalen Lokation-Skalen-Modell $(\mathbb{R}^n, \mathcal{B}^n, N_{\mathbb{R} \times \mathbb{R}_0^+}^n)$ betrachte für jedes $c \in \mathbb{R}_0^+$ den Schätzer $\hat{\sigma}_c^2 := c \sum_{i=1}^n (X_i - \bar{X})^2$ für den Parameter $\sigma^2 \in \mathbb{R}_0^+$, wobei für

$$c = \begin{cases} \frac{1}{n-1} \\ \frac{1}{n} \end{cases} \quad \text{ist} \quad \hat{\sigma}_c^2 \quad \begin{cases} \text{KVS;} \\ \text{MLS (im nächsten Kapitel).} \end{cases}$$

Dann gilt $\mathbb{E}(\hat{\sigma}_c^2) = c(n-1)\sigma^2$ und $\text{Var}(\hat{\sigma}_c^2) = c^2(2n-2)\sigma^4$, so dass bzgl. der quadratischen Verlustfunktion für das Risiko von $\hat{\sigma}_c^2$ gilt

$$\begin{aligned} \mathcal{R}_\nu(\sigma^2, \hat{\sigma}_c^2) &= \text{Var}(\hat{\sigma}_c^2) + (\mathbb{E}(\hat{\sigma}_c^2) - \sigma^2)^2 \\ &= ((n^2 - 1)c^2 - 2(n-1)c + 1)\sigma^4 = h(c)\sigma^4. \end{aligned}$$

Bestimmung der Minimalstelle von h liefert nun $h'(c_o) = 2c_o(n^2 - 1) - 2(n-1) = 0$ genau dann, wenn $c_o = (n+1)^{-1}$. Damit minimiert der Schätzer $(n+1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ das Risiko in der Familie der Schätzer $\{c \sum_{i=1}^n (X_i - \bar{X})^2, c \in \mathbb{R}_0^+\}$ und nicht der Kleinste-Varianz-Schätzer. Außerdem sieht man, dass der MLS in diesem Beispiel besser ist als der Kleinste-Varianz-Schätzer ($c = n^{-1}$ ist näher am Scheitelpunkt als $c = (n-1)^{-1}$).

- (b) Sei $X \odot \text{Poi}_{\mathbb{R}_0^+} := (\text{Poi}_\lambda)_{\lambda \in \mathbb{R}_0^+}$. Gesucht ist der Kleinste-Varianz-Schätzer für den abgeleiteten Parameter $\gamma : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ mit $\lambda \mapsto \gamma(\lambda) := \exp(-2\lambda)$. Es ist bekannt, dass $S(x) = x$ eine vollständige und erschöpfende Statistik für $\text{Poi}_{\mathbb{R}_0^+}$ ist. Daher besitzt der Kleinste-Varianz-Schätzer die Form $\hat{\gamma}_o = h(X)$ für eine Funktion $h : \mathbb{N}_0 \rightarrow \mathbb{R}$, welche Lösung des Gleichungssystems

$$\sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!} e^{-\lambda} = \mathbb{E}_\lambda(\hat{\gamma}_o) = \gamma(\lambda) = \exp(-2\lambda)$$

ist und somit auch $\sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!} = \exp(-\lambda) = \sum_{k=0}^{\infty} (-1)^k \frac{\lambda^k}{k!}$ erfüllt, d.h. aber $\hat{\gamma}_o = (-1)^X$ ist Kleinste-Varianz-Schätzer für $\gamma(\lambda) = \exp(-2\lambda)$, was unsinnig ist. \square

§13 Informationsungleichungen

- §13.01 **Erinnerung.** Seien $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathcal{X})$ Wahrscheinlichkeitsmaße, wobei nicht notwendig $\mathbb{P}_0 \ll \mathbb{P}_1$ gilt. Jede positive numerische Funktion $L \in \overline{\mathcal{X}^+}$ mit $\mathbb{P}_1(L = \infty) = 0$ und $\mathbb{P}_0 = L\mathbb{P}_1 + \mathbb{1}_{\{L=\infty\}}\mathbb{P}_0$ heißt *Dichtequotient (DQ)* von \mathbb{P}_0 bezüglich \mathbb{P}_1 (vgl. **Definition** B04.07 und **Eigenschaft** B04.08). Im Fall $\mathbb{P}_0 \ll \mathbb{P}_1$ ist $d\mathbb{P}_0/d\mathbb{P}_1 = L\mathbb{1}_{\{L \in \mathbb{R}^+\}}$ eine Festlegung der Radon-Nikodym Dichte von \mathbb{P}_0 bzgl. \mathbb{P}_1 (vgl. **Eigenschaft** B04.02 (ii)). \square

§13.02 **Chapman-Robins-Ungleichung.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment, $\hat{\gamma}$ ein erwartungstreuer Schätzer des identifizierbaren Parameters $\gamma : \Theta \rightarrow \mathbb{R}$ und $\theta_o \in \Theta$. Für jedes $\theta \in \Theta$ mit $\mathbb{P}_\theta \ll \mathbb{P}_{\theta_o}$ bezeichne $L_{\theta_o}(\theta) = d\mathbb{P}_\theta/d\mathbb{P}_{\theta_o}$ einen Dichtequotienten von \mathbb{P}_θ bzgl. \mathbb{P}_{θ_o} , also eine \mathbb{P}_{θ_o} -Dichte von \mathbb{P}_θ . Erfüllt dieser zusätzlich $\mathbb{E}_{\theta_o}(L_{\theta_o}^2(\theta)) = \int_{\mathcal{X}} L_{\theta_o}^2(\theta, x) \mathbb{P}_{\theta_o}(dx) \in \mathbb{R}^+$ (kurz $L_{\theta_o}(\theta) \in \mathcal{L}_2(\mathbb{P}_{\theta_o})$), dann gilt

$$\text{Var}_{\theta_o}(\hat{\gamma}) = \mathbb{E}_{\theta_o}(|\hat{\gamma} - \gamma(\theta_o)|^2) \geq \frac{|\gamma(\theta) - \gamma(\theta_o)|^2}{\text{Var}_{\theta_o}(L_{\theta_o}(\theta))}.$$

§13.03 **Beweis von Lemma §13.02.** In der Vorlesung. □

§13.04 **Beispiel.** Betrachte auf $(\mathbb{R}^+, \mathcal{B}^+)$ die Familie $\text{Exp}_{\mathbb{R}^+} := (\text{Exp}_\theta)_{\theta \in \mathbb{R}^+}$ von Exponentialverteilungen. Für $\theta_o, \theta \in \mathbb{R}^+$ ist dann $L_{\theta_o}(\theta, x) = (\theta/\theta_o) \exp(-(\theta - \theta_o)x) \mathbb{1}_{\mathbb{R}^+}(x)$ ein Dichtequotient. Für $\theta \in (\theta_o/2, \infty)$ gilt somit $L_{\theta_o}(\theta) \in \mathcal{L}_2(\mathbb{P}_{\theta_o})$ und $\text{Var}_{\theta_o}(L_{\theta_o}(\theta)) = \frac{(\theta - \theta_o)^2}{\theta_o(2\theta - \theta_o)}$. Sei $\hat{\theta}$ nun ein beliebiger erwartungstreuer Schätzer für θ (d.h. $\gamma = \text{id}_{\mathbb{R}^+}$). Aus der **Chapman-Robins-Ungleichung** §13.02 folgt dann $\text{Var}_{\theta_o}(\hat{\theta}) \geq \sup_{\theta \in (\theta_o/2, \infty)} \theta_o(2\theta - \theta_o) = \infty$. Sofern also beliebig große Werte θ zugelassen sind, existiert kein erwartungstreuer Schätzer von θ mit endlicher Varianz. Betrachten wir dagegen den identifizierbaren Parameter $\theta \mapsto \gamma(\theta) := \theta^{-1}$, so schließen wir mit Hilfe der **Chapman-Robins-Ungleichung** §13.02, dass $\text{Var}_{\theta_o}(\hat{\gamma}) \geq \sup_{\theta \in (\theta_o/2, \infty)} \frac{(2\theta - \theta_o)}{\theta^2 \theta_o} = \theta_o^{-2}$. Der Schätzer $\hat{\gamma}_o := X$ ist erwartungstreu für $\gamma(\theta) = \theta^{-1}$ mit Varianz $\text{Var}_\theta(\hat{\gamma}_o) = \theta^{-2}$ und erreicht somit diese Schranke. □

§13.05 **Erinnerung.** Seien $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathcal{X})$ Wahrscheinlichkeitsmaße dominiert bezüglich $\mu \in \mathcal{M}_\sigma(\mathcal{X})$ (z.Bsp. $\mu = \mathbb{P}_0 + \mathbb{P}_1$) mit Likelihood-Funktion $L_0 := \frac{d\mathbb{P}_0}{d\mu}, L_1 := \frac{d\mathbb{P}_1}{d\mu} \in \mathcal{X}^+$ (vgl. **Definition** §08.01). Dann wird $H(\mathbb{P}_0, \mathbb{P}_1) = \|L_0^{1/2} - L_1^{1/2}\|_{\mathcal{L}_2(\mu)}$ Hellinger Abstand zwischen \mathbb{P}_0 und \mathbb{P}_1 genannt, wobei $\|L_j^{1/2}\|_{\mathcal{L}_2(\mu)}^2 = \mu(L_j) = 1 \in \mathbb{R}^+$, also $L_j^{1/2} \in \mathcal{L}_2(\mu), j \in \{0, 1\}$. □

§13.06 **Definition.** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein μ -dominiertes statistisches Modell mit $\Theta \subseteq \mathbb{R}^k$ und Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu$. Weiterhin bezeichne $\ell(\theta) \in \overline{\mathcal{X}}$ mit $x \mapsto \ell(\theta, x) = \log(L(\theta, x))$ (mit der Konvention $\log(0) := -\infty$) die **Log-Likelihood-Funktion**. Die Familie \mathbb{P}_θ wird **Hellinger-differenzierbar** mit **Ableitung** $\dot{\ell}_\theta$ in $\theta_o \in \text{inn}(\Theta) \subseteq \mathbb{R}^k$ genannt, falls $\dot{\ell}_\theta \in \mathcal{L}_2^k(\mathbb{P}_{\theta_o})$ und somit $\dot{\ell}_\theta L^{1/2}(\theta_o) \in \mathcal{L}_2^k(\mu)$ existiert, derart dass

$$\begin{aligned} \lim_{\theta \rightarrow \theta_o} \int_{\mathcal{X}} \left| \frac{L^{1/2}(\theta, x) - L^{1/2}(\theta_o, x) - \frac{1}{2} \langle \dot{\ell}_\theta(x), \theta - \theta_o \rangle L^{1/2}(\theta_o, x)}{\|\theta - \theta_o\|} \right|^2 \mu(dx) \\ = \lim_{h \rightarrow 0} \frac{\|L^{1/2}(\theta_o + h) - L^{1/2}(\theta_o) - \frac{1}{2} \langle \dot{\ell}_\theta, h \rangle L^{1/2}(\theta_o)\|_{\mathcal{L}_2(\mu)}^2}{\|h\|^2} = 0. \end{aligned}$$

Die Abbildung $\dot{\ell}_\theta \in \mathcal{X}^k$ heißt **Score-Funktion** und (beachte $\dot{\ell}_\theta \in \mathcal{L}_2^k(\mathbb{P}_\theta)$)

$$\mathcal{I}_\theta := \mathbb{E}_\theta(\dot{\ell}_\theta \dot{\ell}_\theta^t) = \int_{\mathcal{X}} (\dot{\ell}_\theta(x) \dot{\ell}_\theta^t(x)) \mathbb{P}_\theta(dx)$$

wird **Fisher-Informationsmatrix** in $\theta_o \in \text{inn}(\Theta)$ genannt. □

§13.07 **Bemerkung.**

(i) Sofern alle folgenden Ausdrücke klassisch differenzierbar sind, so gilt

$$\nabla_\theta L^{1/2}(\theta, x) = \frac{\nabla_\theta L(\theta, x)}{2L^{1/2}(\theta, x)} = \frac{1}{2} L^{1/2}(\theta, x) \nabla_\theta \log(L(\theta, x)) = \frac{1}{2} L^{1/2}(\theta, x) \dot{\ell}_\theta(x)$$

Insbesondere ist also die Score-Funktion $\dot{\ell}$ die Ableitung der Loglikelihood-Funktion ℓ .

- (ii) Die angenommene Differenzierbarkeit im $\mathcal{L}_2(\mu)$ -Mittel ist eine natürliche Verallgemeinerung der klassischen Differenzierbarkeit. Da $\|L^{1/2}(\theta)\|_{\mathcal{L}_2(\mu)}^2 = \int_{\mathcal{X}} L(\theta, x)\mu(dx) = 1$, folgt $L^{1/2}(\theta) \in \mathcal{L}_2(\mu)$, so dass man $\theta \mapsto L^{1/2}(\theta)$ als $\mathcal{L}_2(\mu)$ -wertige Abbildung auffassen kann und die Familie \mathbb{P} im geometrischen Sinne eine Untermannigfaltigkeit des Hilbertraumes $\mathcal{L}_2(\mu)$ bilden.
- (iii) Nach Definition ist die Fisher-Informationsmatrix symmetrisch und positiv-semidefinit, da $\langle \mathcal{I}_\theta v, v \rangle = \mathbb{E}_\theta (|\langle \dot{\ell}_\theta, v \rangle|^2) \geq 0$ für alle $v \in \mathbb{R}^k$ gilt.
- (iv) Die Score-Funktion und die Fisher-Information sind unabhängig vom dominierenden Maß. Sei \mathbb{P}_o ein privilegiertes dominierendes Maß, dann gilt $L(\theta) = \frac{d\mathbb{P}}{d\mathbb{P}_o} \frac{d\mathbb{P}_o}{d\mu}$, so dass in der Definition von $\dot{\ell}$ der Faktor $\frac{d\mathbb{P}}{d\mu}$ aus dem Integranden ausgeklammert werden kann und $\dot{\ell}$ ebenso die Definition bezüglich des dominierenden Maßes \mathbb{P}_o erfüllt. □

§13.08 **Lemma.** Sei \mathbb{P} eine Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ und für $\theta, \theta_o \in \Theta \subseteq \mathbb{R}^k$ bezeichne $L_{\theta_o}(\theta) \in \overline{\mathcal{A}}^+$ einen Dichtequotient von \mathbb{P} bezüglich \mathbb{P}_o . Für alle θ in einer Umgebung von θ_o gelte $\mathbb{P} \ll \mathbb{P}_o$, so dass $L_{\theta_o}(\theta) = d\mathbb{P}/d\mathbb{P}_o$ eine \mathbb{P}_o -Dichte bezeichnet. Falls die Familie \mathbb{P} $\mathcal{L}_2(\mathbb{P}_o)$ -differenzierbar in θ_o ist, also falls $L_{\theta_o}(\theta) \in \mathcal{L}_2(\mathbb{P}_o)$ für alle θ in einer Umgebung von θ_o und eine \mathbb{R}^k -wertige Funktion $\dot{L}_{\theta_o} \in \mathcal{L}_2^k(\mathbb{P}_o)$ mit

$$\begin{aligned} \lim_{\theta \rightarrow \theta_o} \int \left(\frac{L_{\theta_o}(\theta, x) - L_{\theta_o}(\theta_o, x) - \langle \dot{L}_{\theta_o}(x), \theta - \theta_o \rangle}{\|\theta - \theta_o\|} \right)^2 \mathbb{P}_o(dx) \\ = \lim_{h \rightarrow 0} \frac{\|L_{\theta_o}(\theta_o + h) - L_{\theta_o}(\theta_o) - \langle \dot{L}_{\theta_o}, h \rangle\|_{\mathcal{L}_2(\mathbb{P}_o)}^2}{\|h\|^2} = 0 \end{aligned}$$

existiert, dann ist \mathbb{P} auch Hellinger-differenzierbar in θ_o mit Score-Funktion $\dot{\ell}_{\theta_o} = \dot{L}_{\theta_o}$.

§13.09 **Beweis** von Lemma §13.08. In der Vorlesung. □

§13.10 **Beispiel.** Für $\theta \in \mathbb{R}$ und $\sigma \in \mathbb{R}_0^+$ sei X eine reelle Zufallsvariable mit Lebesgue-dichte $f_\theta(x) = \frac{1}{2\sigma} \exp(-|x - \theta|/\sigma)$, $x \in \mathbb{R}$. Der Parameter σ ist bekannt und der Parameter θ ist unbekannt. Für beliebige $\theta, \theta_o \in \mathbb{R}$ und $x \in \mathbb{R}$ gilt

$$L_{\theta_o}(\theta, x) = \exp\left(-(|x - \theta| - |x - \theta_o|)/\sigma\right).$$

und L_{θ_o} ist $\mathcal{L}_2(\mathbb{P}_o)$ -differenzierbar (Nachweis!) mit

$$\dot{L}_{\theta_o}(x) = (\mathbb{1}_{(\theta_o, \infty)}(x) - \mathbb{1}_{(-\infty, \theta_o)}(x))/\sigma$$

Mit Lemma §13.08 gilt $\dot{\ell}_{\theta_o} = \dot{L}_{\theta_o}$ und für die Fisher-Information erhalten wir

$$\mathcal{I}_{\theta_o} = \mathbb{E}_{\theta_o}(\dot{L}_{\theta_o}^2) = \mathbb{E}_{\theta_o}(\mathbb{1}_{(\theta_o, \infty)}(X) + \mathbb{1}_{(-\infty, \theta_o)}(X))/\sigma^2 = 1/\sigma^2.$$

Die Fisher-Information hängt somit nicht vom unbekanntem Parameter ab, was nur selten der Fall ist. □

§13.11 **Cramér-Rao-Schranke.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_o)$ mit $\Theta \subseteq \mathbb{R}^k$ ein statistisches Experiment, $\gamma : \Theta \rightarrow \mathbb{R}$ ein identifizierbarer und in $\theta_o \in \text{inn}(\Theta)$ differenzierbarer Parameter mit Ableitung $\dot{\gamma}_{\theta_o} \in \mathbb{R}^k$ sowie $\hat{\gamma}$ ein erwartungstreuer Schätzer für γ . Für alle θ in einer Umgebung von θ_o gelte $\mathbb{P} \ll \mathbb{P}_o$ sowie die $\mathcal{L}_2(\mathbb{P}_o)$ -Differenzierbarkeit des Dichtequotienten $L_{\theta_o}(\theta) = d\mathbb{P}/d\mathbb{P}_o$ in θ_o . Falls die Fisher-Informationsmatrix \mathcal{I}_{θ_o} strikt positiv-definit ist, gilt die Cramér-Rao-Ungleichung als untere Schranke für das Risiko bezüglich der quadratischen Verlustfunktion

$$\mathbb{E}_\theta \left(|\hat{\gamma} - \gamma(\theta_o)|^2 \right) = \text{Var}_{\theta_o}(\hat{\gamma}) \geq \langle \mathcal{I}_{\theta_o}^{-1} \dot{\gamma}_{\theta_o}, \dot{\gamma}_{\theta_o} \rangle.$$

§13.12 **Beweis** von **Satz** §13.11. In der Vorlesung. □

§13.13 **Bemerkung.** Ist $\hat{\gamma}$ kein erwartungstreuer Schätzer für γ aber $\hat{\gamma} \in \mathcal{L}_1(\mathbb{P}_\theta)$ für alle $\theta \in \Theta$, so ist $\hat{\gamma}$ ein erwartungstreuer Schätzer für $g(\theta) := \mathbb{E}_\theta(\hat{\gamma})$. In dieser Situation liefert die Cramér-Rao-Ungleichung mit Hilfe der Bias-Varianz-Zerlegung

$$\mathbb{E}_\theta \left(|\hat{\gamma} - \gamma(\theta_0)|^2 \right) \geq |g(\theta_0) - \gamma(\theta_0)|^2 + \langle \mathcal{I}_\theta^{-1} \dot{g}_\theta, \dot{g}_\theta \rangle.$$

Diese Abschätzung ist insbesondere hilfreich, in Situationen in denen erwartungstreue Schätzer von γ nicht existieren oder nicht erstrebenswerte Eigenschaften besitzen. □

§13.14 **Lemma.** Sei \mathbb{P}_θ eine μ -dominierte Familie von Wahrscheinlichkeitsmaßen über $(\mathcal{X}, \mathcal{X})$ mit $\Theta \subseteq \mathbb{R}^k$ offen und für $\theta \in \Theta$ entsprechender Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu$. Weiterhin gelten die folgenden Bedingungen:

- (i) für μ -f.a. $x \in \mathcal{X}$ ist die Abbildung $\theta \mapsto s_\theta(x) := L^{1/2}(\theta, x)$ stetig differenzierbar mit Ableitung $\dot{s}_\theta(x)$,
- (ii) für jedes $\theta \in \Theta$ ist \dot{s}_θ in $\mathcal{L}_2(\mu)$, also $\|\dot{s}_\theta\|_{\mathcal{L}_2(\mu)}^2 \in \mathbb{R}^+$, so dass die Matrix $\mathcal{I}_\theta = 4 \int (\dot{s}_\theta \dot{s}_\theta^t) d\mu$, also $\langle \mathcal{I}_\theta v, v \rangle = 4 \int |\langle \dot{s}_\theta, v \rangle|^2 d\mu$ für alle $v \in \mathbb{R}^k$, wohldefiniert ist,
- (iii) die Abbildung $\theta \mapsto \mathcal{I}_\theta$ ist stetig.

Dann ist \mathbb{P}_θ Hellinger-differenzierbar in jedem $\theta_0 \in \Theta$ mit Score $\dot{\ell}_{\theta_0} = \frac{2\dot{s}_{\theta_0}}{L^{1/2}(\theta_0)} \mathbb{1}_{\{L(\theta) \in \mathbb{R}_0^+\}}$.

§13.15 **Beweis** von **Lemma** §13.14. In der Vorlesung Statistik II. □

§13.16 **Lemma.** Es gelten die Annahmen und Notation aus **Satz** §10.10, so dass $\mathbb{P}_{\Theta_{\text{nat}}}$ eine (strikt) k -parametrische natürliche Exponentialfamilie in S mit natürlichem Parameterraum Θ_{nat} bildet. Dann ist $\mathbb{P}_{\Theta_{\text{nat}}}$ im Innern von Θ_{nat} insbesondere $\mathcal{L}_2(\mathbb{P}_\theta)$ und Hellinger-differenzierbar mit Fisher-Information $\mathcal{I}_\theta = \dot{A}\theta$. Sofern \mathcal{I}_{θ_0} strikt positiv-definit ist, so erreicht S_i , $i = 1, \dots, k$, als erwartungstreuer Schätzer von $\gamma_i(\theta) = \mathbb{E}_\theta(S_i)$ die Cramér-Rao-Schranke (ist Cramér-Rao-effizient) in $\theta_0 \in \text{inn}(\Theta_{\text{nat}})$.

§13.17 **Beweis** von **Lemma** §13.16. In der Vorlesung. □

§13.18 **Beispiel** (Beispiel §09.07 fortgesetzt). Für ein $\sigma \in \mathbb{R}_0^+$ sei $X \odot N_{\mathbb{R} \times \{\sigma^2\}}^n := (N_{(\mu, \sigma^2)}^n)_{\mu \in \mathbb{R}}$, das normale Lokations-Modell beschreibt also adäquat die Verteilung von X . Ein erwartungstreuer Schätzer von μ ist $\hat{\mu} = \bar{X}$. Dann gilt $\text{Var}_\mu(\hat{\mu}) = \sigma^2/n$ sowie für die Fisher-Information $\mathcal{I}_\mu = n/\sigma^2$ (da $A(\mu) = \frac{n\mu^2}{2\sigma^2}$, $\ddot{A}_\mu = n/\sigma^2$). Also ist $\hat{\mu}$ effizient im Sinne der Cramér-Rao-Ungleichung. Um nun $\gamma : \mu \mapsto \gamma(\mu) := \mu^2$ zu schätzen, betrachte den erwartungstreuen Schätzer $\hat{\gamma} = (\bar{X})^2 - \sigma^2/n$. Dann gilt $\text{Var}_\mu(\hat{\gamma}) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}$ (Nachweis!), während die Cramér-Rao-Ungleichung die untere Schranke $4\mu^2\sigma^2/n$ liefert. Damit ist $\hat{\gamma}$ nicht Cramér-Rao-effizient. Allerdings ist \bar{X} eine erschöpfende und vollständige Statistik, so dass der **Satz** §12.01 von **Lehmann-Scheffé** zeigt, dass $\hat{\gamma}$ minimale Varianz unter allen erwartungstreuen Schätzern besitzt. Demnach ist die Cramér-Rao-Schranke hier nicht scharf. □

§13.19 **Bemerkung.** Die Cramér-Rao-Schranke wird nur erreicht wenn \mathbb{P}_θ eine Exponentialfamilie in S bildet und $\gamma(\theta) = \mathbb{E}_\theta(S)$ oder eine lineare Funktion davon zu schätzen sind. Wegen der Vollständigkeit der Statistik S könnte man in diesen Fällen auch mit dem **Satz** §12.01 von **Lehmann-Scheffé** argumentieren. Im nächsten Kapitel betrachten wir allgemeinere Schätzverfahren die zu mindestens asymptotisch die Cramér-Rao-Schranke erreichen. □

§13.20 **Lemma.** Es sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ mit $\Theta \subseteq \mathbb{R}^k$ ein in $\theta_0 \in \Theta$ Hellinger-differenzierbares statistisches Experiment. Dann ist die Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu$ insbesondere $\mathcal{L}_1(\mu)$ -differenzierbar mit Ableitung $\dot{L}_\theta = \dot{\ell}_\theta L(\theta)$, und es gilt $\mathbb{E}_\theta(\dot{\ell}_\theta) = 0$.

§13.21 **Beweis** von **Lemma** §13.20. In der Vorlesung. □

§13.22 **Lemma**. Für jedes $i \in \llbracket n \rrbracket$ sei $(\mathcal{X}^{(i)}, \mathcal{X}^{(i)}, \mathbb{P}_\theta^{(i)})$ ein in $\theta_0 \in \Theta$ Hellinger-differenzierbares statistisches Experiment über derselben Parametermenge $\Theta \subseteq \mathbb{R}^k$ mit Fisher-Informationsmatrix $\mathcal{I}_\theta^{(i)}$. Dann ist das statistische Produktexperiment $(\times_{i \in \llbracket n \rrbracket} \mathcal{X}^{(i)}, \otimes_{i \in \llbracket n \rrbracket} \mathcal{X}^{(i)}, (\otimes_{i \in \llbracket n \rrbracket} \mathbb{P}_\theta^{(i)})_{\theta \in \Theta})$ in $\theta_0 \in \Theta$ Hellinger-differenzierbar mit Fisher-Informationsmatrix

$$\forall \theta \in \Theta : \mathcal{I}_\theta = \sum_{i \in \llbracket n \rrbracket} \mathcal{I}_\theta^{(i)}.$$

§13.23 **Beweis** von **Lemma** §13.22. In der Vorlesung. □

§13.24 **Beispiel** (*Beispiel §13.10 fortgesetzt*). Für $\sigma \in \mathbb{R}_0^+$ und $\theta \in \mathbb{R}$ seien $(X_i)_{i \in \llbracket n \rrbracket}$ unabhängige, reelle Zufallsvariablen mit identischer Lebesgue-dichte $f_\theta(x) = \frac{1}{2\sigma} \exp(-|x - \theta|/\sigma)$, $x \in \mathbb{R}$. Der Parameter σ ist bekannt und der Parameter θ unbekannt. Da $\mathcal{I}_\theta^j = 1/\sigma^2$ die Fisher-Information im statistischen Modell der Zufallsvariable X_j für $j \in \llbracket n \rrbracket$ ist, erhalten wir die Fisher-Information $\mathcal{I}_\theta = n/\sigma^2$ im Produktmodell. □

§14 Translations-äquivalente Schätzer

Wir betrachten im Folgenden translations-äquivalente Schätzer für einen unbekanntem Parameter $\theta \in \mathbb{R}$, d.h. Schätzer $\hat{\theta}$ mit der Eigenschaft $\hat{\theta}(X + a\mathbb{1}_n) = \hat{\theta}(X) + a$ wobei wir für $x \in \mathbb{R}^n$ und $a \in \mathbb{R}$ schreiben $x + a\mathbb{1}_n = (x_1 + a, \dots, x_n + a)$. Skalen-äquivalente Schätzer sowie eine allgemeinere Darstellung findet man zum Beispiel in Kapitel 3 in Lehmann and Casella [1998].

§14.01 **Definition**. Es seien $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\theta)$ ein bzgl. eines translations-invarianten, σ -endlichen Maßes μ dominiertes statistisches Experiment mit Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu$, $\Theta = \mathbb{R}$ sowie $0 \in \Theta$. Wir bezeichnen \mathbb{P}_θ als *Lokations-Familie*, falls

$$L(\theta, x) = L(0, x - \theta\mathbb{1}_n) \quad \text{für alle } x \in \mathbb{R}^n \text{ und } \theta \in \Theta$$

gilt. Ein Schätzer $\hat{\theta}$ von θ heißt *translations-äquivalent* (TäS), falls $\hat{\theta}(x + a\mathbb{1}_n) = \hat{\theta}(x) + a$ für alle $x \in \mathbb{R}^n$, $a \in \mathbb{R}$ gilt. Wir bezeichnen eine Verlustfunktion ν und das entsprechende Risiko \mathcal{R}_ν als *translations-invariant*, falls $\nu(\theta, \hat{\theta}) = \nu(\theta + a, \hat{\theta} + a)$ für alle $\theta, \hat{\theta}, a \in \mathbb{R}$ gilt. Ein TäS $\hat{\theta}_0$ heißt *bester translations-äquivalenter Schätzer* bezüglich des translations-invarianten Risikos \mathcal{R}_ν , falls $\mathcal{R}_\nu(\theta, \hat{\theta}_0) \leq \mathcal{R}_\nu(\theta, \hat{\theta})$ für alle $\hat{\theta} \in \Delta_{\text{TäS}}$ und $\theta \in \Theta$ gilt. □

§14.02 **Lemma**. Es seien \mathbb{P}_θ eine Lokations-Familie und \mathcal{R}_ν ein translations-invariantes Risiko. Für jeden translations-äquivalenten Schätzer $\hat{\theta}$ gilt $\mathcal{R}_\nu(\theta, \hat{\theta}) = \mathcal{R}_\nu(0, \hat{\theta})$, für alle $\theta \in \Theta$, d.h. das Risiko ist konstant in (unabhängig von) θ , und somit ist $\hat{\theta}_0$ ein bester translations-äquivalenter Schätzer, falls $\mathcal{R}_\nu(0, \hat{\theta}_0) \leq \mathcal{R}_\nu(0, \hat{\theta})$ für alle translations-äquivalenten Schätzer $\hat{\theta}$ gilt.

§14.03 **Beweis** von **Lemma** §14.02. In der Vorlesung. □

§14.04 **Lemma**. Es seien $n \geq 2$, $V(X) := (X_i - X_n)_{i \in \llbracket n-1 \rrbracket}$, $\hat{\theta}$ ein translations-äquivalenter Schätzer und $\tilde{\theta}$ ein beliebiger Schätzer für θ . Dann sind die folgenden Aussagen äquivalent:

- (i) $\tilde{\theta}$ ist ein translations-äquivalenter Schätzer.
- (ii) Es existiert eine translations-invariante Funktion $u : \mathbb{R}^n \rightarrow \mathbb{R}$, d.h. $u(x + a\mathbb{1}_n) = u(x)$ für alle $x \in \mathbb{R}^n$, $a \in \mathbb{R}$, so dass $\tilde{\theta} = \hat{\theta} + u$ gilt.
- (iii) Es existiert eine Abbildung $h : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ mit $\tilde{\theta} = \hat{\theta} - h(V)$.

§14.05 **Beweis** von **Lemma** §14.04. In der Vorlesung. □

§14.06 **Satz.** Es seien \mathbb{P}_θ eine Lokations-Familie mit $n \geq 2$, $V(X) := (X_i - X_n)_{i \in [n-1]}$, \mathcal{R}_ν ein translations-invariantes Risiko und $\hat{\theta}$ ein translations-äquivarianter Schätzer mit $\mathcal{R}_\nu(0, \hat{\theta}) \in \mathbb{R}^+$. Falls eine Borel-messbare Funktion $h^* : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ mit

$$h^*(v) \in \arg \min_{h \in \mathbb{R}} \mathbb{E}_0(\nu(0, \hat{\theta} - h) | V = v) \quad \text{für } \mathbb{P}_0^V\text{-f.a. } v \in \mathbb{R}^{n-1}$$

existiert, so ist $\hat{\theta}_0 := \hat{\theta} - h^*(V)$ ein bester translations-äquivarianter Schätzer bezüglich des translations-invarianten Risiko \mathcal{R}_ν .

§14.07 **Beweis** von **Satz** §14.06. In der Vorlesung. □

§14.08 **Korollar.** Die Voraussetzungen im **Satz** §14.06 seien erfüllt.

- (i) Ist ν die quadratische Verlustfunktion, d.h. $\nu(\theta, e) = (\theta - e)^2$, so ist der beste translations-äquivariante Schätzer $\hat{\theta}_0$ eindeutig bestimmt mit $h^* = \mathbb{E}_0(\hat{\theta} | V)$.
- (ii) Für den Absolutbetrag $\nu(\theta, e) = |\theta - e|$ ist $\hat{\theta}_0$ ein bester translations-äquivarianter Schätzer falls h^* ein Median der bedingten $\mathbb{P}_0^{\hat{\theta}|V}$ von $\hat{\theta}$ gegeben V ist.

§14.09 **Beweis** von **Korollar** §14.08. In der Vorlesung. □

§14.10 **Beispiel** (*Beispiel* §13.18 fortgesetzt). Betrachte das *normale Lokations-Modell* $X \odot N_{\mathbb{R} \times \{\sigma^2\}}^n$ für bekanntes $\sigma \in \mathbb{R}_0^+$. Das arithmetische Mittel $\hat{\mu} = \bar{X}$ ist ein translations-äquivarianter Schätzer. Da $V = (X_i - X_n)_{i \in [n-1]} \sim N_{(0, 2\sigma^2)}^{n-1}$ gilt, ist die Verteilung $\mathbb{P}^V := \mathbb{P}_\mu^V$ unabhängig von μ und somit ist V eine unwesentliche (ancillary) Statistik für μ . Nach dem **Satz** §11.03 von **Basu** sind V und \bar{X} somit unabhängig (\bar{X} ist erschöpfend und vollständig) und $h^*(V) = h^*$ im **Satz** §14.06 ist konstant. Betrachten wir eine quadratische Verlustfunktion, so sieht man leicht dass $h^* = 0$ gilt. Damit ist $\hat{\mu}$ auch ein bester translations-äquivarianter Schätzer bezüglich des quadratischen Risiko. Allgemeiner, ist $\nu(\theta, e) = \rho(\theta - e)$ für eine konvexe und gerade Funktion ρ , so minimiert h^* den Ausdruck $\mathbb{E}_0 \rho(\bar{X} - v)$. Man kann leicht zeigen, dass ein Minimum für $h^* = 0$ angenommen wird, d.h. $\hat{\mu}$ ist auch in dieser Situation ein bester translations-äquivarianter Schätzer. □

§14.11 **Satz.** Es sei \mathbb{P}_θ eine Lokations-Familie bzgl. des Lebesguemaßes auf $(\mathbb{R}^n, \mathcal{B}^n)$ und \mathcal{R}_ν das Risiko bzgl. der quadratischen Verlustfunktion. Dann ist

$$\hat{\theta}_0(x) = \frac{\int_{\mathbb{R}} u L(0, x - u \mathbf{1}_n) du}{\int_{\mathbb{R}} L(0, x - u \mathbf{1}_n) du}.$$

ein bester translations-äquivarianter Schätzer. In dieser Form heißt der Schätzer $\hat{\theta}_0$ **Pitman-Schätzer**.

§14.12 **Beweis** von **Satz** §14.11. In der Vorlesung. □

§14.13 **Beispiel.** Es seien $(X_i)_{i \in [n]}$ unabhängige und identisch $U_{[\theta - \frac{1}{2b}, \theta + \frac{1}{2b}]}$ -verteilte ZV'en. Dann gilt für die Likelihood-Funktion

$$L(\theta, x) = L(0, x - \theta \mathbf{1}_n) = b^n \prod_{i=1}^n \mathbf{1}_{[-\frac{1}{2b}, \frac{1}{2b}]}(x_i - \theta) = b^n \mathbf{1}_{[x_{(n)} - \frac{1}{2b}, x_{(1)} + \frac{1}{2b}]}(\theta)$$

und $\hat{\theta}_0 = \frac{1}{2}(x_{(1)} + x_{(n)})$ ist der Pitman-Schätzer. □

4 Allgemeine Schätzmethoden

§15 Momentenschätzer

§15.01 **Definition.** Sei $(\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_\theta^n)$ ein statistisches Produktmodell mit $\mathcal{X} \subseteq \mathbb{R}$, $\mathcal{X} \subseteq \mathcal{B}$ und abgeleitetem Parameter $\gamma : \Theta \rightarrow \mathbb{R}^p$. Ferner sei $\psi = (\psi_i)_{i \in \llbracket n \rrbracket} \in \mathcal{X}^q$ eine Statistik in $\mathcal{L}_1(\mathbb{P}_\theta)$ also mit Koordinatenfunktionen $\psi_j \in \mathcal{L}_1(\mathbb{P}_\theta)$, $j \in \llbracket q \rrbracket$, für alle $\theta \in \Theta$, und $\varphi : \Theta \rightarrow \mathbb{R}^q$ mit

$$\theta \mapsto \varphi(\theta) := \mathbb{E}_\theta(\psi) = (\mathbb{E}_\theta(\psi_j))_{j \in \llbracket q \rrbracket} = \left(\int_{\mathcal{X}} \psi_j(x) \mathbb{P}_\theta(dx) \right)_{j \in \llbracket q \rrbracket}.$$

Weiterhin sei $\delta : \varphi(\Theta) \rightarrow \gamma(\Theta)$ eine Borel-messbare Funktion mit $\delta \circ \varphi = \gamma$. Ist $x \in \mathcal{X}^n$ eine Stichprobe mit $\bar{\psi}(x) := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \psi(x_i) \in \varphi(\Theta)$, so wird $\hat{\gamma}(x) := \delta(\bar{\psi}(x))$ **Momentenschätzwert** für $\gamma(\theta)$ mit Momentenfunktionen $\psi = (\psi_i)_{i \in \llbracket q \rrbracket}$ genannt. \square

§15.02 **Beispiel.**

- (a) Es sei $(X_i)_{i \in \llbracket n \rrbracket} \odot \text{Exp}_{\mathbb{R}_0^+}^n := (\text{Exp}_\lambda^n)_{\lambda \in \mathbb{R}_0^+}$ (vgl. **Beispiel A01.06 (h)**). Betrachte die übliche Momentenfunktion $\psi(x) = x^k$ für ein $k \in \mathbb{N}$, für $X_i \sim \text{Exp}_\lambda$, $i \in \llbracket n \rrbracket$, gilt dann $\varphi(\lambda) = \mathbb{E}_\lambda(\psi) = \mathbb{E}_\lambda(X_i^k) = \lambda^{-k} k!$. Ist $\gamma(\lambda) = \lambda$ der abgeleitete Parameter, so ergibt sich $\delta \circ \varphi = \gamma$ für $\delta(x) = (k!/x)^{1/k}$. Der k -te Momentenschätzer für λ ist damit

$$\hat{\lambda}_n := \left(\frac{k!}{\frac{1}{n} \sum_{i=1}^n X_i^k} \right)^{1/k}.$$

- (b) Betrachte einen *autoregressiven Prozess* der Ordnung 1 (AR(1)-Prozess):

$$X_n = aX_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

mit unabhängig und identisch verteilten Fehlertermen $(\varepsilon_i)_{i \in \mathbb{N}}$, für die $\mathbb{E}(\varepsilon_1) = 0$, $\text{Var}(\varepsilon_1) = \sigma^2 < \infty$ gilt, sowie $X_0 = x_0 \in \mathbb{R}$. Insbesondere, motiviert in dieser Situation die Identität $\mathbb{E}_a(X_{n-1}X_n | (\varepsilon_i)_{i \in \llbracket n-1 \rrbracket}) = aX_{n-1}^2$, zum **Yule-Walker-Schätzer**

$$\hat{a}_n := \frac{\frac{1}{n} \sum_{k \in \llbracket n \rrbracket} X_{k-1}X_k}{\frac{1}{n} \sum_{k \in \llbracket n \rrbracket} X_{k-1}^2} = a + \frac{\frac{1}{n} \sum_{k \in \llbracket n \rrbracket} X_{k-1}\varepsilon_k}{\frac{1}{n} \sum_{k \in \llbracket n \rrbracket} X_{k-1}^2}.$$

Im Fall $|a| < 1$ kann man mit Hilfe des Ergodensatzes auf die Konsistenz von \hat{a}_n für $n \rightarrow \infty$ schließen. Allgemeiner zeigt man, dass $M_n := \sum_{k \in \llbracket n \rrbracket} X_{k-1}\varepsilon_k$ ein Martingal bezüglich der Filtration $\mathcal{F}_n := \sigma((\varepsilon_i)_{i \in \llbracket n \rrbracket})$ ist mit quadratischer Variation $\langle M \rangle_n := \sum_{k \in \llbracket n \rrbracket} X_{k-1}^2$. Das starke Gesetz der großen Zahlen für \mathcal{L}_2 -Martingale liefert dann die (starke) Konsistenz

$$\hat{a}_n = a + \frac{M_n}{\langle M \rangle_n} \xrightarrow{\mathbb{P}\text{-f.s.}} a. \quad \square$$

§15.03 **Lemma.** Es seien $(X_i)_{i \in \llbracket n \rrbracket} \odot \mathbb{P}_\theta^n$ und $\bar{\psi}_n := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \psi(X_i)$ in $\varphi(\Theta)$ für hinreichend großes n . Ist δ stetig, so ist der Momentenschätzer $\hat{\gamma}_n := \delta(\bar{\psi}_n)$ (stark) konsistent, d.h. \mathbb{P}_θ -f.s. gilt $\lim_{n \rightarrow \infty} \hat{\gamma}_n = \gamma(\theta)$.

§15.04 **Beweis** von **Lemma** §15.03. In der Vorlesung. \square

§15.05 **Satz (Δ -Methode).** Es seien $(X_i)_{i \in \mathbb{N}}$ eine Folge von zufälligen \mathbb{R}^k -wertigen Vektoren, $\sigma_n \in \mathbb{R}_0^+$ mit $\lim_{n \rightarrow \infty} \sigma_n = 0$, $\theta_0 \in \mathbb{R}^k$ sowie $\Sigma \in \mathbb{R}_{>}^{(k,k)}$ (positiv semi-definit) und es gelte

$$\sigma_n^{-1}(X_n - \theta_0) \xrightarrow{D} N_{(0, \Sigma)}.$$

Ist $\gamma : \mathbb{R}^k \rightarrow \mathbb{R}$ in einer Umgebung von θ_0 stetig differenzierbar, so folgt

$$\sigma_n^{-1}(f(X_n) - f(\theta_0)) \xrightarrow{D} N_{(0, \langle \Sigma \dot{\gamma}_{\theta_0}, \dot{\gamma}_{\theta_0} \rangle)},$$

wobei $N_{(0,0)}$ gegebenenfalls als Punktmaß δ_0 in der Null zu verstehen ist.

§15.06 **Beweis von Satz §15.05.** In der Vorlesung. □

§15.07 **Beispiel (Beispiel §12.07 (b) fortgesetzt).** Sei $(X_i)_{i \in \mathbb{N}} \odot \text{Poi}_{\mathbb{R}_0^+}^n$ mit unbekanntem Parameter $\lambda \in \mathbb{R}_0^+$. Da die vollständige und erschöpfende Statistik $\hat{\lambda}_n := \bar{X}$ ein erwartungstreuer Schätzer für λ ist, ist $\hat{\lambda}_n$ der KVS. Nach dem zentralen Grenzwertsatz gilt $\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{D} N_{(0, \lambda)}$ unter $\text{Poi}_{\lambda}^{\mathbb{N}}$. Ist das Ziel nun ein asymptotisches Konfidenzintervall herzuleiten, so stört die Abhängigkeit der asymptotischen Varianz vom unbekanntem Parameter. Betrachtet man nun $\gamma(x) = 2x^{1/2}$ mit $\dot{\gamma}_x = x^{-1/2}$, so folgt mit Hilfe der Δ -Methode, dass $\sqrt{n}(2\hat{\lambda}_n^{1/2} - 2\lambda^{1/2}) \xrightarrow{D} N_{(0,1)}$, so dass $[2\hat{\lambda}_n^{1/2} \pm n^{-1/2}z_{1-\alpha/2}]$ ein asymptotisches $(1 - \alpha)$ -Konfidenzintervall für $2\lambda^{1/2}$ bildet, wobei $z_{1-\alpha/2}$ das $(1 - \alpha/2)$ Quantil einer Standardnormalverteilung bezeichnet. Eine Rücktransformation ergibt dann für λ selbst das asymptotische $(1 - \alpha)$ -Konfidenzintervall $[(\hat{\lambda}_n^{1/2} - (4n)^{-1/2}z_{1-\alpha/2})^2, (\hat{\lambda}_n^{1/2} + (4n)^{-1/2}z_{1-\alpha/2})^2]$. Die Idee, mittels Δ -Transformation eine vom unbekanntem Parameter unabhängige asymptotische Varianz zu erhalten, ist in vielen Situation sehr erfolgreich und wird *Varianz-stabilisierende Transformation* genannt.

Alternativ kann man die asymptotische Varianz mittels $\hat{\lambda}_n$ konsistent schätzen und mit Hilfe des Slutsky-Lemma dann auf $(n/\hat{\lambda}_n)^{1/2}(\hat{\lambda}_n - \lambda) \xrightarrow{D} N_{(0,1)}$ schließen. Daraus ergibt sich $[\hat{\lambda}_n \pm (\hat{\lambda}_n/n)^{1/2}z_{1-\alpha/2}]$ als asymptotisches Konfidenzintervall. □

§15.08 **Satz.** Es seien $(X_i)_{i \in \mathbb{N}} \odot \mathbb{P}_0^n$, $\theta_0 \in \Theta$, $\gamma : \Theta \rightarrow \mathbb{R}$ und $\bar{\psi}_n := \frac{1}{n} \sum_{i \in \mathbb{N}} \psi(X_i)$ in $\varphi(\Theta)$ für hinreichend großes n mit Momentenfunktionen $\psi = (\psi_j)_{j \in \llbracket q \rrbracket} \in \mathcal{L}_2(\mathbb{P}_0)$. Bezeichne mit $\Sigma_{\theta_0} \in \mathbb{R}_{>}^{q \times q}$ die Kovarianzmatrix von ψ mit den Einträgen $(\Sigma_{\theta_0})_{ij} = \text{Cov}_{\theta_0}(\psi_i, \psi_j)$ für $i, j \in \llbracket q \rrbracket$. Sofern δ in einer Umgebung von $\varphi(\theta_0)$ stetig differenzierbar ist, ist der standardisierte Momentenschätzer $\hat{\gamma}_n := \delta(\bar{\psi}_n)$ unter $\mathbb{P}_0^{\mathbb{N}}$ asymptotisch normalverteilt mit Rate $n^{-1/2}$, asymptotischem Erwartungswert Null und asymptotischer Varianz $\langle \Sigma_{\theta_0} \dot{\delta}_{\varphi(\theta_0)}, \dot{\delta}_{\varphi(\theta_0)} \rangle$:

$$\sqrt{n}(\hat{\gamma}_n - \gamma(\theta_0)) \xrightarrow{D} N_{(0, \langle \Sigma_{\theta_0}(\psi) \dot{\delta}_{\varphi(\theta_0)}, \dot{\delta}_{\varphi(\theta_0)} \rangle)} \quad (\text{unter } \mathbb{P}_0).$$

§15.09 **Beweis von Satz §15.08.** In der Vorlesung. □

§15.10 **Bemerkung.** Die Begriffe *asymptotischer Erwartungswert* und *asymptotische Varianz* sind leicht irreführend, da nicht notwendigerweise gilt, dass die Momente von $\sqrt{n}(\hat{\gamma}_n - \gamma(\theta_0))$ gegen die entsprechenden Momente der asymptotischen Verteilung konvergieren (dafür wird gleichgradige Integrierbarkeit benötigt). □

§15.11 **Beispiel (Beispiel §15.02 (a) fortgesetzt).** Es gilt $\Sigma_{\lambda_0}(\psi) = \text{Var}_{\lambda_0}(X_i^k) = ((2k)! - (k!)^2)/\lambda_0^{2k}$ und $\dot{\delta}_x = -(k!/x)^{1/k}(kx)^{-1}$. Für jedes $k \in \mathbb{N}$ ist der Momentenschätzer $\hat{\lambda}_n$ asymptotisch normalverteilt mit Rate $n^{-1/2}$ und asymptotischer Varianz $\sigma_k^2 = \lambda_0^2 k^{-2} ((2k)!/(k!)^2 - 1)$. Da für $k = 1$ der Momentenschätzer $\hat{\lambda}_n := (\bar{X})^{-1}$ die gleichmäßig kleinste asymptotische Varianz besitzt und auf der erschöpfenden Statistik \bar{X} basiert, wird dieser Schätzer im Allgemeinen vorgezogen. □

§15.12 **Bemerkung.** Die Momentenmethode kann unter folgendem allgemeinem Gesichtspunkt betrachtet werden. Sind $(X_i)_{i \in \llbracket n \rrbracket}$ unabhängige und identisch \mathbb{P}_θ -verteilte reelle Zufallsvariablen, so ist die *empirische Verteilungsfunktion* $\widehat{\mathbb{F}}_n(x) := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \mathbb{1}_{(-\infty, x]}(X_i)$ eine erschöpfende Statistik und nach dem Satz von Glivenko-Cantelli gilt $\widehat{\mathbb{F}}_n(x) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{F}_\theta(x) = \mathbb{P}_\theta((-\infty, x])$ gleichmäßig in $x \in \mathbb{R}$. Ist nun $\gamma(\theta)$ als Funktional $\delta(\mathbb{F}_\theta)$ darstellbar, so ist die empirische Version $\delta(\widehat{\mathbb{F}}_n)$ ein natürlicher Schätzer für $\gamma(\theta)$. Falls das Funktional δ stetig bezüglich der Supremumsnorm ist, so folgt die Konsistenz.

Der Satz von Donsker für empirische Prozesse zeigt $\sqrt{n}(\widehat{\mathbb{F}}_n - \mathbb{F}_\theta) \xrightarrow{w} G_\theta$ gleichmäßig auf \mathbb{R} für einen zentrierten Gaußprozess G_θ mit Kovarianzstruktur $\text{Cov}(G_\theta(x), G_\theta(y)) = \mathbb{F}_\theta(x \wedge y) - \mathbb{F}_\theta(x)\mathbb{F}_\theta(y)$. Ist δ ein Hadamard-differenzierbares Funktional, so folgt $\sqrt{n}(\delta(\widehat{\mathbb{F}}_n) - \gamma(\theta)) \xrightarrow{D} \dot{\delta}(\mathbb{F}_\theta)G_\theta$ unter \mathbb{P}_θ , also insbesondere asymptotische Normalverteilung mit Rate $n^{-1/2}$ und explizit bestimmbarer asymptotischer Varianz. Eine detaillierte Darstellung findet man zum Beispiel in van der Vaart [1998].

Als einfaches (lineares) Beispiel sei $\gamma(\theta) = \mathbb{E}_\theta(\psi)$ zu schätzen und $X_1 \geq 0$ \mathbb{P}_θ -f.s.. Dann folgt informell $\delta(\mathbb{F}_\theta) = \int_{\mathbb{R}^+} \psi(x) d\mathbb{F}_\theta(x) = \int_{\mathbb{R}^+} \dot{\psi}_x(1 - \mathbb{F}_\theta(x)) dx$. Aus der Linearität folgt weiterhin $\dot{\delta}(\mathbb{F}_\theta)G_\theta = \int_{\mathbb{R}^+} \dot{\psi}_x(-G_\theta(x)) dx$, welches normalverteilt ist mit Erwartungswert Null und Varianz

$$\begin{aligned} & \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \dot{\psi}_x \dot{\psi}_y (\mathbb{F}_\theta(x \wedge y) - \mathbb{F}_\theta(x)\mathbb{F}_\theta(y)) dx dy \\ &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \psi(x)\psi(y) \partial_{xy} (\mathbb{F}_\theta(x \wedge y) - \mathbb{F}_\theta(x)\mathbb{F}_\theta(y)) dx dy \\ &= \int_{\mathbb{R}^+} \psi^2(x) d\mathbb{F}_\theta(x) - \left(\int_{\mathbb{R}^+} \psi^2(x) d\mathbb{F}_\theta(x) \right)^2 \end{aligned}$$

was gerade der Varianz von $\delta(\widehat{\mathbb{F}}_n) = \frac{1}{\sqrt{n}} \sum_{i \in \llbracket n \rrbracket} \psi(X_i)$ entspricht. \square

§16 Maximum-Likelihood-Schätzer

§16.01 **Beispiel.** (a) Auf dem Stichprobenraum \mathcal{X} sei jede Verteilung \mathbb{P}_θ , $\theta \in \Theta$, durch eine Zähldichte \mathbb{p}_θ gegeben. Die Verlustfunktion $\nu(\theta, \delta)$ sei homogen in $\theta \in \Theta$, dann ist eine plausible Schätzmethode für θ , bei Vorliegen einer Realisation x als Schätzwert $\widehat{\theta}(x)$ denjenigen Parameter $\theta \in \Theta$ zu wählen, für den die Wahrscheinlichkeit $\mathbb{p}_\theta(x)$ des Eintretens von x maximiert wird, d.h. $\widehat{\theta}(x) := \arg \max_{\theta \in \Theta} \mathbb{p}_\theta(x)$. Dieser Schätzer wird *Maximum-Likelihood-Schätzer* (MLS) genannt, wobei weder Existenz noch Eindeutigkeit ohne Weiteres garantiert sind. Bei Mehrdeutigkeit wählt man einen maximierenden Parameter θ nach Belieben. Im Fall $(X_i)_{i \in \llbracket n \rrbracket} \rightsquigarrow \text{Poi}_{\mathbb{R}_0^+}^n$ mit unbekanntem Parameter $\lambda \in \mathbb{R}_0^+$, ergibt sich für $\bar{X} \in \mathbb{R}_0^+$ beispielsweise

$$\widehat{\lambda} = \arg \max_{\lambda \in \mathbb{R}_0^+} \prod_{i \in \llbracket n \rrbracket} \left(e^{-\lambda} \frac{\lambda^{X_i}}{X_i!} \right) = \bar{X}.$$

Für $\bar{X} = 0$, d.h. $X_1 = \dots = X_n = 0$ wird das Supremum nur asymptotisch für $\lambda \rightarrow 0$ erreicht. Hier könnte man sich behelfen, indem man die Verteilungsfamilie mit Poi_0 als Punktmaß in der Null stetig ergänzt.

(b) Ist jede Verteilung \mathbb{P}_θ , $\theta \in \Theta$, durch eine Lebesgue-Dichte f_θ gegeben, so führt der Maximum-Likelihood-Ansatz zu dem Schätzwert $\widehat{\theta}(x) := \arg \max_{\theta \in \Theta} f_\theta(x)$. Sei $Y \rightsquigarrow \text{N}_{\mathbb{R} \times \{1\}}$ mit unbekanntem Parameter $\mu \in \mathbb{R}$. Dann heißt $X = \exp(Y)$ log-normalverteilt, und für den MLS

gilt

$$\hat{\mu} = \arg \max_{\mu \in \mathbb{R}} \frac{\exp(-(\log(X) - \mu)^2/2)}{\sqrt{2\pi}X} = \log(X).$$

Auf der anderen Seite unter Verwendung von Y erhält man den MLS $\hat{\mu} = Y$. Der MLS ist somit invariant unter Parametertransformationen, da Einsetzen von $Y = \log(X)$ auf das selbe Ergebnis führt. Interessanterweise führt die Momentenmethode unter Benutzung von $\mathbb{E}_\mu(X) = \exp(\mu + 1/2)$ auf den Momentenschätzer $\tilde{\mu} = \log(X) - 1/2$, während der MLS $\hat{\mu} = Y$ auch Momentenschätzer ist, da $\mathbb{E}_\mu(Y) = \mu$. Momentenschätzer bezüglich der selben Momentenfunktion sind also im Allgemeinen nicht transformationsinvariant. \square

§16.02 **Definition.** Es sei $(\mathcal{X}, \mathcal{X}, \mathbb{P})$ ein bzgl. eines σ -endlichen Maßes $\mu \in \mathcal{M}_\sigma(\mathcal{X})$ dominiertes statistisches Modell mit Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu$ für $\theta \in \Theta$ und (Θ, \mathcal{T}) ein messbarer Raum. Eine Statistik $\hat{\theta}$ auf $(\mathcal{X}, \mathcal{X})$ mit Werten in (Θ, \mathcal{T}) wird *Maximum-Likelihood-Schätzer (MLS)* für θ genannt, wenn $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$ μ -f.ü. gilt. Ist $\gamma(\hat{\theta})$ eine Statistik auf $(\mathcal{X}, \mathcal{F})$ mit Werten in (Γ, \mathcal{G}) , so wird $\gamma(\hat{\theta})$ *Maximum-Likelihood-Schätzer (MLS)* für den abgeleiteten Parameter $\gamma : \Theta \rightarrow \Gamma$ genannt. \square

§16.03 **Bemerkung.** $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$ μ -f.ü. meint $L(\hat{\theta}(x), x) = \sup_{\theta \in \Theta} L(\theta, x)$ für μ -f.a. alle $x \in \mathcal{X}$. Der MLS braucht weder zu existieren noch eindeutig zu sein, falls er existiert. Er hängt von der gewählten Version der Radon-Nikodym-Dichte ab; es gibt jedoch häufig eine kanonische Wahl, wie beispielsweise im Fall stetiger Lebesguedichten. Außerdem ist eine Abänderung auf einer Nullmenge bezüglich aller \mathbb{P}_θ irrelevant, weil der Schätzer vor Realisierung des Experiments festgelegt wird und diese Realisierung damit fast sicher zum selben Schätzwert führen wird. \square

§16.04 **Lemma.** Es sei $\mathbb{P}_{\Theta_{\text{nat}}}$ eine natürliche Exponentialfamilie in S , dann ist der MLS $\hat{\theta}$ implizit durch die Momentengleichung $\mathbb{E}_{\hat{\theta}(x)}(S) = S(x)$ gegeben, vorausgesetzt der MLS existiert und liegt im Innern $\text{inn}(\Theta_{\text{nat}})$ von Θ_{nat} .

§16.05 **Beweis** von Lemma §16.04. In der Vorlesung. \square

§16.06 **Bemerkung.** Liegt ein eindeutiger abgeleiteter Parameter $\gamma : \theta \mapsto \gamma(\theta)$ vor, so ist natürlich $\hat{\gamma} := \gamma(\hat{\theta})$ der MLS für $\gamma(\theta)$. \square

§16.07 **Beispiel.**

(a) (*Beispiel §10.09 (a) fortgesetzt*) Betrachte das *normale Lokations-Skalen-Modell* $X \odot N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, d.h. eine natürliche zwei-parametrische Exponentialfamilie in $S(x) = (\sum_{i \in \llbracket n \rrbracket} x_i, -\sum_{i \in \llbracket n \rrbracket} x_i^2)$ und natürlichem Parameter $\theta = \sigma^{-2}(\mu, 1/2)$. Der MLS $\hat{\theta}$ ist somit implizit durch die Momentengleichungen $\mathbb{E}_{\hat{\theta}(x)}(\bar{X}) = \bar{x}$ und $\mathbb{E}_{\hat{\theta}(x)}(\bar{X}^2) = \bar{x}^2$ gegeben, also $\hat{\mu} = \bar{X}$ und $\widehat{\mu^2 + \sigma^2} = \bar{X}^2$. Mittels Reparametrisierung $(\mu, \mu^2 + \sigma^2) \mapsto (\mu, \sigma^2)$ erhalten wir $\hat{\sigma}^2 = \bar{X}^2 - (\bar{X})^2 = \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} (X_i - \bar{X})^2$. Damit ist der MLS für σ^2 nicht erwartungstreu.

(b) Sei (X_0, X_1, \dots, X_n) eine Markovkette auf dem Zustandsraum $S = \llbracket M \rrbracket$ mit vom Parameter unabhängigen Anfangswert $X_0 = x_0$ und unbekanntem Übergangswahrscheinlichkeiten $\mathbb{P}^{X_{k+1}|X_k=i}(\{j\}) = p_{ij}$ ergibt sich die Likelihood-Funktion (bzgl. des Zählmaßes) durch

$$L((p_{kl}), x) = \prod_{i \in \llbracket n \rrbracket} p_{x_{i-1}, x_i} = \prod_{k, l \in \llbracket M \rrbracket} p_{kl}^{N_{kl}(x)},$$

wobei $N_{kl}(x) = |\{i \in \llbracket n \rrbracket : x_{i-1} = k, x_i = l\}|$ die Anzahl der beobachteten Übergänge

von Zustand k nach l angibt. Als MLS erhalten wir damit die relative Häufigkeit $\hat{p}_{ij} = N_{ij} / (\sum_{l \in \llbracket m \rrbracket} N_{il})$ der Übergänge.

(c) In einem parametrischen Regressionsmodell mit Beobachtungen

$$Y_i = g_\theta(x_i) + \varepsilon_i, \quad i \in \llbracket n \rrbracket$$

ergibt sich unter der Normalverteilungsannahme $(\varepsilon_i)_{i \in \llbracket n \rrbracket} \odot N_{(\{0\} \times \mathbb{R}_+^n)}$ als MLS der Kleinste-Quadrate-Schätzer $\hat{\theta} \in \arg \min_{\theta \in \Theta} \sum_{i \in \llbracket n \rrbracket} (Y_i - g_\theta(x_i))^2$. \square

§17 Minimum-Kontrast-Schätzer

§17.01 **Definition.** Es sei $((\mathcal{X}^{(n)}, \mathcal{X}^{(n)}, \mathbb{P}_\theta^{(n)} = (\mathbb{P}_\theta^{(n)})_{\theta \in \Theta})_{n \in \mathbb{N}}$ eine Folge statistische Modelle über demselben Parameterraum Θ sowie $\gamma : \Theta \rightarrow \Gamma$ ein interessierender Parameter. Eine Funktion $K : \Theta \times \Gamma \rightarrow \overline{\mathbb{R}}$ heißt *Kontrastfunktion*, falls für alle $\theta \in \Theta$ die Funktion $K_\theta : \Gamma \mapsto K(\theta, \gamma)$ ein eindeutiges Minimum in $\gamma(\theta)$ besitzt. Eine Folge $(\hat{K}_n)_{n \in \mathbb{N}}$ von Funktionen $\hat{K}_n : \Gamma \times \mathcal{X}^{(n)} \rightarrow \overline{\mathbb{R}}$ heißt zugehöriger *Kontrastprozess* (oder kurz *Kontrast*), falls folgende Bedingungen gelten:

(KP1) Für alle $\gamma \in \Gamma$ ist $\hat{K}_n(\gamma) : x \mapsto \hat{K}_n(\gamma, x)$ eine Statistik, also $\hat{K}_n(\gamma) \in \overline{\mathcal{X}^{(n)}}$.

(KP2) Für alle $\gamma \in \Gamma$ und $\theta \in \Theta$ gilt $\hat{K}_n(\gamma) \xrightarrow{\mathbb{P}_\theta^{(n)}} K_\theta(\gamma)$.

Jede (messbare) Wahl $\hat{\gamma}_n : \mathcal{X}^{(n)} \rightarrow \Gamma$ (sofern existent) mit

$$\hat{K}_n(\hat{\gamma}_n) = \inf_{\gamma \in \Gamma} \hat{K}_n(\gamma) \quad \mathbb{P}_\theta^{(n)}\text{-f.ü.}$$

(nicht notwendigerweise eindeutiger) heißt *Minimum-Kontrast-Schätzer (MKS)* für γ . \square

§17.02 **Definition.** Für zwei Wahrscheinlichkeitsmaße $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathcal{X})$ wird die Funktion

$$\text{KL}(\mathbb{P}_0 | \mathbb{P}_1) = \begin{cases} \mathbb{E}_0 \left(\log \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \right) = \int \log \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_1} \right) d\mathbb{P}_0, & \text{falls } \mathbb{P}_0 \ll \mathbb{P}_1, \\ +\infty, & \text{sonst} \end{cases}$$

Kullback-Leibler-Divergenz (oder auch *Kullback-Leibler-Abstand*, *relative Entropie*) von \mathbb{P}_0 bezüglich \mathbb{P}_1 genannt. \square

§17.03 **Lemma.** Für die Kullback-Leibler-Divergenz gilt:

(i) $\text{KL}(\mathbb{P}_0 | \mathbb{P}_1) \geq 0$ sowie $\text{KL}(\mathbb{P}_1 | \mathbb{P}_1) = 0$ genau dann wenn $\mathbb{P}_0 = \mathbb{P}_1$, aber $\text{KL}(\bullet | \bullet)$ ist nicht symmetrisch;

(ii) für Produktmaße ist die Kullback-Leibler-Divergenz additiv:

$$\text{KL}(\mathbb{P}_{0,1} \otimes \mathbb{P}_{0,2} | \mathbb{P}_{1,1} \otimes \mathbb{P}_{1,2}) = \text{KL}(\mathbb{P}_{0,1} | \mathbb{P}_{1,1}) + \text{KL}(\mathbb{P}_{0,2} | \mathbb{P}_{1,2});$$

(iii) bildet \mathbb{P}_θ eine natürliche Exponentialfamilie und ist θ_0 ein innerer Punkt von Θ , so gilt

$$\text{KL}(\mathbb{P}_\theta | \mathbb{P}_{\theta_0}) = A(\theta) - A(\theta_0) + \langle \dot{A}_{\theta_0}, \theta_0 - \theta \rangle.$$

§17.04 **Beweis** von Lemma §17.03. Übungsaufgabe. \square

§17.05 **Bemerkung.** Betrachte eine natürliche Exponentialfamilie \mathbb{P} in S wie in Lemma §17.03 (iii), so gilt $\dot{A}_{\theta_0} = \text{Cov}_{\theta_0}(S)$ und für $\theta, \theta_0 \in \text{inn}(\Theta)$ erhalten wir mit Hilfe einer Taylorentwicklung $\text{KL}(\mathbb{P}_\theta | \mathbb{P}_{\theta_0}) = \frac{1}{2} \langle \text{Cov}_{\theta_0}(S)(\theta - \theta_0), \theta - \theta_0 \rangle$ für eine Zwischenstelle θ_* zwischen θ und θ_0 . Zur Erinnerung, $\text{Cov}_{\theta_0}(S)$ gibt gerade die Fisher-Information in θ_* an. Im Fall einer mehrdimensionalen $N_{(\mu, \Sigma)}$ mit strikt positiver Kovarianzmatrix Σ folgt nun aus $A(\mu) = \langle \Sigma^{-1} \mu, \mu \rangle / 2$, dass $\ddot{A}_\mu = \Sigma^{-1}$ unabhängig von μ ist und somit $\text{KL}(N_{(\mu, \Sigma)} | N_{(\mu_0, \Sigma)}) = \frac{1}{2} \langle \Sigma^{-1}(\mu - \mu_0), \mu - \mu_0 \rangle$ gilt. \square

§17.06 **Korollar.** Es sei \mathbb{P}_θ eine dominierte Verteilungsfamilie mit Likelihood-Funktion L bzgl. eines privilegierten dominierenden Maßes \mathbb{P}_0 , Log-Likelihood-Funktion $\ell = \log(L)$ und interessierendem Parameter θ (also $\gamma = \text{id}_\Theta$). Des Weiteren, seien \mathbb{P}_θ und \mathbb{P}_{θ_0} äquivalent ($\mathbb{P}_\theta \ll \mathbb{P}_{\theta_0}$ und $\mathbb{P}_{\theta_0} \ll \mathbb{P}_\theta$) für alle $\theta, \theta_0 \in \Theta$. Im Produktexperiment $(\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_\theta^n)$ ist $\hat{K}_n(\theta) \in \overline{\mathcal{X}^n}$ mit

$$x \mapsto \hat{K}_n(\theta, x) := -\frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \ell(\theta, x_i)$$

ein Kontrastprozess zur Kontrastfunktion $K(\theta_0, \theta) := \text{KL}(\mathbb{P}_\theta | \mathbb{P}_{\theta_0}) - \text{KL}(\mathbb{P}_{\theta_0} | \mathbb{P}_{\theta_0})$. Ein zugehöriger Minimum-Kontrast-Schätzer ist dann auch ein Maximum-Likelihood-Schätzer.

§17.07 **Beweis von Korollar §17.06.** In der Vorlesung. \square

§17.08 **Beispiel.**

(a) (*Beispiel §16.07 (c) fortgesetzt*) Zusätzlich seien der interessierende Parameter θ ($\gamma = \text{id}_\Theta$) identifizierbar, d.h. $\theta \neq \theta_0$ impliziert $g_\theta \neq g_{\theta_0}$, die Regressionsfunktionen $g_\theta : [0, 1] \rightarrow \mathbb{R}$ stetig, das Design $x_i = i/n$, $i \in \llbracket n \rrbracket$ äquidistant und die Fehlerterme $(\varepsilon_i)_{i \in \llbracket n \rrbracket}$ unabhängig und identisch verteilt mit $\mathbb{E}(\varepsilon_1) = 0$ und $\mathbb{E}(\varepsilon_1^4) \in \mathbb{R}^+$ (nicht notwendigerweise normalverteilt). Mit Hilfe der Tchebyscheff-Ungleichung und der Riemannschen Summen-Approximation zeigt man, dass $\hat{K}_n(\theta) \in \mathcal{B}^n$ mit

$$y = (y_i)_{i \in \llbracket n \rrbracket} \mapsto \hat{K}_n(\theta, y) := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} (y_i - g_\theta(x_i))^2$$

einen Kontrastprozess zur Kontrastfunktion $K(\theta_0, \theta) := \int_0^1 (g_\theta(x) - g_{\theta_0}(x))^2 ds + \mathbb{E}(\varepsilon^2)$ bildet. Der zugehörige Minimum-Kontrast-Schätzer ist der Kleinste-Quadrate-Schätzer.

(b) Betrachte das Regressionsmodell aus (a) im Fall einer Modellmisspezifikation, d.h. das Modell ist nicht adäquat für die Beobachtung $(Y_i)_{i \in \llbracket n \rrbracket}$, in dem Sinne, dass die Beobachtungen dem Regressionsmodell $Y_i = f(i/n) + \varepsilon_i$, $i \in \llbracket n \rrbracket$, genügen, aber die Funktion $f : [0, 1] \rightarrow \mathbb{R}$ nicht Element der Funktionenklasse $\{g_\theta, \theta \in \Theta\}$ ist. Identifiziert man die Regressionsfunktion mit dem interessierenden Parameter θ , d.h. $\Theta \subseteq \mathcal{L}_2([0, 1])$, so liefert der Kleinste-Quadrate-Ansatz $\hat{K}_n(\hat{\theta}) = \min_{\theta \in \Theta} \hat{K}_n(\theta)$. Dann erhält man nach obiger Herleitung im Grenzwert eine „Kontrast-Typ-Funktion“ $K(f, \theta) := \|\theta - f\|_{\mathcal{L}_2}^2 + \mathbb{E}(\varepsilon_1^2)$. Für $f \notin \Theta$ wird das Minimum natürlich nicht in f angenommen. Man kann die Kontrasttheorie durch Wahl der Funktion γ jedoch trotzdem anwenden. Dazu nehmen wir an, dass die Menge interessierender Parameter Γ Riemann-integrierbare Funktionen enthält, sowie abgeschlossen in $\mathcal{L}_2([0, 1])$ und konvex ist, so dass für jede Funktion $\theta_0 \in \mathcal{L}_2([0, 1])$ eine eindeutige \mathcal{L}_2 -Orthogonalprojektion $\gamma_0 := \gamma(\theta_0)$ von θ_0 auf Γ existiert. Beispielsweise kann Γ die Menge aller Polynome vom Grad $\leq d$ sein. Bezeichnet Θ die Menge der quadratisch Riemann-integrierbaren Funktionen in $\mathcal{L}_2([0, 1])$, so ist $\hat{K}_n(\gamma) \in \mathcal{B}^n$ mit

$$y \mapsto \hat{K}_n(\gamma, y) := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} (y_i - \gamma(i/n))^2$$

$\gamma \in \Gamma$ ein Kontrastprozess zur Kontrastfunktion $K_{\theta_0} : \gamma \mapsto K(\theta_0, \gamma) = \|\theta_0 - \gamma\|_{\mathcal{L}_2}^2 + \mathbb{E}(\varepsilon_1^2)$, welche genau in $\gamma_0 = \gamma(\theta_0)$ ihr Minimum in Γ annimmt. Unter geeigneten Bedingungen

konvergiert der KQS $\hat{\gamma}_n$ mit $\hat{K}_n(\hat{\gamma}_n) = \min_{\gamma \in \Gamma} \hat{K}_n(\gamma)$ unter $\mathbb{P}_\theta^{(n)}$ gegen $\gamma_o = \gamma(\theta_o)$ (Übung). Im derart misspezifizierten Modell wird also die beste \mathcal{L}_2 -Approximation γ_o an die wahre Funktion θ_o geschätzt, z.Bsp. die beste Approximation durch ein Polynom vom Grad $\leq d$. \square

§17.09 **Definition.** Eine Folge $(X_n)_{n \in \mathbb{N}}$ numerischer Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ heißt *stochastisch beschränkt* oder *straff*, wenn es für jedes $\delta \in \mathbb{R}_0^+$ ein $K \in \mathbb{N}$ mit $\sup_{n \in \mathbb{N}} \mathbb{P}(|X_n| > K) \leq \delta$ gibt. \square

§17.10 **Landau-O-Schreibweise.** Für eine Folge reeller Zahlen $(x_n)_{n \in \mathbb{N}}$ schreiben wir $x_n = O(1)$ bzw. $x_n = o(1)$, wenn sie beschränkt, also $\sup_{n \in \mathbb{N}} |x_n| < \infty$, ist bzw. $x_n \xrightarrow{n \rightarrow \infty} 0$ gilt. Ist weiterhin $(a_n)_{n \in \mathbb{N}}$ eine Folge in \mathbb{R}_0^+ , so schreiben wir $x_n = O(a_n)$ bzw. $x_n = o(a_n)$ wenn $x_n/a_n = O(1)$ bzw. $x_n/a_n = o(1)$ gilt. Entsprechend für eine Folge $(X_n)_{n \in \mathbb{N}}$ numerischer Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ schreiben wir $X_n = O_p(a_n)$ bzw. $X_n = o_p(a_n)$, wenn die Folge $(X_n/a_n)_{n \in \mathbb{N}}$ stochastisch beschränkt ist bzw. wenn $X_n/a_n \xrightarrow{\mathbb{P}} 0$ gilt. Allgemeiner, ist $(A_n)_{n \in \mathbb{N}}$ eine Folge strikt positiver numerischer Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, dann schreiben wir $X_n = O_p(A_n)$ bzw. $X_n = o_p(A_n)$, wenn die Folge $(X_n/A_n)_{n \in \mathbb{N}}$ stochastisch beschränkt ist bzw. wenn $X_n/A_n \xrightarrow{\mathbb{P}} 0$ gilt. \square

§17.11 **Rechenregeln.** Unter den Voraussetzungen der *Landau-O-Schreibweise* §17.10 gelten:

- (i) für $X_n = o_p(1)$ and $Y_n = o_p(1)$ gilt $X_n + Y_n = o_p(1)$, kurz $o_p(1) + o_p(1) = o_p(1)$;
- (ii) $O_p(1) + o_p(1) = O_p(1)$; (iii) $O_p(1) \cdot o_p(1) = o_p(1)$; (iv) $(1 + o_p(1))^{-1} = O_p(1)$;
- (v) $o_p(A_n) = A_n o_p(1)$; (vi) $O_p(A_n) = A_n O_p(1)$; (vii) für $X_n = o_p(A_n)$ und $A_n = O_p(1)$ gilt $X_n = o_p(1)$, kurz $o_p(O_p) = o_p(1)$. \square

§17.12 **Satz.** Es gelten die Annahmen und Notationen aus *Definition* §17.01. Der Minimum-Kontrast-Schätzer $\hat{\gamma}_n$ ist konsistent für $\gamma(\theta_o)$ unter $\mathbb{P}_\theta^{(n)}$, wenn die folgenden Bedingungen erfüllt sind:

$$(Ko1) \quad \inf_{\gamma \in \Gamma: |\gamma - \gamma_o| \geq \varepsilon} K_\theta(\gamma) > K_\theta(\gamma_o) \text{ für alle } \varepsilon \in \mathbb{R}_0^+;$$

$$(Ko2) \quad \|\hat{K}_n - K_\theta\|_\infty := \sup_{\gamma \in \Gamma} |\hat{K}_n(\gamma) - K(\theta_o, \gamma)| = o_{\mathbb{P}^{(n)}}(1).$$

§17.13 **Beweis** von **Satz** §17.12. In der Vorlesung. \square

§17.14 **Lemma.** Eine Kontrastfunktion K erfüllt (Ko1) in **Satz** §17.12, wenn zusätzlich die folgenden zwei Bedingungen gelten: (i) Γ ist kompakt und (ii) K_θ ist stetig.

§17.15 **Beweis** von **Lemma** §17.14. Übung. \square

§17.16 **Lemma.** Eine Kontrastfunktion K und ein Kontrastprozess $(\hat{K}_n)_{n \in \mathbb{N}}$ erfüllen (Ko2) in **Satz** §17.12, wenn zusätzlich die folgenden drei Bedingungen gelten: (i) Γ ist kompakt; (ii) K_θ ist stetig und (iii) $(\hat{K}_n)_{n \in \mathbb{N}}$ ist **gleichgradig stetig** (in $\mathbb{P}_\theta^{(n)}$ Wahrscheinlichkeit), also

$$\forall \varepsilon \in \mathbb{R}_0^+ : \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}_\theta^{(n)} \left(\sup_{|\gamma_1 - \gamma_2| < \delta} |\hat{K}_n(\gamma_1) - \hat{K}_n(\gamma_2)| \geq \varepsilon \right) = 0.$$

§17.17 **Beweis** von **Lemma** §17.16. In der Vorlesung. \square

§17.18 **Bemerkung.** Beachte, dass $\hat{\gamma}_n$ als Minimum einer fast sicher stetigen Funktion auf einem Kompaktum stets fast sicher existiert. Es kann außerdem messbar gewählt werden (vgl. Witting and Müller-Funk [1995], Satz 6.7). \square

§17.19 **Satz.** Es gelten die Annahmen und Notationen aus *Definition* §17.01. Der Minimum-Kontrast-Schätzer sei konsistent für $\gamma_o := \gamma(\theta_o)$ unter $\mathbb{P}_\theta^{(n)}$ (z.Bsp. unter (Ko1)-(Ko2) von **Satz** §17.12), mit $\Gamma \subseteq \mathbb{R}^k$ und $\gamma_o \in \text{inn}(\Gamma)$. Der Kontrastprozess \hat{K}_n sei $\mathbb{P}_\theta^{(n)}$ -f.ü. zweimal stetig differenzierbar in

einer Umgebung von γ_o , so dass mit

$$U_\gamma^{(n)} := (\hat{K}_n)_\gamma \in \overline{\mathcal{X}}^k \quad (\text{Score}), \quad V_\gamma^{(n)} := (\ddot{K}_n)_\gamma \in \overline{\mathcal{X}}^{(k,k)}$$

die folgenden Konvergenzen unter $\mathbb{P}_o^{(n)}$ gelten:

(aN1) $\sqrt{n} U_{\gamma_o}^{(n)} \xrightarrow{D} N_{(0, \Sigma_o)}$ für ein $\Sigma_o \in \mathbb{R}^{(k,k)}$ positiv semi-definit;

(aN2) für jede Folge $(\tilde{\gamma}_n)_{n \in \mathbb{N}}$ von Statistiken mit $\tilde{\gamma}_n \xrightarrow{\mathbb{P}_o^{(n)}} \gamma_o$ gilt $V^{(n)} \tilde{\gamma}_n \xrightarrow{\mathbb{P}_o^{(n)}} V_o \in \mathbb{R}^{(k,k)}$ regulär.

Dann gilt $\sqrt{n}(\hat{\gamma}_n - \gamma_o) = -V_o^{-1} \sqrt{n} U_{\gamma_o}^{(n)} + o_{\mathbb{P}_o^{(n)}}(1)$ und $\hat{\gamma}_n$ ist unter $\mathbb{P}_o^{(n)}$ asymptotisch normalverteilt:

$$\sqrt{n}(\hat{\gamma}_n - \gamma_o) \xrightarrow{D} N_{(0, V_o^{-1} \Sigma_o V_o^{-1})}.$$

§17.20 **Beweis** von **Satz** §17.19. In der Vorlesung. □

§17.21 **Beispiel** (*Beispiel* §01.06 (b) fortgesetzt). Im Lokations-Modell $(Y_i)_{i \in \llbracket n \rrbracket} \odot \text{IM}_{\mathbb{R} \times \{1\}}^n$ mit unbekanntem Parameter γ , also $Y_i = \gamma + \varepsilon_i, i \in \llbracket n \rrbracket$, mit Fehlertermen $(\varepsilon_i)_{i \in \llbracket n \rrbracket} \odot \text{IM}_{(0,1)}^n$, betrachte einen **M-Schätzer** $\hat{\gamma}_n$ mit

$$\hat{\gamma}_n \in \arg \min_{\gamma \in \mathbb{R}} \sum_{i \in \llbracket n \rrbracket} \rho(Y_i - \gamma)$$

mit einer Funktion $\rho : \mathbb{R} \rightarrow [0, \infty)$, so dass die Funktion $x \mapsto \mathbb{E}(\rho(x + \varepsilon_1))$ minimal (nur) bei $x = 0$ ist. Zu dem Kontrast $\hat{K}_n(\gamma) := \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \rho(Y_i - \gamma)$ erhalten wir dann die Kontrastfunktion $K(\theta_o, \gamma) = \mathbb{E}(\rho(\gamma_o - \gamma + \varepsilon_1))$, wobei wir $\theta_o = (\gamma_o, \mathbb{P}^{\varepsilon_1})$ als unbekanntem Parameter auffassen. Im Fall symmetrisch verteilter Fehlerterme, also $\mathbb{P}^{\varepsilon_1} = \mathbb{P}^{-\varepsilon_1}$, führt als zugehörigen Minimum-Kontrast-Schätzer die Funktion $\rho(x) = \frac{1}{2}x^2$ auf das Stichprobenmittel $\hat{\gamma}_n = \bar{Y}$ und für $\rho(x) = |x|$ auf den Stichprobenmedian $\hat{\gamma}_n$. Ein Kompromiss zwischen beiden Schätzern ist der Huber-Schätzer für $\kappa \in \mathbb{R}_0^+$

$$\hat{\gamma}_n \in \arg \min_{\gamma \in \mathbb{R}} \sum_{i \in \llbracket n \rrbracket} \rho(Y_i - \gamma), \quad \rho(x) = \begin{cases} \frac{1}{2}x^2, & \text{falls } |x| \leq \kappa, \\ \kappa|x| - \frac{\kappa^2}{2}, & \text{falls } |x| > \kappa. \end{cases}$$

Setzt man die Regularitätsannahmen im obigen Satz voraus, so erhält man für den M-Schätzer

$$\sqrt{n}(\hat{\gamma}_n - \gamma_o) \xrightarrow{D} N_{(0, \frac{\mathbb{E}(|\rho(\varepsilon_1)|^2)}{\mathbb{E}\hat{\rho}(\varepsilon_1)^2})}.$$

Im Fall des Stichprobenmittels ist die asymptotische Varianz also gerade $\mathbb{E}(\varepsilon_1^2) = \text{Var}(\varepsilon_1)$. Einsetzen im Fall einer Dichte f^{ε_1} von ε_1 liefert heuristisch für den Stichprobenmedian die asymptotische Varianz $\mathbb{E}(\text{sgn}(\varepsilon_1)^2) / \mathbb{E}(|2\delta_0(\varepsilon_1)|^2) = (4f^{\varepsilon_1}(0))^{-1}$ sowie für den Huber-Schätzer $\mathbb{E}(\varepsilon_1^2 \wedge \kappa^2) / P(|\varepsilon_1| \leq \kappa)^2$. □

§17.22 **Satz**. Es sei $((\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_o^n))_{n \in \mathbb{N}}$ eine Folge μ -dominierter Produktexperimente mit marginaler Loglikelihood-Funktion $\ell(\theta) = \log(d\mathbb{P}_\theta / d\mu)$ und es gelten die Bedingungen:

- (C1) $\Theta \subseteq \mathbb{R}^k$ ist kompakt und θ_o liegt im Innern $\text{inn}(\Theta)$ von Θ .
- (C2) Der Parameter θ ist identifizierbar, also aus $\theta \neq \theta_o$ folgt $\mathbb{P}_\theta \neq \mathbb{P}_{\theta_o}$.
- (C3) Für alle $x \in \mathcal{X}$ ist $\theta \mapsto \ell(\theta, x)$ stetig und $\sup_{\theta \in \Theta} \ell(\theta) \in \mathcal{L}_1(\mathbb{P}_o)$.

Dann sind (Ko1)-(Ko2) von **Satz** §17.12 erfüllt, und somit ist der Maximum-Likelihood-Schätzer $\hat{\theta}_n$ konsistent für θ_o unter \mathbb{P}_o^n . Gelten weiterhin die Bedingungen

- (C4) Für alle $x \in \mathcal{X}$ ist die Abbildung $\theta \mapsto \ell(\theta, x)$ zweimal stetig differenzierbar in einer Umgebung U von θ_0 mit Ableitungen $\dot{\ell}_\theta(x)$ und $\ddot{\ell}_\theta(x)$.
- (C5) Es gilt $\sup_{\theta \in U} \|\dot{\ell}_\theta\| \in \mathcal{L}_2(\mathbb{P}_0)$ und $\sup_{\theta \in U} \|\ddot{\ell}_\theta\| \in \mathcal{L}_1(\mathbb{P}_0)$.
- (C6) Die Fisher-Informationsmatrix (zu einer Beobachtung) $\mathcal{I}_{\theta_0} = \mathbb{E}_{\theta_0}(\dot{\ell}_{\theta_0}^t)$ ist strikt positiv definit.

Dann sind (aN1)-(aN2) von Satz §17.19 erfüllt, und für den Maximum-Likelihood-Schätzer $\hat{\theta}_n$ gilt unter \mathbb{P}_0^n somit

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i \in \llbracket n \rrbracket} \mathcal{I}_{\theta_0}^{-1} \dot{\ell}(\theta_0, X_i) + o_{\mathbb{P}^n}(1).$$

Insbesondere ist $\hat{\theta}_n$ unter \mathbb{P}_0^n asymptotisch normalverteilt mit Rate $n^{-1/2}$ und asymptotischer Kovarianzmatrix $\mathcal{I}_{\theta_0}^{-1}$:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N_{(0, \mathcal{I}_{\theta_0}^{-1})}.$$

Ferner gilt die Identität $\mathcal{I}_{\theta_0} = -\mathbb{E}_{\theta_0}(\ddot{\ell}(\theta_0))$.

§17.23 **Beweis** von Satz §17.22. In der Vorlesung. □

§17.24 **Bemerkung.**

- (i) Die Fisher-Information \mathcal{I}_{θ_0} gibt gerade sowohl die asymptotische Varianz der Score-Funktion als auch die lokale Krümmung der Kontrastfunktion $\text{KL}(\mathbb{P}_\theta | \mathbb{P}_\theta) - \text{KL}(\mathbb{P}_0 | \mathbb{P}_\theta)$ im Minimum θ_0 an.
- (ii) Unter Regularitätsbedingungen gilt für die asymptotische Verteilung des MLS sowohl Unverzerrtheit als auch Cramér-Rao-Effizienz. Es ist aber weder klar noch im Allgemeinen korrekt, dass die Momente ebenfalls konvergieren und dass die Cramér-Rao-Schranke auch asymptotisch gilt.
- (iii) Oft ist Θ nicht kompakt, aber man kann durch eine separate Untersuchung die Konsistenz von $\hat{\theta}_n$ nachweisen. Dann gelten die Konvergenzresultate weiterhin.
- (iv) Die Regularitätsbedingungen lassen sich in natürlicher Weise abschwächen. Es reicht aus, dass \mathbb{P}_0 in θ_0 Hellinger-differenzierbar ist, sowie dass die Loglikelihood-Funktion ℓ in einer Umgebung von θ_0 Lipschitz-stetig in θ ist mit Lipschitz-Konstante in $\mathcal{L}_2(\mathbb{P}_0)$. Einen Beweis unter Verwendung von empirischer Prozesstheorie findet man z.Bsp. in van der Vaart [1998], Satz 5.39.
- (v) Im Fall einer Modellmisspezifikation, in der die zu Grunde liegende Verteilung \mathbb{P}_0 nicht in \mathbb{P}_θ enthalten ist (nicht aber die Annahme unabhängig und identisch \mathbb{P}_0 -verteilter ZV'en verletzt ist), konvergiert der MLS $\hat{\theta}_n$ gegen $\theta^* := \arg \max_{\theta \in \Theta} \int_{\mathcal{X}} \ell(\theta, x) \mathbb{P}_0(dx)$, sofern θ^* existiert und eindeutig ist. Es gilt entsprechend $\theta^* = \arg \min_{\theta \in \Theta} \text{KL}(\mathbb{P}_0 | \mathbb{P}_\theta)$ sofern $\mathbb{P}_0 \ll \mathbb{P}_\theta$. θ^* heißt Kullback-Leibler-Projektion von \mathbb{P}_0 auf \mathbb{P}_θ . Satz §17.19 liefert unter Regularitätsbedingungen die asymptotische Normalität, $\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{D} N_{(0, V^{-1}UV^{-1})}$ mit $U = \mathbb{E}_{\mathbb{P}_0}[\dot{\ell}_{\theta^*}^t]$, $V = \mathbb{E}_{\mathbb{P}_0}[\ddot{\ell}_{\theta^*}]$. Im Allgemeinen wird dabei $U \neq V$ gelten. □

§17.25 **Beispiel.** Bei einer Exponentialfamilie mit natürlichem Parameterraum und natürlicher erschöpfender Statistik S erfüllt der MLS (sofern existent und in $\text{inn}(\Theta)$ enthalten) $\mathbb{E}_{\hat{\theta}(x)}(S) = S(x)$ und die Fisher-Information $\mathcal{I}_\theta = \text{Cov}_\theta(S)$ (Kovarianzmatrix von S). Es gilt unter Regularitätsannahmen $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N_{(0, \text{Cov}_{\theta_0}(S)^{-1})}$ unter \mathbb{P}_0^n . Im Fall einer Bernoullikette

$(X_i)_{i \in \llbracket 1, n \rrbracket} \odot (\mathbb{B}_\pi^n)_{\pi \in (0,1)}$ ist $\theta = \log(\pi/(1-\pi))$ der natürliche Parameter sowie $S(x) = x$ die natürliche erschöpfende Statistik. Aus $\sqrt{n}(\hat{\theta}_n - \theta_o) \xrightarrow{D} N_{(0, \pi(\theta_o)^{-1}(1-\pi(\theta_o))^{-1})}$ unter $\mathbb{B}_{\pi(\theta_o)}^n$ folgt mittels der Δ -Methode für die π -Parametrisierung $\sqrt{n}(\hat{\pi}_n - \pi_o) \xrightarrow{D} N_{(0, \pi_o(1-\pi_o))}$ unter $\mathbb{B}_{\pi_o}^n$. Da $\hat{\pi}_n = \bar{X}$ gilt, ist dies natürlich einfach direkt nachprüfbar. \square

5 Testtheorie

§18 Neyman-Pearson-Theorie

§18.01 **Beispiel.** Ihnen wird folgendes Glückspiel angeboten: sie werfen einmal einen Würfel, im Fall einer „6“ erhalten sie 900,- EUR, anderenfalls zahlen sie 100,- EUR. Das zufällige Experiment sei durch eine Bernoulli, also B_θ -verteilte, Zufallsvariable X beschrieben, wobei der unbekannte Parameter $\theta \in [0, 1]$ die Wahrscheinlichkeit des Auftretens einer „6“ angibt. Eine natürliche Frage ist nun, für welche Parameterwerte lohnt sich das Spiel im Mittel für sie. Betrachte dazu den Gewinn $G := 900X - 100(1 - X)$, so dass für den erwarteten Gewinn gilt $\mathbb{E}_\theta G = 1000\theta - 100$. Der erwartete Gewinn ist offensichtlich positiv, falls $\theta > 1/10 =: \theta_0$ gilt. Unter Verwendung einer Stichprobe $(X_i)_{i \in [n]} \sim (B_\theta^n)_{\theta \in [0,1]}$ möchten Sie nun entscheiden, ob sich das Spiel für sie lohnt, so dass Sie an einer Entscheidungsregel für das Testproblem der Nullhypothese $H_0 : \theta \leq \theta_0$ gegen die Alternative $H_1 : \theta > \theta_0$ interessiert sind. \square

§18.02 **Definition.** Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Modell, $\gamma : \Theta \rightarrow \Gamma$ ein abgeleiteter Parameter und $\{\mathcal{H}^0, \mathcal{H}^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ , also $\mathcal{H}^0 \sqcup \mathcal{H}^1 = \Gamma$ und $\mathcal{H}^0 \neq \emptyset \neq \mathcal{H}^1$. Jede Statistik $\varphi : \mathcal{X} \rightarrow [0, 1]$, also \mathcal{X} - $\mathcal{B}_{[0,1]}$ -messbare Abbildung, heißt *randomisierter Test* der *Nullhypothese* $H_0 : \mathcal{H}^0$ gegen die *Alternative* $H_1 : \mathcal{H}^1$. Die Abbildung $\beta_\varphi : \Theta \rightarrow [0, 1]$ mit $\theta \mapsto \beta_\varphi(\theta) := \mathbb{E}_\theta(\varphi)$ heißt *Gütefunktion* von φ und ihre Werte $\beta_\varphi(\theta)$ werden unter der Alternative, also für $\theta \in \Theta$ mit $\gamma(\theta) \in \mathcal{H}^1$, kurz $\theta \in \gamma^{-1}(\mathcal{H}^1)$, *Macht* von φ genannt. Ein Test φ hält das (*Signifikanz-*) *Niveau* $\alpha \in [0, 1]$ ein (oder kurz ist ein α -Test), falls unter der Nullhypothese, also für jedes $\theta \in \gamma^{-1}(\mathcal{H}^0)$, die Gütefunktion $\beta_\varphi(\theta) \leq \alpha$ erfüllt. Ein α -Test φ heißt *unverfälscht*, wenn seine Macht nicht kleiner als α ist, also $\beta_\varphi(\theta) \geq \alpha$ für alle $\theta \in \gamma^{-1}(\mathcal{H}^1)$ gilt. Ein Test φ_0 der Nullhypothese $H_0 : \mathcal{H}^0$ gegen die Alternative $H_1 : \mathcal{H}^1$ heißt *gleichmäßig bester (unverfälschter) α -Test*, falls φ_0 ein (unverfälschter) α -Test ist und für jedes $\theta \in \gamma^{-1}(\mathcal{H}^1)$ die Macht $\beta_{\varphi_0}(\theta)$ eines jeden anderen (unverfälschten) α -Tests $\tilde{\varphi}$ nicht größer ist, dass heißt, $\beta_{\varphi_0}(\theta) \geq \beta_{\tilde{\varphi}}(\theta)$ gilt. \square

§18.03 **Bemerkung.** Der Wert $\varphi(x)$ eines randomisierten Tests φ wird als bedingte Wahrscheinlichkeit, die Nullhypothese $H_0 : \mathcal{H}^0$ abzulehnen, bei Vorliegen einer Stichprobe x interpretiert (vgl. **Beispiel** §05.15 (b)). Nimmt ein randomisierter Test φ nur die Werte in $\{0, 1\}$ an, so wird er nicht-randomisiert genannt. Im Fall eines nicht-randomisierten Tests φ ergibt sich für eine Stichprobe x damit folgende Entscheidungsregel; lehne die Nullhypothese ab, falls $\varphi(x) = 1$ gilt, anderenfalls, lehne die Nullhypothese nicht ab. Offensichtlich können in dieser Situation nur zwei Fehlentscheidungen auftreten, die Nullhypothese $H_0 : \mathcal{H}^0$ wird abgelehnt, obwohl das statistische Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ für ein $\theta \in \gamma^{-1}(\mathcal{H}^0)$ adäquat ist, kurz die Nullhypothese vorliegt, oder die Nullhypothese wird nicht abgelehnt, obwohl die Alternative $H_1 : \mathcal{H}^1$ vorliegt, also das statistische Experiment $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ für ein $\theta \in \gamma^{-1}(\mathcal{H}^1)$ adäquat ist. Unter der Nullhypothese $H_0 : \mathcal{H}^0$, also für $\theta \in \gamma^{-1}(\mathcal{H}^0)$, wird der Wert der Gütefunktion $\beta_\varphi(\theta) = \mathbb{P}_\theta(\varphi = 1)$, d.h. die Wahrscheinlichkeit die Nullhypothese abzulehnen, Irrtumswahrscheinlichkeit 1. Art oder Fehler 1. Art genannt. Analog unter der Alternative $H_1 : \mathcal{H}^1$, also für $\theta \in \gamma^{-1}(\mathcal{H}^1)$, heißt $1 - \mathbb{E}_\theta(\varphi) = \mathbb{P}_\theta(\varphi = 0)$, also die Wahrscheinlichkeit die Nullhypothese nicht abzulehnen, Irrtumswahrscheinlichkeit 2. Art oder Fehler 2. Art. \square

§18.04 **Beispiel (Beispiel §14.10 fortgesetzt).** Für ein bekanntes $\sigma_0 \in \mathbb{R}_0^+$ betrachte das *normale Lokations-Modell* $(\mathbb{R}^n, \mathcal{B}^n, N_{\mathbb{R} \times \{\sigma_0^2\}}^n)$ mit unbekanntem Parameter $\mu \in \mathbb{R}$. Für ein vorgegebenes $\mu_0 \in \mathbb{R}$ soll

das *einseitige* Testproblem der Nullhypothese $H_0 : \mu \leq \mu_o$ gegen die Alternative $H_1 : \mu > \mu_o$ entschieden werden, das heißt, die Partition $\{\mathcal{H}^0, \mathcal{H}^1\}$ der Parametermenge \mathbb{R} ist gegeben durch $\mathcal{H}^0 = (-\infty, \mu_o]$ und $\mathcal{H}^1 = (\mu_o, \infty)$. Der *einseitige Gauß-Test* beruht auf der Teststatistik $Z : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $x \mapsto Z(x) := \sqrt{n}(\bar{x} - \mu_o)/\sigma_o$, die unter $N_{(\mu_o, \sigma_o^2)}^n$ standardnormalverteilt ist, also $Z \sim N_{(0,1)}$. Zu vorgegebenen Niveau $\alpha \in (0, 1)$ sei z^α das obere α -Fraktile der Standardnormalverteilung, also $N_{(\mu_o, \sigma_o^2)}^n(Z > z^\alpha) = \alpha$. Dann hält der einseitige Gauß-Test $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$ mit $x \mapsto \varphi(x) = \mathbb{1}_{\{Z(x) > z^\alpha\}}$ das Niveau α ein, da nach Konstruktion $N_{(\mu_o, \sigma_o^2)}^n(\varphi = 1) = \alpha$ sowie aus Monotoniegründen $N_{(\mu, \sigma_o^2)}^n(\varphi = 1) < \alpha$ für $\mu < \mu_o$ gilt. \square

§18.05 **Lemma.** Φ und $\tilde{\Phi}$ seien Teilgesamtheiten von Testfunktionen mit $\Phi \subseteq \tilde{\Phi}$. φ_o sei gleichmäßig bester (unverfälschter) Test bzgl. $\tilde{\Phi}$ und es sei $\varphi_o \in \Phi$. Dann ist φ_o auch gleichmäßig bester (unverfälschter) Test bzgl. Φ .

§18.06 **Beweis** von Lemma §18.05. In der Vorlesung. \square

§18.07 **Korollar.** Ist φ_o gleichmäßig bester α -Test für $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$, so ist φ_o auch gleichmäßig bester unverfälschter α -Test für $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$.

§18.08 **Beweis** von Korollar §18.07. In der Vorlesung. \square

§18.09 **Definition.** Seien $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{W}(\mathcal{X})$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{X})$, wobei nicht notwendig $\mathbb{P}_1 \ll \mathbb{P}_0$ gilt. Dann heißt jede positive numerische Zufallsvariable $L \in \overline{\mathcal{X}}^+$ mit

$$\mathbb{P}_0(L < \infty) = 1 \text{ und } \mathbb{P}_1(B) = \int_B L(x) \mathbb{P}_0(dx) + \mathbb{P}_1(B \cap \{L = \infty\}), \quad \forall B \in \mathcal{X} \quad (18.01)$$

ein *Dichtequotient (DQ)*, von \mathbb{P}_1 bzgl. \mathbb{P}_0 .

§18.10 **Bemerkung.** Sei $\mu \in \mathcal{M}_o(\mathcal{X})$ ein \mathbb{P}_0 und \mathbb{P}_1 dominierendes Maß (z.Bsp. $\mu = \mathbb{P}_0 + \mathbb{P}_1$), und bezeichnet f_i eine μ -Dichte von $\mathbb{P}_i, i \in \{0, 1\}$, so ist

$$L_* = \frac{f_1}{f_0} \mathbb{1}_{\mathbb{R}_o^+} + \infty \mathbb{1}_{\{0\}} \quad \mu\text{-f.ü.} \quad (18.02)$$

ein Dichtequotient von \mathbb{P}_1 bzgl. \mathbb{P}_0 . Da $L_* \in \overline{\mathcal{X}}^+$ eine positive numerische Zufallsvariable mit $\mathbb{P}_0(L_* = \infty) = \mathbb{P}_0(f_0 = 0) = 0$ ist, die für alle $B \in \mathcal{X}$ erfüllt

$$\begin{aligned} \int_B L_*(x) \mathbb{P}_0(dx) + \mathbb{P}_1(B \cap \{L_* = \infty\}) &= \int \frac{f_1}{f_0} \mathbb{1}_{B \cap \{f_0 \in \mathbb{R}_o^+\} \cap \{f_1 \in \mathbb{R}_o^+\}} f_0 d\mu + \mathbb{P}_1(B \cap \{f_0 = 0\}) \\ &= \int f_1 \mathbb{1}_{B \cap \{f_0 \neq 0\} \cap \{f_1 \in \mathbb{R}_o^+\}} d\mu + \mathbb{P}_1(B \cap \{f_0 = 0\}) = \mathbb{P}_1(B). \end{aligned}$$

Somit ist L_* stets eine spezielle Festlegung des DQ. Der DQ L von \mathbb{P}_1 bzgl. \mathbb{P}_0 ist durch (18.01) $(\mathbb{P}_0 + \mathbb{P}_1)$ -f.ü. eindeutig bestimmt (Witting [1985] Satz 1.110 a), S. 112). Für eine beliebige Festlegung L des DQ und L_* gemäß (18.02) gilt für jedes $s \in \mathbb{R}^+$ damit

$$\{L > s\} = \{L_* > s\} = \{f_1 > s f_0\} \quad (\mathbb{P}_0 + \mathbb{P}_1)\text{-f.s.} \quad (18.03)$$

sowie $(\mathbb{P}_0 + \mathbb{P}_1)$ -f.s. auch $\{L = s\} = \{L_* = s\} = \{f_1 = s f_0\}$. \square

§18.11 **Lemma.** Es seien $\mathbb{P}_0 \in \mathcal{W}(\mathcal{X})$, $\alpha \in (0, 1)$ und $S \in \overline{\mathcal{X}}$ eine numerische Zufallsvariable mit $\mathbb{P}_0(|S| = \infty) = 0$. Dann existiert ein rechtsseitiger Test der Form

$$\varphi = \mathbb{1}_{\{S > k^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{S = k^\alpha\}} \quad (18.04)$$

mit $\mathbb{E}_0(\varphi) = \alpha$. Dabei lassen sich $k^\alpha \in \mathbb{R}$ und $\zeta_\alpha \in [0, 1]$ explizit angeben, nämlich gemäß

$$k^\alpha := \inf \{s \in \mathbb{R} : \mathbb{P}_0(S > s) \leq \alpha\}, \tag{18.05}$$

$$\zeta_\alpha := \begin{cases} \frac{\alpha - \mathbb{P}_0(S > k^\alpha)}{\mathbb{P}_0(S = k^\alpha)} & \text{für } \mathbb{P}_0(S = k^\alpha) > 0 \\ 0 & \text{für } \mathbb{P}_0(S = k^\alpha) = 0. \end{cases} \tag{18.06}$$

Insbesondere ist der Test φ nicht-randomisiert wählbar, falls gilt $\mathbb{P}_0(S = k^\alpha) = 0$. Ist $S \in \overline{\mathcal{X}}^+$ eine positive numerische Zufallsvariable, so gilt $k^\alpha \in \mathbb{R}^+$.

§18.12 **Beweis** von Lemma §18.11. In der Vorlesung. □

§18.13 **Bemerkung.** Sind in Lemma §18.11 Werte $k^\alpha \in \overline{\mathbb{R}}$ zugelassen, so läßt sich die Festsetzung (18.05) auch für $\alpha = 0$ wählen, bei $\alpha = 1$ setze man $k^\alpha = \sup \{s \in \mathbb{R} : \mathbb{P}_0(S \leq s) = 0\}$. Aus (18.05)-(18.06) folgt weiterhin, wann ein Test der Form (18.04) mit $\mathbb{P}_0(S = k^\alpha) > 0$ nicht-randomisiert gewählt werden kann:

$$\begin{aligned} \varphi(x) = \mathbb{1}_{\{S(x) > k^\alpha\}} &\Leftrightarrow \zeta_\alpha = 0 \Leftrightarrow \exists k^\alpha \in \mathbb{R} : \mathbb{P}_0(S > k^\alpha) = \alpha, \\ \varphi(x) = \mathbb{1}_{\{S(x) \geq k^\alpha\}} &\Leftrightarrow \zeta_\alpha = 1 \Leftrightarrow \exists k^\alpha \in \mathbb{R} : \mathbb{P}_0(S \geq k^\alpha) = \alpha. \end{aligned} \tag{18.07}$$

Betrachten wir einen linksseitigen Test der Form

$$\varphi = \mathbb{1}_{\{S < k_\alpha\}} + \zeta_\alpha \mathbb{1}_{\{S = k_\alpha\}} \tag{18.08}$$

so gilt für $k_\alpha := \sup \{s \in \mathbb{R} : \mathbb{P}_0(S < s) \leq \alpha\}$ und $\zeta_\alpha := \frac{\alpha - \mathbb{P}_0(S < k_\alpha)}{\mathbb{P}_0(S = k_\alpha)}$ für $\mathbb{P}_0(S = k_\alpha) > 0$, sowie $\zeta_\alpha := 0$ für $\mathbb{P}_0(S = k_\alpha) = 0$. Die Werte $k_\alpha = \sup \{s \in \mathbb{R} : \mathbb{P}_0^S([-\infty, s]) = \mathbb{P}_0(S < s) \leq \alpha\}$ bzw. $k_\alpha = \inf \{s \in \mathbb{R} : \mathbb{P}_0^S((s, \infty]) \leq \alpha\}$ werden *unteres* bzw. *oberes α -Fraktile* der Verteilung \mathbb{P}_0^S von S genannt. Das obere α -Fraktile ist der kleinste Wert s , oberhalb dessen höchstens die \mathbb{P}_0^S -Wahrscheinlichkeit α , und es gilt insbesondere $\mathbb{P}_0(S > k_\alpha) \leq \alpha \leq \mathbb{P}_0(S \geq k_\alpha)$. □

§18.14 **Definition.** Sei $(\mathcal{X}, \mathcal{X}, (\mathbb{P}_\theta)_{\theta \in \{0,1\}})$ ein (binäres) statistisches Experiment und L ein Dichtequotient von \mathbb{P}_1 bzgl. \mathbb{P}_0 . Jeder randomisierte Test der Form

$$\varphi := \mathbb{1}_{\{L > k\}} + \zeta \mathbb{1}_{\{L = k\}} \quad (\mathbb{P}_0 + \mathbb{P}_1)\text{-f.ü.} \tag{18.09}$$

für einen *kritischen Wert* $k \in \mathbb{R}^+$ und messbarer Randomisierung $\zeta : \{L = k\} \rightarrow [0, 1]$ wird *Neyman-Pearson-Test* genannt.

§18.15 **Bemerkung.** Da $L \in \overline{\mathcal{X}}^+$ eine numerische Zufallsvariable ist, gilt $\{L > k\}, \{L = k\} \in \mathcal{X}$ nach Eigenschaft A02.06 (i) und φ ist somit \mathcal{X} - $\mathcal{B}_{[0,1]}$ -messbar. Mit anderen Worten ein *Neyman-Pearson-Test* ist in der Tat ein randomisierter Test. Wir halten weiterhin fest, dass sich ein Neyman-Pearson-Test φ der Form (18.09) äquivalent mit μ -Dichten $f_0, f_1 \in \overline{\mathcal{X}}$ von \mathbb{P}_0 und \mathbb{P}_1 bzgl. eines beliebigen domierenden Maes μ formulieren läßt und zwar gemäß

$$\varphi = \mathbb{1}_{\{f_1 > k f_0\}} + \zeta \mathbb{1}_{\{f_1 = k f_0\}} \quad (\mathbb{P}_0 + \mathbb{P}_1)\text{-f.ü.} \tag{18.10}$$

Bezeichnet nämlich L_* die spezielle Festlegung (18.02) des DQ (bei verwenden der μ -Dichten f_0 und f_1), so läßt sich in (18.09) wegen (18.03) $\{L \geq k\}$ durch $\{L_* \geq k\}$ und damit $\{f_1 \geq k f_0\}$ ersetzen, die Gleichheit gilt nach (18.09) $(\mathbb{P}_0 + \mathbb{P}_1)$ -f.ü. □

§18.16 (**Neyman-Pearson-Lemma**). Es seien $(\mathcal{X}, \mathcal{X}, (\mathbb{P})_{\theta \in \{0,1\}})$ ein (binäres) statistisches Experiment, L ein Dichtequotient von \mathbb{P} bzgl. \mathbb{P}_0 und $\Gamma = \{\mathbb{P}_0, \mathbb{P}_1\}$ die Menge der interessierenden Parameter $\gamma(\theta) := \mathbb{P}$, $\theta \in \{0,1\}$. Für das Testproblem der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternative $H_1 : \{\mathbb{P}_1\}$ zu vorgegebenem Niveau $\alpha \in (0,1)$ gilt:

(i) Es existiert ein Neyman-Pearson-Test φ_o der Form (18.09), nämlich

$$\varphi_o := \mathbb{1}_{\{L > k^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{L = k^\alpha\}}, \quad (18.11)$$

wobei das obere α -Fraktile $k^\alpha \in \mathbb{R}^+$ der Verteilung \mathbb{P}_0^L und die konstante Randomisierung $\zeta^\alpha \in [0,1]$ gemäß (18.05) bzw. (18.06) bestimmt ist. Dieser Test schöpft das Signifikanzniveau α aus, also $\mathbb{E}_0(\varphi_o) = \alpha$.

(ii) **Hinreichend und notwendig** für die Optimalität eines α -Tests φ ist:
Es gibt eine Zahl $k \in \mathbb{R}^+$ mit

$$\varphi(x) := \begin{cases} 1 & \text{für } L(x) > k \\ 0 & \text{für } L(x) < k \end{cases} \quad (\mathbb{P}_0 + \mathbb{P}_1)\text{-f.a. } x \in \mathcal{X} \quad (18.12)$$

oder dazu äquivalent für μ -Dichten f_0 und f_1 von \mathbb{P}_0 bzw. \mathbb{P}_1 bzgl. eines beliebigen domierenden Maßes μ gilt

$$\varphi(x) := \begin{cases} 1 & \text{für } f_1(x) > k f_0(x) \\ 0 & \text{für } f_1(x) < k f_0(x) \end{cases} \quad \mu\text{-f.a. } x \in \mathcal{X} \quad (18.13)$$

und es gilt

$$k \in \mathbb{R}_0^+ \quad \text{und} \quad \mathbb{E}_0(\varphi) = \alpha \quad (18.14)$$

oder

$$k = 0 \quad \text{und} \quad \mathbb{E}_0(\varphi) \leq \alpha. \quad (18.15)$$

(iii) Jeder Neyman-Pearson-Test φ der Form (18.09) ist ein bester randomisierter Test zu seinem Niveau $\mathbb{E}_0(\varphi) \in [0,1]$.

§18.17 **Beweis** von Satz §18.16. In der Vorlesung. □

§18.18 **Bemerkung**. Nach Satz §18.16 (ii) ist ein α -Test mit (18.12) und (18.14) oder (18.15) ein bester α -Test unabhängig von seinen Werten auf $\{L = k\}$, so dass die Randomisierung wie in (18.11) konstant gewählt werden kann. Ein Neyman-Pearson-Test φ kann nicht-randomisiert gewählt werden (vgl. Bemerkung §18.13), also $\text{Bild}(\varphi) = \{0,1\}$, wenn der Rand $\{L = k\}$ eine \mathbb{P}_0 Nullmenge ist oder eine der Bedingungen (18.07) erfüllt ist. Die Aussagen des Satz §18.16 für das Signifikanzniveau $\alpha = 0$ oder $\alpha = 1$ lassen sich durch Sonderbetrachtungen beweisen, dabei ist das α -Fraktile wie in Bemerkung §18.13 erklärt. □

§18.19 **Korollar**. Für die Macht $\beta_o := \mathbb{E}_1(\varphi_o)$ eines besten Tests φ_o für $H_0 : \{\mathbb{P}_0\}$ gegen $H_1 : \{\mathbb{P}_1\}$ zu vorgegebenem Niveau $\alpha \in (0,1)$ gilt

- (i) $\beta_o \geq \alpha$,
- (ii) ist der interessierende Parameter $\gamma(\theta) := \mathbb{P}$, $\theta \in \{0,1\}$, identifizierbar, also $\mathbb{P}_0 \neq \mathbb{P}_1$, so gilt $\beta_o > \alpha$,
- (iii) $\mathbb{E}_0(\varphi_o) < \alpha$ impliziert $\beta_o = 1$, also aus $\beta_o < 1$ folgt $\mathbb{E}_0(\varphi_o) = \alpha$ durch Negation,
- (iv) $\mathbb{P}_0 \ll \mathbb{P}_1$ impliziert $\beta_o < 1$ und damit $\mathbb{E}_0(\varphi_o) = \alpha$.

§18.20 **Beweis** von **Korollar** §18.19. In der Vorlesung. □

§18.21 **Bemerkung**. Nach **Satz** §18.16 existiert stets ein bester α -Test, der das zugelassene Niveau ausschöpft, und dass jeder beste Test dies tut, falls die Verteilungen \mathbb{P} und \mathbb{P}_0 äquivalent sind. Die Tatsache, dass unter den α -Tests, also Tests φ mit $\mathbb{E}_0(\varphi) \leq \alpha$, immer einen besten Test φ_0 mit $\mathbb{E}_0(\varphi_0) = \alpha$ gibt, impliziert mit **Lemma** §18.05 die Optimalität von φ_0 unter allen Tests mit $\mathbb{E}_0(\varphi_0) = \alpha$. □

§18.22 **Definition**. Es seien $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein statistisches Experiment mit $\Theta \subseteq \mathbb{R}$, L_{θ_0, θ_1} ein Dichtequotient von \mathbb{P}_{θ_0} bzgl. \mathbb{P}_{θ_1} für alle $\theta_0, \theta_1 \in \Theta$, sowie $S \in \mathcal{X}$ eine reelle Statistik. Die Verteilungsfamilie \mathbb{P}_θ besitzt einen (streng) isotonen Dichtequotienten in S , falls gilt:

(iDQ1) Der Parameter θ ist identifizierbar, also $\theta \neq \theta_0$ impliziert $\mathbb{P}_\theta \neq \mathbb{P}_{\theta_0}$.

(iDQ2) Zu jedem Paar $\theta_0, \theta_1 \in \Theta$ mit $\theta_0 < \theta_1$ existiert eine (streng) isotone Funktion¹ $h_{\theta_0, \theta_1} \in \overline{\mathcal{B}}^+$ mit $L_{\theta_0, \theta_1} = h_{\theta_0, \theta_1}(S)$ ($\mathbb{P}_{\theta_0} + \mathbb{P}_{\theta_1}$)-f.ü.. □

§18.23 **Satz**. Ist \mathbb{P}_θ mit $\Theta \subseteq \mathbb{R}$ eine einparametrische Exponentialfamilie in $\eta(\theta)$ und S , so besitzt sie einen (streng) isotonen Dichtequotienten in S , sofern η (streng) isoton ist.

§18.24 **Beweis** von **Satz** §18.23. In der Vorlesung. □

§18.25 **Beispiel**. In einem Binomialmodell $X \odot (\text{Bin}_{(n, \pi)})_{\pi \in (0, 1)}$ liegt eine Exponentialfamilie in $\eta(\pi) = \log(\pi/(1 - \pi))$ und $S = X$ vor. η wächst streng monoton, so dass diese Familie einen streng isotonen Dichtequotienten in X besitzt. Da

$$L_{\pi_0, \pi}(x) = \frac{\binom{n}{x} \pi^x (1 - \pi)^{n-x}}{\binom{n}{x} \pi_0^x (1 - \pi_0)^{n-x}} = \left(\frac{\pi(1 - \pi_0)}{\pi_0(1 - \pi)} \right)^x \left(\frac{1 - \pi}{1 - \pi_0} \right)^n, \quad x \in \llbracket 0, n \rrbracket$$

folgt die Aussage auch direkt aus der Isotonie in x des Dichtequotienten für $\pi > \pi_0$. □

§18.26 **Beispiel** (Beispiel §18.04 fortgesetzt). Eine normale Lokations-Familie $N_{\mathbb{R} \times \{\sigma_0^2\}}^n$ ist eine einparametrische Exponentialfamilie in $\eta(\mu) = \mu/\sigma_0^2$ und $S(x) = \sum_{i \in \llbracket n \rrbracket} x_i$. η ist streng isoton, so dass diese Familie einen streng isotonen Dichtequotienten in $\sum_{i \in \llbracket n \rrbracket} x_i$ besitzt. Da

$$\begin{aligned} L_{\mu_0, \mu}(x) &= \exp\left(\frac{1}{2\sigma_0^2} \sum_{i \in \llbracket n \rrbracket} \{(x_i - \mu_0)^2 - (x_i - \mu)^2\}\right) \\ &= \exp\left(\frac{(\mu - \mu_0)}{\sigma_0^2} \sum_{i \in \llbracket n \rrbracket} x_i\right) \exp\left(\frac{n}{2\sigma_0^2} (\mu_0^2 - \mu^2)\right), \quad x \in \mathbb{R}^n. \end{aligned}$$

folgt die Aussage auch direkt aus der Isotonie in $\sum_{i \in \llbracket n \rrbracket} x_i$ des Dichtequotienten für $\mu > \mu_0$. □

§18.27 **Satz**. Sei \mathbb{P}_θ mit $\Theta \subseteq \mathbb{R}$ eine Verteilungsfamilie mit isotonem Dichtequotienten in S . Für $\alpha \in (0, 1)$ und $\theta_0 \in \Theta$ gilt dann:

(i) Unter allen Test φ für das einseitige Testproblem $H_0 : \theta \leq \theta_0$ gegen $H_1 : \theta > \theta_0$, die das Signifikanzniveau α ausschöpfen, d.h. $\mathbb{E}_{\theta_0}(\varphi) = \alpha$, existiert ein Test $\varphi_0 := \varphi^0(S)$, der die Irrtumswahrscheinlichkeiten erster und zweiter Art gleichmäßig minimiert, nämlich

$$s \mapsto \varphi^0(s) := \mathbb{1}_{\{s > k^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{s = k^\alpha\}} \quad (18.16)$$

wobei $k^\alpha \in \mathbb{R}$ und $\zeta^\alpha \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, also gemäß

$$\mathbb{E}_{\theta_0}(\varphi_0) = \mathbb{E}_{\theta_0}(\varphi^0) = \mathbb{P}_{\theta_0}(S > k^\alpha) + \zeta^\alpha \mathbb{P}_{\theta_0}(S = k^\alpha) = \alpha. \quad (18.17)$$

¹Jede isotone Funktion $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}^+$ ist Borel-messbar, also $h \in \overline{\mathcal{B}}^+$.

(ii) Der Test φ_o ist ein gleichmäßig bester Test zum Niveau α für das einseitige Testproblem $H_0 : \theta \leq \theta_o$ gegen $H_1 : \theta > \theta_o$.

§18.28 **Beweis** von Satz §18.27. In der Vorlesung. □

§18.29 **Bemerkung.** Die Form des Test (18.16) sowie die Nebenbedingung (18.17) sind invariant gegenüber streng isotonen Transformationen $h : \mathbb{R} \rightarrow \mathbb{R}$ der Prüfgröße S . Besitzt die Verteilungsfamilie \mathbb{P}_θ einen isotonen Dichtequotienten in S , so auch in $\tilde{S} := h(S)$. Wir bestimmen den kritischen Wert $\tilde{k}^\alpha \in \mathbb{R}$ durch (18.05) bzgl. \tilde{S} , also gemäß $\tilde{k}^\alpha = \inf \{s \in \mathbb{R} : \mathbb{P}_\theta(\tilde{S} > s) \leq \alpha\} = \inf \{h(s) \in \mathbb{R} : \mathbb{P}_\theta(\tilde{S} > h(s)) \leq \alpha\} = \inf \{h(s) \in \mathbb{R} : \mathbb{P}_\theta(S > s) \leq \alpha\} = h(k^\alpha)$. Damit erfüllt der Test $\tilde{\varphi}_o := \tilde{\varphi}^o(\tilde{S})$ mit $\tilde{s} \mapsto \tilde{\varphi}^o(\tilde{s}) := \mathbb{1}_{\{\tilde{s} > \tilde{k}^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{\tilde{s} = \tilde{k}^\alpha\}}$ die Nebenbedingung (18.17), also $\mathbb{E}_\theta(\tilde{\varphi}_o) = \mathbb{E}_\theta^{\tilde{S}}(\tilde{\varphi}^o) = \alpha$, und der Test $\tilde{\varphi}_o$ besitzt mit Satz §18.27 die gleichen Optimalitätseigenschaften wie φ_o aus (18.16). Andererseits gelten die Aussagen trivialerweise, da offensichtlich $\tilde{\varphi}_o = \tilde{\varphi}^o(\tilde{S}) = \varphi^o(S) = \varphi_o$ gilt. □

§18.30 **Beispiel** (Beispiel §18.26 fortgesetzt). Im normalen Lokations-Modell ist der einseitige Gauß-Test gegeben in Beispiel §18.04 ein gleichmäßig bester α -Test, da die Normalverteilungsfamilie $N_{\mathbb{R} \times \{\sigma_o^2\}}^n$ einen monotonen Likelihood-Quotienten in $S(x) = \sum_{i \in [n]} x_i$ also auch in $Z(x) = \frac{1}{\sigma_o \sqrt{n}}(S(x) - n\mu_o)$ besitzt. □

§18.31 **Bemerkung.**

- (i) Die Gütefunktion $\beta_{\varphi_o}(\theta) = \mathbb{E}_\theta(\varphi_o)$ ist isoton und zwar streng isoton für alle Werte θ mit $\beta_{\varphi_o}(\theta) \in (0, 1)$ (Übung).
- (ii) Im Beweis wurde eine Konvexkombination $\tilde{\varphi}$ von Tests betrachtet. Dieses Argument lässt sich wie folgt geometrisch veranschaulichen. Zu jedem Paar $\mathbb{P}_\theta, \mathbb{P}_\theta$ ist die Menge $C := \{(\mathbb{E}_\theta(\varphi), \mathbb{E}_\theta(\varphi)) : \varphi \text{ Test}\} \subseteq [0, 1]^2$ konvex (Menge der Tests ist konvex), abgeschlossen (folgt aus dem Satz von Banach-Alaoglu) und enthält die Diagonale (betrachte konstante Tests). Neyman-Pearson-Tests entsprechen dann gerade der oberen Begrenzungskurve von C . □

§18.32 (**Verallgemeinerung des Neyman-Pearson-Lemmas**). Es seien $(\mathcal{X}, \mathcal{X}, (\mathbb{P}_\theta)_{\theta \in \{0,1\}})$ ein (binäres) statistisches Experiment, f_θ Dichte von \mathbb{P}_θ , $\theta \in \{0, 1\}$, bzgl. eines dominierenden Maßes μ , $\Gamma = \{\mathbb{P}_0, \mathbb{P}_1\}$ die Menge der interessierenden Parameter $\gamma(\theta) := \mathbb{P}_\theta$, $\theta \in \{0, 1\}$, sowie $S \in \mathcal{L}_1(\mathbb{P})$ eine reelle Statistik. Ein Test der Form

$$\varphi = \mathbb{1}_{\{f_1 > k^\alpha f_0 + l^\alpha S f_0\}} + \zeta \mathbb{1}_{\{f_1 = k^\alpha f_0 + l^\alpha S f_0\}} \quad (18.18)$$

mit $k^\alpha, l^\alpha \in \mathbb{R}$ und messbarer Randomisierung $\zeta : \{f_1 = k^\alpha f_0 + l^\alpha S f_0\} \rightarrow [0, 1]$, der für $\alpha \in [0, 1]$ die Nebenbedingungen

$$\mathbb{E}_0(\varphi) = \alpha \quad \text{und} \quad \mathbb{E}_0(S\varphi) = \alpha \mathbb{E}_0(S) \quad (18.19)$$

erfüllt, besitzt maximale Macht $\mathbb{E}_1(\varphi)$ in der Menge aller Tests, die (18.19) erfüllen.

§18.33 **Beweis** von Satz §18.32. Übungsaufgabe. □

§18.34 **Bemerkung.** Der Test (18.18) besitzt maximale Macht $\mathbb{E}_1(\varphi)$ in der Menge aller Tests, die (18.19) erfüllen, unabhängig von seinen Werten auf $\{f_1 = k^\alpha f_0 + l^\alpha S f_0\}$, so dass die Randomisierung auch konstant gewählt werden kann. □

§18.35 **Definition.** Ein Test φ heißt α -ähnlich auf $\Theta' \subseteq \Theta$, falls $\mathbb{E}_\theta(\varphi) = \alpha$ für alle $\theta \in \Theta'$ gilt. □

§18.36 **Lemma.** Betrachte das Testproblem $H_0 : \mathcal{H}^0$ gegen $H_1 : \mathcal{H}^1$. Die Parametermenge $\Theta = \mathcal{H}^0 \uplus \mathcal{H}^1$ bilde einen metrischen Raum, $\partial\mathcal{H}^0 \subseteq \mathcal{H}^0$ bezeichne den topologischen Rand zwischen Hypothese und Alternative, und jeder Test besitze eine stetige Gütefunktion in allen $\theta \in \partial\mathcal{H}^0$. Ist φ_o ein α -ähnlicher Test auf $\partial\mathcal{H}^0$, der bester unter allen auf $\partial\mathcal{H}^0$ α -ähnlichen Tests ist, so ist φ_o ein bester unverfälschter α -Test.

§18.37 **Beweis** von Lemma §18.36. In der Vorlesung. □

§18.38 **Satz.** Sei \mathbb{P} eine einparametrische Exponentialfamilie in $\eta(\theta)$ und S . Weiterhin seien $\Theta \subseteq \mathbb{R}$, $\theta_o \in \text{inn}(\Theta)$ und η streng monoton (wachsend oder fallend) und stetig differenzierbar in einer Umgebung von θ_o mit $\dot{\eta}_{\theta_o} \neq 0$. Für $\alpha \in (0, 1)$, $k_1, k_2 \in \mathbb{R}$ mit $k_1 < k_2$ und $\zeta_1, \zeta_2 \in [0, 1]$ erfülle der Test $\varphi_o := \varphi^o(S)$ mit

$$s \mapsto \varphi^o(s) = \mathbb{1}_{\{s < k_1\}} + \zeta_1 \mathbb{1}_{\{s = k_1\}} + \mathbb{1}_{\{s > k_2\}} + \zeta_2 \mathbb{1}_{\{s = k_2\}} \tag{18.20}$$

die Nebenbedingungen

$$\mathbb{E}_{\theta_o}(\varphi_o) = \mathbb{E}_{\theta_o}^S(\varphi^o) = \alpha \quad \text{und} \quad \mathbb{E}_{\theta_o}(S\varphi_o) = \alpha \mathbb{E}_{\theta_o}(S). \tag{18.21}$$

Dann ist φ_o ein *gleichmäßig bester unverfälschter α -Test* für das *zweiseitige Testproblem* $H_0 : \theta = \theta_o$ gegen $H_1 : \theta \neq \theta_o$.

§18.39 **Beweis** von Satz §18.38. In der Vorlesung. □

§18.40 **Bemerkung.** Die Form des Test φ_o (18.20) unter den Nebenbedingungen (18.21) ist invariant gegenüber affinen, streng isotonen Transformationen der Prüfgröße. Besitzt die Verteilungsfamilie \mathbb{P} einen isotonen Dichtequotienten in S , so auch in $\tilde{S} := aS + b$ für $a \in \mathbb{R}_0^+$ und $b \in \mathbb{R}$. Betrachten wir für die kritischen Werte $\tilde{k}_i = ak_i + b$ und $\tilde{\zeta}_i = \zeta_i$, $i \in \{1, 2\}$, den Test $\tilde{\varphi}_o := \tilde{\varphi}^o(\tilde{S})$ mit $\tilde{s} \mapsto \tilde{\varphi}^o(\tilde{s}) = \mathbb{1}_{\{\tilde{s} < \tilde{k}_1\}} + \tilde{\zeta}_1 \mathbb{1}_{\{\tilde{s} = \tilde{k}_1\}} + \mathbb{1}_{\{\tilde{s} > \tilde{k}_2\}} + \tilde{\zeta}_2 \mathbb{1}_{\{\tilde{s} = \tilde{k}_2\}}$, so gilt offensichtlich $\tilde{\varphi}_o = \tilde{\varphi}^o(\tilde{S}) = \varphi^o(S) = \varphi_o$ und $\tilde{\varphi}_o$ erfüllt die Nebenbedingungen $\mathbb{E}_{\theta_o}(\tilde{\varphi}_o) = \mathbb{E}_{\theta_o}(\varphi_o) = \alpha$ sowie $\mathbb{E}_{\theta_o}(\tilde{S}\tilde{\varphi}_o) = \alpha \mathbb{E}_{\theta_o}(\tilde{S})$. In der Tat es gilt $\mathbb{E}_{\theta_o}(\tilde{S}\tilde{\varphi}_o) = a\mathbb{E}_{\theta_o}(S\varphi_o) + b\mathbb{E}_{\theta_o}(\varphi_o) = a\alpha\mathbb{E}_{\theta_o}(S) + b\alpha = \alpha\mathbb{E}_{\theta_o}(\tilde{S})$. Damit besitzt der Test $\tilde{\varphi}_o$ die gleichen Optimalitätseigenschaften wie φ_o aus (18.20). Die Bestimmung von $(\tilde{k}_1, \tilde{\zeta}_1)$ und $(\tilde{k}_2, \tilde{\zeta}_2)$ vereinfacht sich, wenn zusätzlich $\mathbb{P}_{\theta_o}^{\tilde{S}} = \mathbb{P}_{\theta_o}^{-\tilde{S}}$ und $\tilde{S} = h(S) \in \mathcal{L}_1(\mathbb{P}_{\theta_o})$, also $\mathbb{P}_{\theta_o}^{\tilde{S}} \in \mathcal{W}_1(\mathcal{B})$, für eine steng isotone Funktion gilt. Dann ist $\mathbb{E}_{\theta_o}(\tilde{S}) = 0$ und die zweite Bedingung in (18.21) ist für jeden Test $\tilde{\varphi}_o = \tilde{\varphi}^o(\tilde{S})$ mit $\tilde{\varphi}^o(\tilde{S}) = \tilde{\varphi}^o(-\tilde{S})$ erfüllt, also zum Beispiel für eine Test der Form

$$\begin{aligned} \tilde{\varphi}^o(\tilde{s}) &= \mathbb{1}_{\{\tilde{s} < -k^{\alpha/2}\}} + \zeta^{\alpha/2} \mathbb{1}_{\{\tilde{s} = -k^{\alpha/2}\}} + \mathbb{1}_{\{\tilde{s} > k^{\alpha/2}\}} + \zeta^{\alpha/2} \mathbb{1}_{\{\tilde{s} = k^{\alpha/2}\}} \\ &= \mathbb{1}_{\{|\tilde{s}| > k^{\alpha/2}\}} + \zeta^{\alpha/2} \mathbb{1}_{\{|\tilde{s}| = k^{\alpha/2}\}} \end{aligned} \tag{18.22}$$

wobei das obere $\alpha/2$ Fraktile $k^{\alpha/2}$ von $\mathbb{P}_{\theta_o}^{\tilde{S}}$ und $\zeta^{\alpha/2}$ durch (18.05) bzw. (18.06) bestimmt ist, also gemäß $\mathbb{E}_{\theta_o}^{\tilde{S}}(\tilde{\varphi}^o) = \alpha$. □

§18.41 **Beispiel** (Beispiel §18.30 fortgesetzt). Eine normale Lokations-Familie $N_{\mathbb{R} \times \{\sigma_o^2\}}^n$ ist eine einparametrische Exponentialfamilie in $\eta(\mu) = \mu/\sigma_o^2$ und strikt isotonen Dichtequotienten in $S(x) = \sum_{i \in [n]} x_i$, da für $\mu_o \in \mathbb{R}$ gilt $\dot{\eta}(\mu_o) = \sigma_o^{-2} > 0$. Unter Verwendung von Satz §18.38 bestimmen wir einen gleichmäßig besten unverfälschten Test von $H_0 : \mu = \mu_o$ gegen $H_1 : \mu \neq \mu_o$. Betrachten wir den Test φ_o (18.20) mit kritischen Werten k_1, k_2 und Randomisierung ζ_1, ζ_2 unter den Nebenbedingungen (18.21), so gilt unter Verwendung von $Z = \frac{1}{\sigma_o\sqrt{n}}(S - n\mu_o)$ und $\tilde{k}_i = \frac{1}{\sigma_o\sqrt{n}}(k_i - n\mu_o)$, $i \in \{1, 2\}$ gerade $\tilde{\varphi}^o(z) = \mathbb{1}_{\{z < \tilde{k}_1\}} + \zeta_1 \mathbb{1}_{\{z = \tilde{k}_1\}} + \mathbb{1}_{\{z > \tilde{k}_2\}} + \zeta_2 \mathbb{1}_{\{z = \tilde{k}_2\}}$ (vgl.

Bemerkung §18.40). Unter $N_{(\mu, \sigma^2)}^n$ besitzt Z eine symmetrische $N_{(0,1)}$ Verteilung. Wählen wir einen Test der Form (18.22), wobei wir auf Grund der stetigen Verteilung auf eine Randomisierung verzichten, also $\zeta = 0$, sowie $k = z^{\alpha/2}$ das obere $\alpha/2$ Fraktile von $N_{(0,1)}$ ist. Damit lässt sich der Test φ_o in der Form $\varphi_o = \mathbb{1}_{\{Z > z^{\alpha/2}\}} + \mathbb{1}_{\{Z < -z^{\alpha/2}\}} = \mathbb{1}_{\{\frac{\sqrt{n}}{\sigma_o} |\bar{X} - \mu_o| > z^{\alpha/2}\}}$ schreiben. Der **zweiseitige Gauß-Test** φ_o ist also ein gleichmäßig bester unverfälschter α -Test. Die Hypothese im zweiseitigen Testproblem wird also zum Niveau α abgelehnt genau dann, wenn die Hypothese im entsprechenden rechtseitigen oder linkseitigen Testproblem zum Niveau $\alpha/2$ abgelehnt wird (vgl. **Beispiel §18.04**). □

§19 Bedingte Tests

In vielen Fällen sind bestimmte Parameter der Verteilung für einen Test nicht von Interesse (*unwesentlich*), aber sie beeinflussen die Güte eines Tests (sogenannte Störparameter oder *nuisance*-Parameter). Ein wichtiges Beispiel ist der t-Test für Hypothesen über den Mittelwert μ im normalen Lokations-Skalen-Modell $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$. Eine weitere wichtige Klasse bilden Mehrstichprobentestprobleme, in denen das Verhältnis von Kenngrößen (wie Mittelwert) zwischen den Stichproben getestet wird.

§19.01 **Lemma.** Ist T eine vollständige und erschöpfende Statistik für \mathbb{P}_θ , und φ ein auf Θ' α -ähnlicher Test, so gilt $\mathbb{E}_\theta(\varphi|T) = \alpha \mathbb{P}_\theta^T$ -f.ü..

§19.02 **Beweis** von **Lemma §19.01**. In der Vorlesung. □

§19.03 **Vorbemerkung.** Gegeben sei eine natürliche $1 + k$ -parametrische Exponentialfamilie $\mathbb{P}_{\Theta_{\text{nat}}}$ in $(\theta, \tau) \in \Theta_{\text{nat}} \subseteq \mathbb{R} \times \mathbb{R}^k$ und $(S, T) \in \mathcal{X} \otimes \mathcal{X}^k$ mit der Darstellung (10.02). Wir zerlegen die Parametermenge $\Theta_{\text{nat}} = \bigsqcup \{ \{\theta\} \times \Theta_{\text{nat}}^\theta : \theta \in \Theta_\bullet \}$ mit $\Theta_{\text{nat}}^\theta := \{ \tau \in \mathbb{R}^k : (\theta, \tau) \in \Theta_{\text{nat}} \} \subseteq \mathbb{R}^k$ und $\Theta_\bullet := \{ \theta \in \mathbb{R} : \Theta_{\text{nat}}^\theta \neq \emptyset \} \subseteq \mathbb{R}$. Für $\theta_o \in \Theta_\bullet$ betrachten wir die Zerlegung $\Theta_\bullet = \mathcal{H}_{\Theta_{\text{nat}}}^0 \sqcup \mathcal{H}_{\Theta_{\text{nat}}}^1$ mit

$$\mathcal{H}_{\Theta_{\text{nat}}}^0 := \Theta_\bullet \cap (-\infty, \theta_o] \quad \text{und} \quad \mathcal{H}_{\Theta_{\text{nat}}}^1 := \Theta_\bullet \cap (\theta_o, \infty),$$

sowie $\Theta_{\text{nat}} = \mathcal{H}_{\Theta_{\text{nat}}}^0 \sqcup \mathcal{H}_{\Theta_{\text{nat}}}^1$ mit

$$\mathcal{H}_{\Theta_{\text{nat}}}^0 = \bigsqcup \{ \{\theta\} \times \Theta_{\text{nat}}^\theta : \theta \in \mathcal{H}_{\Theta_\bullet}^0 \} \quad \text{und} \quad \mathcal{H}_{\Theta_{\text{nat}}}^1 = \bigsqcup \{ \{\theta\} \times \Theta_{\text{nat}}^\theta : \theta \in \mathcal{H}_{\Theta_\bullet}^1 \}.$$

Wir untersuchen das Testproblem der Nullhypothese $H_0 : \mathcal{H}_{\Theta_{\text{nat}}}^0$ gegen die Alternative $H_1 : \mathcal{H}_{\Theta_{\text{nat}}}^1$, wobei wir auch kurz schreiben $H_0 : \theta \leq \theta_o, \tau \in \Theta_{\text{nat}}^\theta$ gegen $H_0 : \theta > \theta_o, \tau \in \Theta_{\text{nat}}^\theta$. Für ein fest gewähltes aber beliebiges $(\theta^\circ, \tau^\circ) \in \Theta_{\text{nat}}$ bezeichne $\mathbb{P}_\theta := \mathbb{P}_{\theta^\circ, \tau^\circ}$ eine spezielle Verteilung und $\mathbb{P}_\bullet(\bullet|T)$ eine reguläre Festlegung der bedingten Verteilung gegeben T bzgl. \mathbb{P}_θ . Nach **Korollar §10.13 (i)** gibt es für jedes $\theta \in \Theta_\bullet$ eine von τ unabhängige reguläre Festlegung $\mathbb{P}_\bullet(\bullet|T)$ der bedingten Verteilung $\mathbb{P}_{\theta, \tau}(\bullet|T)$, die für \mathbb{P}_θ^T -f.a. $t \in \mathbb{R}^k$ einen isotonen Dichtequotienten in S besitzt. Für ein vorgegebenes $t \in \mathbb{R}^k$ betrachten wir wie in **Satz §18.27** den Test $\varphi_t^\circ(S)$ mit

$$s \mapsto \varphi_t^\circ(s) = \mathbb{1}_{\{s > k^\alpha(t)\}} + \zeta^\alpha(t) \mathbb{1}_{\{s = k^\alpha(t)\}} \tag{19.01}$$

wobei das α -Fraktile $k^\alpha(t) \in \mathbb{R}$ von $\mathbb{P}_{\theta_\bullet}^{S|T=t}$ und $\zeta^\alpha(t) \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, also $\mathbb{E}_{\theta_\bullet}(\varphi_t^\circ(S)|T = t) = \mathbb{E}_{\theta_\bullet}^S(\varphi_t^\circ) = \mathbb{P}_{\theta_\bullet}^{S|T=t}((k^\alpha(t), \infty)) + \zeta^\alpha(t) \mathbb{P}_{\theta_\bullet}^{S|T=t}(\{k^\alpha(t)\}) = \alpha$. Nehmen wir die Erwartung bzgl. $\mathbb{P}_{\theta_\bullet, \tau}^T$ auf beiden Seiten der letzten Gleichung, so erhalten wir, dass der Test $\varphi^\circ := \varphi_t^\circ(S)$ auf dem Rand $\partial \mathcal{H}_{\Theta_{\text{nat}}}^0 = \{\theta_o\} \times \Theta_{\text{nat}} \subseteq \mathcal{H}_{\Theta_{\text{nat}}}^0$ α -ähnlich ist. Wir zeigen

weiterhin, dass für jeden auf $\partial \mathcal{H}_{\Theta_{\text{nat}}}^0$ α -ähnlichen Test φ auch $\mathbb{E}_{\theta, \bullet}(\varphi | T = t) = \alpha$ für \mathbb{P}_\bullet^T -f.a. $t \in \mathbb{R}^k$ gilt. Somit schöpft für \mathbb{P}_\bullet^T -f.a. $t \in \mathbb{R}^k$ der Test φ das Signifikanzniveau α für das Testproblem $H_0 : \mathcal{H}_{\Theta_{\text{nat}}}^0$ gegen $H_1 : \mathcal{H}_{\Theta_{\text{nat}}}^1$ im statistischen Experiment $(\mathcal{X}, \mathcal{X}, (\mathbb{P}_{\theta, \bullet}(\bullet | T = t))_{\theta \in \Theta_{\text{nat}}})$ aus. Unter Verwendung von **Satz §18.27 (i)** sind die Fehler 1. und 2. Art des Tests φ nicht kleiner als die des Tests $\varphi_i^\circ(S)$ gemäß (19.01), also für \mathbb{P}_\bullet^T -f.a. $t \in \mathbb{R}^k$ gilt

$$\begin{aligned} \mathbb{E}_{\theta, \bullet}^{S|T=t}(\varphi_i^\circ) &= \mathbb{E}_{\theta, \bullet}(\varphi_i^\circ(S) | T = t) \leq \mathbb{E}_{\theta, \bullet}(\varphi | T = t) \quad \forall \theta \in \mathcal{H}_{\Theta_{\text{nat}}}^0 \quad \text{und} \\ \mathbb{E}_{\theta, \bullet}^{S|T=t}(\varphi_i^\circ) &= \mathbb{E}_{\theta, \bullet}(\varphi_i^\circ(S) | T = t) \geq \mathbb{E}_{\theta, \bullet}(\varphi | T = t) \quad \forall \theta \in \mathcal{H}_{\Theta_{\text{nat}}}^1. \end{aligned}$$

Nehmen wir erneut die Erwartung bzgl. $\mathbb{P}_{\theta, \tau}^T$ auf beiden Seiten der letzten beiden Ungleichungen, so gilt für $\varphi_o = \varphi_T^\circ(S)$ und jeden anderen auf $\{\theta_o\} \times \Theta_{\text{nat}}^\theta$ α -ähnlichen Test φ

$$\mathbb{E}_{\theta, \tau}(\varphi_o) \leq \mathbb{E}_{\theta, \tau}(\varphi) \quad \forall (\theta, \tau) \in \mathcal{H}_{\Theta_{\text{nat}}}^0 \quad \text{und} \quad \mathbb{E}_{\theta, \tau}(\varphi_o) \geq \mathbb{E}_{\theta, \tau}(\varphi) \quad \forall (\theta, \tau) \in \mathcal{H}_{\Theta_{\text{nat}}}^1.$$

Also ist φ_o ein gleichmäßig bester auf $\partial \mathcal{H}_{\Theta_{\text{nat}}}^0$ α -ähnlicher Test. Ein Vergleich mit dem konstanten Test, sowie die Stetigkeit der Gütefunktion erlauben uns dann, die Aussage von **Satz §19.06** zu zeigen. Vorher halten wir fest, dass mit φ_i° aus (19.01) die Abbildung $x \mapsto \varphi_o(x) = \varphi_{T(x)}^\circ(S(x))$ wie oben angenommen ein Test ist, dazu muss die Abbildung $(s, t) \mapsto \varphi_i^\circ(s)$ und damit die Funktionen $k^\alpha : \mathbb{R}^k \rightarrow \mathbb{R}$ mit $t \mapsto k^\alpha(t)$ sowie $\zeta^\alpha : \mathbb{R}^k \rightarrow [0, 1]$ mit $t \mapsto \zeta^\alpha(t)$ Borel-messbar sein. Dies ist die Aussage von **Lemma §19.04**. □

§19.04 **Lemma.** *Es seien $\mathbb{P}^{S|T}$ eine reguläre bedingte Verteilung von $S \in \mathcal{X}$ gegeben $T \in \mathcal{X}^k$ und $\alpha \in (0, 1)$. Dann sind die folgenden Abbildungen Borel-messbare Funktionen:*

- (i) $(s, t) \mapsto \mathbb{P}^{S|T=t}((-\infty, s])$ und $(s, t) \mapsto \mathbb{P}^{S|T=t}((-\infty, s))$;
- (ii) $t \mapsto k^\alpha(t) := \inf \{s \in \mathbb{R} : \mathbb{P}^{S|T=t}((s, \infty)) \leq \alpha\}$ und $t \mapsto k_\alpha(t) := \sup \{s \in \mathbb{R} : \mathbb{P}^{S|T=t}((-\infty, s)) \leq \alpha\}$;
- (iii) $t \mapsto \zeta^\alpha(t) := \frac{\alpha - \mathbb{P}(S > k^\alpha(t) | T=t)}{\mathbb{P}^{S|T=t}(\{k^\alpha(t)\})} \mathbb{1}_{\{\mathbb{P}^{S|T=t}(\{k^\alpha(t)\}) > 0\}}$ und $t \mapsto \zeta_\alpha(t) := \frac{\alpha - \mathbb{P}(S < k_\alpha(t) | T=t)}{\mathbb{P}^{S|T=t}(\{k_\alpha(t)\})} \mathbb{1}_{\{\mathbb{P}^{S|T=t}(\{k_\alpha(t)\}) > 0\}}$;
- (iv) $(s, t) \mapsto \varphi_t^\circ(s) = \mathbb{1}_{\{s > k^\alpha(t)\}} + \zeta^\alpha(t) \mathbb{1}_{\{s = k^\alpha(t)\}}$ und $(s, t) \mapsto \varphi_t^\circ(s) = \mathbb{1}_{\{s < k_\alpha(t)\}} + \zeta_\alpha(t) \mathbb{1}_{\{s = k_\alpha(t)\}}$.

§19.05 **Beweis von Lemma §19.04.** Witting [1985], Hilfssatz 3.61, S. 376 □

§19.06 **Satz.** *Gegeben sei eine natürliche $1+k$ -parametrische Exponentialfamilie $\mathbb{P}_{\theta, \tau}$ in $(\theta, \tau) \in \Theta_{\text{nat}} \subseteq \mathbb{R} \times \mathbb{R}^k$, $(S, T) \in \mathcal{X} \otimes \mathcal{X}^{1+k}$ mit der Darstellung (10.02) und $\alpha \in (0, 1)$ sowie $\theta_o \in \mathbb{R}$ mit $(\theta_o, \tau) \in \text{inn}(\Theta)$ für ein $\tau \in \mathbb{R}^k$. Nach **Korollar §10.13 (i)** gibt es eine von τ unabhängige Festlegung $\mathbb{P}_{\theta_o, \bullet}^{S|T}$ der regulären bedingten Verteilung $\mathbb{P}_{\theta_o, \tau}^{S|T}$. Dann ist $\varphi_o := \varphi_T^\circ(S)$ mit*

$$(s, t) \mapsto \varphi_i^\circ(s) = \mathbb{1}_{\{s > k^\alpha(t)\}} + \zeta^\alpha(t) \mathbb{1}_{\{s = k^\alpha(t)\}}, \tag{19.02}$$

wobei für $t \in \mathbb{R}^k$ das obere α -Fraktile $k^\alpha(t) \in \mathbb{R}$ von $\mathbb{P}_{\theta_o, \bullet}^{S|T=t}$ und $\zeta^\alpha(t) \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, also $\mathbb{E}_{\theta_o, \bullet}^{S|T=t}(\varphi_i^\circ) = \alpha$, ein gleichmäßig bester unverfälschter α -Test für das Testproblem der Nullhypothese $H_0 : \theta \leq \theta_o, \tau \in \Theta_{\text{nat}}^\theta$ gegen die Alternative $H_1 : \theta > \theta_o, \tau \in \Theta_{\text{nat}}^\theta$.

§19.07 **Beweis von Satz §19.06.** In der Vorlesung. □

§19.08 **Beispiel.** Für $n, m \in \mathbb{N}$ seien $X \odot (\text{Poi}_a^n)_{a \in \mathbb{R}_0^+}$ und $Y \odot (\text{Poi}_b^m)_{b \in \mathbb{R}_0^+}$ unabhängig. Es soll die Hypothese $H_0 : a \leq b$ gegen die Alternative $H_1 : a > b$ getestet werden. Für $(a, b) \in (\mathbb{R}_0^+)^2$ bezeichnen wir mit $\text{Poi}_{a,b}^{n,m} := \text{Poi}_a^n \otimes \text{Poi}_b^m$ die gemeinsame Verteilung von $X \sim \text{Poi}_a^n$ und

$Y \sim \text{Poi}_b^m$. Dann ist $(\text{Poi}_{a,b}^{n,m})_{(a,b) \in (\mathbb{R}_0^+)^2}$ eine 2-parametrische Exponentialfamilie bezüglich des Zählmaßes μ auf \mathbb{N}_0^{m+n} , da für $x \in \mathbb{N}_0^n$ und $y \in \mathbb{N}_0^m$ gilt

$$\frac{d\text{Poi}_{a,b}^{n,m}}{d\mu}(x, y) = \frac{e^{-na-mb}}{(\prod_i x_i!)(\prod_j y_j!)} \exp\left(\log(a/b) \sum_{i \in [n]} x_i + \log(b) \left(\sum_{i \in [n]} x_i + \sum_{j \in [m]} y_j\right)\right).$$

Absorbieren wir die Funktion $h(x, y) = ((\prod_i x_i!)(\prod_j y_j!))^{-1}$ im dominierenden Maß $\nu := h\mu$, so liegt eine 2-parametrische natürliche Exponentialfamilie \mathbb{P}_{nat} mit der Darstellung (10.02) in $S(x, y) = \sum_{i \in [n]} x_i$, $T(x, y) = \sum_{i \in [n]} x_i + \sum_{j \in [m]} y_j$ mit $\theta = \log(a/b)$ und $\tau = \log(b)$ vor. Wir können also das reparametrisierte Testproblem $H_0 : \theta \leq 0, \tau \in \mathbb{R}$ gegen $H_1 : \theta > 0, \tau \in \mathbb{R}$ bei Vorliegen einer Beobachtung der erschöpfenden Statistiken S und T betrachten. Nach **Satz §19.06** ist $\varphi_\circ := \varphi_\tau^\circ(S)$ ein gleichmäßig bester unverfälschter Test $\varphi_\circ = \varphi_\tau^\circ(S)$ mit

$$(s, t) \mapsto \varphi_t^\circ(s) = \mathbb{1}_{\{s > k^\alpha(t)\}} + \zeta^\alpha(t) \mathbb{1}_{\{s = k^\alpha(t)\}},$$

wobei für $t \in \mathbb{R}^k$ das obere α -Fraktile $k^\alpha(t) \in \mathbb{R}$ von $\mathbb{P}_{0,\bullet}^{S|T=t}$ und $\zeta^\alpha(t) \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist. Es bleibt eine Festlegung der regulären bedingten Verteilung von S gegeben T bzgl. $\mathbb{P}_{0,\tau}$ zu bestimmen, die nicht von $\tau \in \mathbb{R}$ abhängt. Zur Erinnerung, für $(X, Y) \sim \text{Poi}_a^n \otimes \text{Poi}_b^m$ sind $S \sim \text{Poi}_{na}$ und $T - S \sim \text{Poi}_{mb}$ unabhängig, sowie $T \sim \text{Poi}_{na+bm}$. Für $t \in \mathbb{N}$ und $s \in \mathbb{N}_0$ mit $s \leq t$ gilt

$$\begin{aligned} \text{Poi}_{a,b}^{n,m}(S = s | T = t) &= \frac{\text{Poi}_{a,b}^{n,m}(S = s, T - S = t - s)}{\text{Poi}_{a,b}^{n,m}(T = t)} = \frac{\frac{(an)^s}{s!} e^{-an} \frac{(bm)^{t-s}}{(t-s)!} e^{-bm}}{\frac{(an+bm)^t}{t!} e^{-an-bm}} \\ &= \frac{\binom{t}{s} (an)^s (bm)^{t-s}}{\sum_{i=0}^t \binom{t}{i} (an)^i (bm)^{t-i}} = \binom{t}{s} \left(\frac{an}{an+bm}\right)^s \left(\frac{bm}{an+bm}\right)^{t-s} \\ &= \binom{t}{s} \left(\frac{\rho}{1+\rho}\right)^s \left(\frac{1}{1+\rho}\right)^{t-s} \quad \text{mit } \rho = \frac{na}{mb}. \end{aligned}$$

In dem Fall $t = 0$, also $T = 0$, ist natürlich $S = 0$, so dass $\text{Poi}_{a,b}^{n,m}(S = s | T = 0) = \delta_0(s)$, $s \in \mathbb{N}_0$, gilt. Für $\pi := \frac{na}{mb} / (1 + \frac{na}{mb})$ erhalten wir $\text{Poi}_{a,b}^{n,m}(S = s | T = t) = \text{Bin}_{(t,\pi)}(\{s\})$ für alle $t \in \mathbb{N}_0$, wobei $\text{Bin}_{(0,\pi)} := \delta_0$. Damit hängt für alle $\theta = \log(a/b) \in \mathbb{R}$ und $\tau = \log(b) \in \mathbb{R}$ diese Festlegung der bedingten Verteilung von S gegeben T bzgl. $\mathbb{P}_{\theta,\tau} = \text{Poi}_{a,b}^{n,m}$ in der Tat nur von θ und nicht von τ ab. Im Fall $\theta = 0$, also für alle $a = b \in \mathbb{R}_0^+$, vereinfacht sich diese für alle $s \in \mathbb{N}_0$ zu

$$\mathbb{P}_{0,\bullet}^{S|T=t}(\{s\}) = \mathbb{P}_{0,\tau}^{S|T=t}(\{s\}) = \text{Poi}_{b,b}^{n,m}(S = s | T = t) = \text{Bin}_{(t,\pi_\circ)}(s) \quad \text{mit } \pi_\circ = \frac{n}{n+m}.$$

Für $t \in \mathbb{N}_0$ sei also gemäß (18.05) $k^\alpha(t) = \text{Bin}_{(t,\pi_\circ)}^\alpha$ das obere α -Fraktile einer $\text{Bin}_{(t,\pi_\circ)}$ -Verteilung und $\zeta^\alpha(t) \in [0, 1]$ wie in (18.06), so dass der Test $\varphi_t^\circ(S)$ mit

$$s \mapsto \varphi_t^\circ(s) = \mathbb{1}_{\{s > \text{Bin}_{(t,\pi_\circ)}^\alpha\}} + \zeta^\alpha(t) \mathbb{1}_{\{s = \text{Bin}_{(t,\pi_\circ)}^\alpha\}}$$

für $t \in \mathbb{N}_0$ dass Niveau α ausschöpft, also $\mathbb{E}_{0,\bullet}^{S|T=t}(\varphi_t^\circ) = \mathbb{E}_{0,\bullet}(\varphi_t^\circ(S) | T = t) = \alpha$ gilt. Im Fall $t = 0$ ist $\text{Bin}_{(0,\pi_\circ)}^\alpha = 0$ und $\zeta^\alpha(0) = \alpha$, also $\varphi_0^\circ(S)$ mit $s \mapsto \varphi_0^\circ(s) = \alpha$ der konstante Test. Damit ist der Test $\varphi_\circ = \varphi_\tau^\circ(S)$ gegeben durch

$$(x, y) \mapsto \varphi_\circ(x, y) = \mathbb{1}_{\{n\bar{x} > \text{Bin}_{(n\bar{x}+m\bar{y}, n/[n+m])}^\alpha\}} + \zeta^\alpha(n\bar{x} + m\bar{y}) \mathbb{1}_{\{n\bar{x} = \text{Bin}_{(n\bar{x}+m\bar{y}, n/[n+m])}^\alpha\}}$$

ein gleichmäßig bester unverfälschter Test von $H_0 : a \leq b$ gegen $H_1 : a > b$ zum Niveau $\alpha \in (0, 1)$. Der kritische Wert als auch die Randomisierung hängen aber von $n\bar{x} + m\bar{y}$ ab, und müssen also bei der Anwendung des Test φ_\circ für jede Stichprobe neu berechnet werden \square

§19.09 **Beispiel.** Wir bestimmen einen gleichmäßigen besten unverfälschten Test für den Mittelwert einer *normalen Lokations-Skalen-Familie* $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ für das einseitige Testproblem $H_0 : \mu \leq 0$, $\sigma \in \mathbb{R}_0^+$ gegen $H_1 : \mu > 0, \sigma \in \mathbb{R}_0^+$ unter Verwendung des **Satz §19.06** (für allgemeines $H_0 : \mu \leq \mu_0, \sigma \in \mathbb{R}_0^+$ gegen $H_1 : \mu > \mu_0, \sigma \in \mathbb{R}_0^+$ verschiebe die Beobachtungen entsprechend, $\tilde{X}_i = X_i - \mu_0$). Die Beobachtungen folgen einer zweiparametrischen natürlichen Exponentialfamilie in $S(x) = \bar{x}$, $T(x) = \frac{1}{n} \sum_{i \in [n]} x_i^2$ mit $\theta = n\mu/\sigma^2$, $\tau = -n/(2\sigma^2)$. Wir können also das reparametrisierte Testproblem $H_0 : \theta \leq 0, \tau \in \mathbb{R}_0^-$ gegen $H_1 : \theta > 0, \tau \in \mathbb{R}_0^-$ bei Vorliegen der Beobachtung der erschöpfenden Statistiken S und T betrachten. Nach **Satz §19.06** ist $\varphi_0 = \mathbb{1}_{\{S > k^\alpha(T)\}}$ ein gleichmäßig bester unverfälschter Test, wobei wir auf Grund der stetigen Verteilung auf eine Randomisierung verzichten und gemäß (18.05) für jedes $t \in \mathbb{R}$ das obere α -Fraktile $k^\alpha(t) \in \mathbb{R}$ einer von τ unabhängigen regulären Version $\mathbb{P}_{\theta, \tau}^{S|T=t}$ der bedingten Verteilung von S gegeben $T = t$ zu bestimmen ist. \square

§19.10 **Bemerkung.** Die letzte Aussage zeigt, dass die reguläre bedingte Verteilung $\mathbb{P}_{\theta, \tau}^{S|T}$ nicht von dem Störparameter τ abhängt, dieser also für die Wahl von $k^\alpha(t) \in \mathbb{R}$ und $\zeta^\alpha(t) \in [0, 1]$ unerheblich ist. Der kritische Wert als auch die Randomisierung hängen aber von t ab, und müssen also bei der Anwendung des Test φ_0 für jede Stichprobe neu berechnet werden. Der Test wäre nicht-bedingt wählbar, wenn es eine von t unabhängige Festlegung der Prüfverteilung $\mathbb{P}_{\theta, \tau}^{S|T=t}$ gäbe. Die nächste Aussage gibt eine hinreichende und notwendige Bedingung an, so dass die bedingte Verteilung unabhängig von t wählbar ist. Diese benötigt die Suffizienz und Vollständigkeit der Statistik T für τ , nicht jedoch die Eigenschaft einer Exponentialfamilie. Dieses ist wesentlich, da nur die erste, nicht dagegen die zweite dieser Eigenschaften unter gewissen Transformationen erhalten bleibt. Die nächste Aussage folgt aus **Satz §09.22** und dem **Lemma von Basu §11.03**. \square

§19.11 **Korollar.** Es seien $T \in \mathcal{X}^k$ eine erschöpfende und vollständige Statistik für die Verteilungsfamilie $\mathbb{P}_{\{\theta, \tau\} \times \Theta^\theta}$ auf $(\mathcal{X}, \mathcal{X})$ und $S \in \mathcal{X}$ eine weitere Statistik. Dann sind äquivalent:

(i) Die Statistik S ist unwesentlich, ihre Verteilung $\mathbb{P}_{\theta, \tau}^S = \mathbb{P}_{\theta, \tau}^S$ hängt nicht von $\tau \in \Theta^\theta$ ab.

(ii) Die bedingte Verteilung $\mathbb{P}_{\theta, \tau}^{S|T=t}$ ist unabhängig von t wählbar.

Jede dieser Aussage impliziert die Gültigkeit der folgenden beiden Aussagen:

(iii) S und T sind stochastisch unabhängig unter jedem $\mathbb{P}_{\theta, \tau}$, $\tau \in \Theta^\theta$.

(iv) Für alle $\alpha \in (0, 1)$ ist der Test aus (19.02) als nicht-bedingter Test wählbar.

§19.12 **Beweis von Korollar §19.11.** In der Vorlesung. \square

§19.13 **Lemma.** Es seien \mathbb{P}_θ eine dominierte Verteilungsfamilie auf $(\mathcal{X}, \mathcal{X})$, $T \in \mathcal{X}^k$, $S \in \mathcal{X}$, $g : \mathbb{R}^{1+k} \rightarrow \mathbb{R}^{1+k}$ mit $(s, t) \mapsto g(s, t) := (h_t(s), t)$ eine Borel-messbare Abbildung, $\tilde{S} := h_T(S) \in \mathcal{X}$ und $(\tilde{S}, T) = g(S, T) \in \mathcal{X}^{1+k}$. Dann gilt: Ist T erschöpfend und vollständig für $\mathbb{P}_\theta^{(S, T)}$, so auch für $\mathbb{P}_\theta^{(\tilde{S}, T)}$.

§19.14 **Beweis von Lemma §19.13.** In der Vorlesung. \square

§19.15 **Bemerkung.** Der einseitige Test aus (19.01) ist invariant gegenüber Transformationen $s \mapsto \tilde{s} := h_t(s)$ bei festem t , falls h_t streng isoton ist (vgl. **Bemerkung §18.29**). Dann ist nämlich $s > k^\alpha(t)$ äquivalent mit $\tilde{s} = h_t(s) > h_t(k^\alpha(t)) =: \tilde{k}^\alpha(t)$, und analog $s < k^\alpha(t)$ bzw. $s = k^\alpha(t)$ mit $\tilde{s} < \tilde{k}^\alpha(t)$ bzw. $\tilde{s} = \tilde{k}^\alpha(t)$. Die Möglichkeit den Test φ_0 aus (19.02) nicht-bedingt zu wählen, beruht nun darauf, eine streng isotone Transformation h_t zu finden, derart dass die Verteilung von $\tilde{S} = h_T(S)$ auf $(\mathcal{X}, \mathcal{X}, \mathbb{P}_{\{\theta, \tau\} \times \Theta^\theta})$ unwesentlich ist. \square

§19.16 **Korollar.** Zusätzlich zu den Annahmen von **Satz §19.06** sei $(s, t) \mapsto g(s, t) := (h_t(s), t)$ eine Borel-messbare Abbildung und $\tilde{S} = h_T(S)$, derart dass die Verteilung $\mathbb{P}_{\theta, \tau}^{\tilde{S}} = \mathbb{P}_{\theta, \tau}^{\tilde{S}}$ von \tilde{S} unabhän-

gig von $\tau \in \Theta_{\text{nat}}^\theta$, also \tilde{S} unwesentlich für $\mathbb{P}_{\{\theta_o\} \times \Theta_{\text{nat}}^\theta}$ ist. Existiert eine $\mathbb{P}_{\Theta_{\text{nat}}^{(S,T)}}$ Einsmenge \mathcal{E} , so dass auf \mathcal{E} die Abbildung $s \mapsto h_t(s)$ streng isoton ist, so ist der Test

$$\tilde{\varphi}_o = \mathbb{1}_{\{\tilde{S} > \tilde{k}^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{\tilde{S} = \tilde{k}^\alpha\}}, \tag{19.03}$$

wobei das obere α -Fraktile $\tilde{k}^\alpha \in \mathbb{R}$ von $\mathbb{P}_{\Theta_{\text{nat}}^{\tilde{S}}}$ und $\zeta^\alpha \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, ein gleichmäßig bester unverfälschter α -Test für das Testproblem der Nullhypothese $H_0 : \theta \leq \theta_o, \tau \in \Theta_{\text{nat}}^\theta$ gegen die Alternative $H_1 : \theta > \theta_o, \tau \in \Theta_{\text{nat}}^\theta$.

§19.17 **Beweis** von **Korollar** §19.16. In der Vorlesung. □

§19.18 **Beispiel** (Beispiel §19.09 fortgesetzt). Sei $(X_i)_{i \in [n]} \odot N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, so besitzt $\hat{T}_n := \frac{\sqrt{n}}{\hat{S}_n} \bar{X}_n$ mit $\bar{X}_n = \frac{1}{n} \sum_{i \in [n]} X_i$, $\hat{S}_n := (\hat{S}_n^{(2)})^{1/2}$ und $\hat{S}_n^{(2)} := \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X}_n)^2$ unter $N_{(0, \sigma^2)}^n$ für jedes $\sigma \in \mathbb{R}_0^+$ eine Student- t_{n-1} -Verteilung mit $n-1$ Freiheitsgraden, kurz $\hat{T}_n \sim t_{n-1}$ (vgl. **Eigenschaft** A05.09 (iii)). Für den gleichmäßig besten unverfälschten Test $\varphi_o = \mathbb{1}_{\{S > k^\alpha(T)\}}$ mit $S = \bar{X}_n$ und $T = \frac{1}{n} \sum_{i \in [n]} X_i^2$ im **Beispiel** §19.09 blieb für jedes $t \in \mathbb{R}$ die Bestimmung des oberen α -Fraktiles $k^\alpha(t) \in \mathbb{R}$ einer von τ unabhängigen regulären Version $\mathbb{P}_{\Theta_{\text{nat}}^{S|T=t}}$ der bedingten Verteilung von S gegeben $T = t$. Wir halten fest, dass $T(x) - (S(x))^2 = \frac{1}{n} \sum_{i \in [n]} (x_i - \bar{x})^2 \in \mathbb{R}_0^+$ für alle $x \in \text{Bild}(\mathbb{1}_n)^c$ gilt, wobei $\text{Bild}(\mathbb{1}_n)^c$ eine $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ -Einsmenge ist. Da $\hat{S}_n^{(2)} = \frac{n}{n-1}(T - S^2)$ gilt, ist $\hat{T}_n = \frac{S\sqrt{n-1}}{\sqrt{T-S^2}} =: h_T(S)$ auf $\text{Bild}(\mathbb{1}_n)^c$, also $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ -f.ü., erklärt. Weiterhin gilt $0 \leq S^2 < T$ auf $\text{Bild}(\mathbb{1}_n)^c$ und für alle $t \in \mathbb{R}_0^+$ ist die Funktion $s \mapsto h_t(s)$ auf $(-t^{1/2}, t^{1/2})$ ungerade und streng isoton. Da T vollständig und suffizient für die Verteilungsfamilie $\mathbb{P}_{\{\theta_o\} \times \Theta_{\text{nat}}^\theta}$ ist und die Verteilung von \hat{T}_n , nämlich $\hat{T}_n \sim t_{n-1}$, unter $\mathbb{P}_{\Theta_{\text{nat}}^\tau}$, nicht von τ abhängt, sind \hat{T}_n und T unabhängig unter jedem $\mathbb{P}_{\Theta_{\text{nat}}^\tau}$ (**Korollar** §19.11). Bezeichnen wir mit t_{n-1}^α das obere α -Fraktile einer t_{n-1} -Verteilung, wobei wir auf die Randomisierung auf Grund der stetigen Verteilung von \hat{T}_n verzichten, so sind der ursprüngliche gleichmäßig beste unverfälschte α -Test φ_o (19.02) und der Test $\tilde{\varphi}_o = \mathbb{1}_{\{\hat{T}_n > t_{n-1}^\alpha\}}$ $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ -f.ü. äquivalent. Zusammenfassend, für jedes $\mu_o \in \mathbb{R}$ ist der **einseitige t-Test** $\varphi = \mathbb{1}_{\{\bar{X} - \mu_o > n^{-1/2} \hat{S}_n t_{n-1}^\alpha\}}$ also ein gleichmäßig bester unverfälschter Test von $H_0 : \mu \leq \mu_o, \sigma \in \mathbb{R}_0^+$ gegen $H_1 : \mu > \mu_o, \sigma \in \mathbb{R}_0^+$. □

§19.19 **Satz**. Es gelten die Voraussetzungen des **Satz** §19.06. Der Test $\varphi_o := \varphi_T^\circ(S)$ mit

$$(s, t) \mapsto \varphi_i^\circ(s) = \mathbb{1}_{\{s < k_1^\alpha(t)\}} + \zeta_1^\alpha(t) \mathbb{1}_{\{s = k_1^\alpha(t)\}} + \mathbb{1}_{\{s > k_2^\alpha(t)\}} + \zeta_2^\alpha(t) \mathbb{1}_{\{s = k_2^\alpha(t)\}}, \tag{19.04}$$

wobei für $\alpha \in (0, 1)$ und für jedes $t \in \mathbb{R}^k$ die kritischen Werte $k_1^\alpha(t), k_2^\alpha(t) \in \mathbb{R}$ und die Randomisierung $\zeta_1^\alpha(t), \zeta_2^\alpha(t) \in [0, 1]$ gemäß der Nebenbedingungen

$$\begin{aligned} \mathbb{E}_{\Theta_{\text{nat}}^{S|T=t}}(\varphi_i^\circ) &= \mathbb{E}_{\Theta_{\text{nat}}^\bullet}(\varphi_i^\circ(S)|T=t) = \alpha \quad \text{und} \\ \mathbb{E}_{\Theta_{\text{nat}}^\bullet}(S\varphi_i^\circ(S)|T=t) &= \alpha \mathbb{E}_{\Theta_{\text{nat}}^\bullet}(S|T=t) \end{aligned} \tag{19.05}$$

bestimmt sind, ein gleichmäßig bester unverfälschter α -Test für das Testproblem der Nullhypothese $H_0 : \theta = \theta_o, \tau \in \Theta_{\text{nat}}^\theta$ gegen die Alternative $H_1 : \theta \neq \theta_o, \tau \in \Theta_{\text{nat}}^\theta$.

§19.20 **Beweis** von **Satz** §19.19. In der Vorlesung. □

§19.21 **Bemerkung**. Der zweiseitige Test aus (19.04) ist invariant gegenüber affinen, streng isotonen Transformationen (vgl. **Bemerkung** §18.40). Dies gibt uns erneut die Möglichkeit, den Test nicht-bedingt zu wählen. Eine explizite Angabe der kritischen Werte $k_1^\alpha(t), k_2^\alpha(t) \in \mathbb{R}$ die Lösung der Nebenbedingungen (19.05) sind, ist im Allgemeinen nicht möglich. Im Spezialfall einer symmetrischen bedingten Verteilung $\mathbb{E}_{\Theta_{\text{nat}}^{S|T=t}}$ können wir wie in **Bemerkung** §18.40 die kritischen Werte $k_1^\alpha(t), k_2^\alpha(t) \in \mathbb{R}$ als unteres bzw. oberes $\alpha/2$ Fraktile von $\mathbb{E}_{\Theta_{\text{nat}}^{S|T=t}}$ wählen. □

§19.22 **Korollar.** Zusätzlich zu den Annahmen von Satz §19.06 sei $(s, t) \mapsto g(s, t) := (h_t(s), t)$ eine Borel-messbare Abbildung und $\tilde{S} = h_T(S)$, derart dass die Verteilung $\mathbb{P}_{\theta, \tau}^{\tilde{S}} = \mathbb{P}_{\theta, \tau}^{\tilde{S}} \in \mathcal{W}_1(\mathcal{B})$ von \tilde{S} symmetrisch und unabhängig von $\tau \in \Theta_{\text{nat}}^\theta$, also \tilde{S} unwesentlich für $\mathbb{P}_{\{\theta, \tau\} \times \Theta_{\text{nat}}^\theta}$ ist. Existiert eine $\mathbb{P}_{\Theta_{\text{nat}}}^{(S, T)}$ Einsmenge \mathcal{E} , so dass auf \mathcal{E} die Abbildung $s \mapsto h_t(s)$ streng isoton ist, so ist der Test

$$\tilde{\varphi}_\circ = \mathbb{1}_{\{|\tilde{S}| > \tilde{k}^{\alpha/2}\}} + \zeta^{\alpha/2} \mathbb{1}_{\{|\tilde{S}| = \tilde{k}^{\alpha/2}\}}, \tag{19.06}$$

wobei das obere $\alpha/2$ -Fraktile $\tilde{k}^{\alpha/2} \in \mathbb{R}$ von $\mathbb{P}_{\theta, \tau}^{\tilde{S}}$ und $\zeta^{\alpha/2} \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, ein gleichmäßig bester unverfälschter α -Test für das Testproblem der Nullhypothese $H_0 : \theta = \theta_\circ, \tau \in \Theta_{\text{nat}}^\theta$ gegen die Alternative $H_1 : \theta \neq \theta_\circ, \tau \in \Theta_{\text{nat}}^\theta$.

§19.23 **Beweis** von Korollar §19.22. In der Vorlesung. □

§19.24 **Beispiel** (Beispiel §19.18 fortgesetzt). Sei $(X_i)_{i \in [1, n]} \odot N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, so besitzt $\hat{T}_n = \frac{\sqrt{n}}{\hat{S}_n} \bar{X}_n = h_T(S)$ unter $N_{(0, \sigma^2)}^n$ für alle $\sigma \in \mathbb{R}_0^+$ eine Student- t_{n-1} -Verteilung. Die Abbildung $s \mapsto h_t(s)$ ist ungerade und streng isoton auf dem Bild von $\text{Bild}(\mathbb{1}_n)^c$ unter (S, T) , wobei $\text{Bild}(\mathbb{1}_n)^c$ eine $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ -Einsmenge ist. Weiterhin ist T vollständig und suffizient für die Verteilungsfamilie $\mathbb{P}_{\{0\} \times \Theta^\circ}$ und die Verteilung von \hat{T}_n , nämlich $\hat{T}_n \sim t_{n-1}$, unter $\mathbb{P}_{0, \tau}$, hängt nicht von τ ab, und \hat{T}_n und T sind somit unabhängig unter jedem $\mathbb{P}_{0, \tau}$ (vgl. Korollar §19.11 und Beispiel §19.18). Bezeichnen wir mit $t_{n-1}^{\alpha/2}$ das obere $\alpha/2$ -Fraktile einer t_{n-1} -Verteilung, wobei wir auf die Randomisierung auf Grund der stetigen Verteilung von \hat{T}_n verzichten, so sind der ursprüngliche gleichmäßig beste unverfälschte α -Test φ_\circ (19.04) und der Test $\tilde{\varphi}_\circ = \mathbb{1}_{\{|\hat{T}_n| > t_{n-1}^{\alpha/2}\}} N_{\mathbb{R} \times \mathbb{R}_0^+}^n$ -f.ü. äquivalent. Zusammenfassend, für jedes $\mu_\circ \in \mathbb{R}$ ist der **zweiseitige t-Test** $\varphi = \mathbb{1}_{\{|\bar{X} - \mu_\circ| > n^{-1/2} \hat{S}_n t_{n-1}^{\alpha/2}\}}$ also ein gleichmäßig bester unverfälschter Test von $H_0 : \mu = \mu_\circ, \sigma \in \mathbb{R}_0^+$ gegen $H_1 : \mu \neq \mu_\circ, \sigma \in \mathbb{R}_0^+$. □

§20 Likelihood-Quotienten-Test

§20.01 **Definition.** Sei $(\mathcal{X}, \mathcal{X}, \mathbb{P}_\theta)$ ein μ -dominiertes statistisches Experiment mit Likelihood-Funktion $L(\theta) = d\mathbb{P}_\theta/d\mu \in \overline{\mathcal{X}^+}$ für $\theta \in \Theta$. Für nicht-leere Teilmengen $\Theta^0, \Theta^1 \subseteq \Theta$ heißt

$$\Lambda_{\Theta^0, \Theta^1} := \frac{\sup_{\theta \in \Theta^1} L(\theta)}{\sup_{\theta \in \Theta^0} L(\theta)} \mathbb{1}_{\{\sup_{\theta \in \Theta^0} L(\theta) \in \mathbb{R}_0^+\}} \in \overline{\mathcal{X}^+} \tag{20.01}$$

Likelihood-Quotienten-Statistik. Jeder randomisierte Test der Form

$$\varphi := \mathbb{1}_{\{\Lambda_{\Theta^0, \Theta^1} > k\}} + \zeta \mathbb{1}_{\{\Lambda_{\Theta^0, \Theta^1} = k\}} \quad \mu\text{-f.ü.} \tag{20.02}$$

für einen **kritischen Wert** $k \in \mathbb{R}^+$ und **messbarer Randomisierung** $\zeta : \{\Lambda_{\Theta^0, \Theta^1} = k\} \rightarrow [0, 1]$ wird **Likelihood-Quotienten-Test** (bzgl. $\Lambda_{\Theta^0, \Theta^1}$) genannt. □

§20.02 **Bemerkung.** Liegt Θ^1 dicht in Θ und ist $\theta \mapsto L(\theta)$ μ -f.ü. stetig, so gilt für den Maximum-Likelihood-Schätzer $\hat{\theta}$, falls er existiert, $\sup_{\theta \in \Theta^1} L(\theta) = \sup_{\theta \in \Theta} L(\theta) = L(\hat{\theta})$. Dies ist der Ausgangspunkt für die folgende asymptotische Theorie. □

§20.03 **Beispiel.** Es seien $\mathbb{P}_{\Theta_{\text{nat}}}$ eine natürliche Exponentialfamilie in S , Θ_{nat}^1 dicht in Θ_{nat} und $\hat{\theta}$ ein MLS. In der Darstellung (10.01) mit $\nu = h\mu$ gilt dann

$$\Lambda_{\Theta_{\text{nat}}^0, \Theta_{\text{nat}}^1}(x) = \inf_{\theta_\circ \in \Theta_{\text{nat}}^0} \exp(\langle \hat{\theta}(x) - \theta_\circ, S(x) \rangle - A(\hat{\theta}(x)) + A(\theta_\circ)) \quad \nu\text{-f.a. } x \in \mathcal{X}.$$

Gilt weiterhin $\hat{\theta} \in \text{inn}(\Theta_{\text{nat}})$, so folgt $\mathbb{E}_{\hat{\theta}(x)}(S) = T(x)$ gemäß **Lemma** §16.04 für ν -f.a. $x \in \mathcal{X}$ und daher mit **Lemma** §17.03 (iii)

$$\log(\Lambda_{\Theta_{\text{nat}}, \Theta_{\text{nat}}}) = \inf_{\theta_o \in \Theta_{\text{nat}}^0} \text{KL}(\mathbb{P}_{\hat{\theta}} | \mathbb{P}_{\theta_o}) \quad \nu\text{-f.ü.}$$

Mit anderen Worten, die Log-Likelihood-Quotienten-Statistik $\log(\Lambda_{\Theta_{\text{nat}}, \Theta_{\text{nat}}})$ gibt die Divergenz der zum MLS $\hat{\theta} \in \Theta_{\text{nat}}$ gehörenden Verteilung $\mathbb{P}_{\hat{\theta}}$ zur Familie der Verteilungen $(\mathbb{P}_{\theta})_{\theta \in \Theta_{\text{nat}}}$ an. \square

§20.04 **Lemma.** Sei \mathbb{P}_{θ} mit $\Theta \subseteq \mathbb{R}$ eine Verteilungsfamilie mit isotonem Dichtequotienten in S . Für $\alpha \in (0, 1)$ und $\theta_o \in \Theta$ betrachten wir das einseitige Testproblem $H_0 : \mathcal{H}^0 = \Theta \cap (-\infty, \theta_o]$ gegen $H_1 : \mathcal{H}^1 = \Theta \cap (\theta_o, \infty)$ und die Likelihood-Quotienten-Statistik $\Lambda := \Lambda_{\mathcal{H}^0, \mathcal{H}^1}$, so gilt:

- (i) Der Likelihood-Quotienten-Test $\varphi_o := \mathbb{1}_{\{\Lambda > k^\alpha\}} + \zeta^\alpha \mathbb{1}_{\{\Lambda = k^\alpha\}}$, wobei das obere α -Fraktile $k^\alpha \in \mathbb{R}^+$ von $\mathbb{P}_{\theta_o}^\Lambda$ und $\zeta^\alpha \in [0, 1]$ durch (18.05) bzw. (18.06) bestimmt ist, minimiert die Irrtumswahrscheinlichkeiten erster und zweiter Art gleichmäßig unter allen auf $\{\theta_o\}$ α -ähnlichen Tests.
- (ii) Der Test φ_o ist ein gleichmäßig bester Test zum Niveau α für das einseitige Testproblem $H_0 : \theta \leq \theta_o$ gegen $H_1 : \theta > \theta_o$.

§20.05 **Beweis** von **Lemma** §20.04. In der Vorlesung. \square

§20.06 **Bemerkung.** Sei \mathbb{P}_{θ} mit $\Theta \subseteq \mathbb{R}$ wie in **Lemma** §20.04 eine Verteilungsfamilie mit isotonem Dichtequotienten in S . Betrachten wir für $\theta_o \in \Theta$ ein zweiseitiges Testproblem $H_0 : \mathcal{H}^0 = \{\theta_o\}$ gegen $H_1 : \mathcal{H}^1 = \Theta \setminus \{\theta_o\}$ sowie die Likelihood-Quotienten-Statistik $\Lambda_{\mathcal{H}^0, \mathcal{H}^1}$. Dann enthält der Ablehnbereich eines Likelihood-Quotienten-Tests bzgl. $\Lambda_{\mathcal{H}^0, \mathcal{H}^1}$ ein Ereignis der Form $\{S \notin [k_1, k_2]\}$, allerdings ist der Test im Allgemeinen nicht mehr unverfälscht, wie folgendes Beispiel zeigt. Sei $X \sim \text{Poi}_{\mathbb{R}_0^+}$, dann ist $\Lambda_{\mathcal{H}^0, \mathcal{H}^1} = \exp(X(\log(X/\theta_o) - 1) + \theta_o)$ und der Ablehnbereich des entsprechenden Likelihood-Quotienten-Tests mit kritischem Wert $k \in \mathbb{R}_0^+$ enthält $\{X(\log(X/\theta_o) - 1) > \log(k) - \theta_o\}$. Da die Funktion $x \mapsto x(\log(x/\theta_o) - 1)$ auf $(0, \theta_o)$ nicht-positiv und streng antiton sowie auf (θ_o, ∞) streng isoton ist, entspricht für $\log(k) > \theta_o$ das Ereignis $\{X(\log(X/\theta_o) - 1) > \log(k) - \theta_o\}$ einem einseitigen Ablehnbereich $\{X > \tilde{k}\}$. Hingegen sind im Fall einer Normalverteilung ein- und zweiseitige Gauß- und t-Tests Likelihood-Quotienten-Tests. \square

§20.07 **Vorbemerkung.** Im Folgenden betrachten wir wie im **Satz** §17.22 eine Folge μ -dominierter Produktexperimente $((\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_{\theta}^n))_{n \in \mathbb{N}}$ mit $\Theta \subseteq \mathbb{R}^k$, marginaler Likelihood-Funktion $L(\theta) = d\mathbb{P}_{\theta}/d\mu \in \overline{\mathcal{X}^+}$ und marginaler Loglikelihood-Funktion $\ell(\theta) = \log(L(\theta)) \in \overline{\mathcal{X}}$. Existieren weiterhin ein MLS $\hat{\theta}_n$ und $\hat{\theta}_n^0$ im statistischen Modell $(\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_{\theta}^n)$ bzw. $(\mathcal{X}^n, \mathcal{X}^n, \mathbb{P}_{\theta}^n)$ für eine Teilmenge $\Theta^0 \subseteq \Theta$, dann gilt $\Lambda_{\Theta^0, \Theta}(x) = \prod_{i \in [n]} (L(\hat{\theta}_n(x), x_i) / L(\hat{\theta}_n^0(x), x_i))$ für die entsprechende Likelihood-Quotienten-Statistik. Unter Verwendung des Kontrastprozesses $\widehat{K}_n(\theta) \in \overline{\mathcal{X}^n}$ mit $\widehat{K}_n(\theta, x) := -\frac{1}{n} \sum_{i \in [n]} \ell(\theta, x_i)$ (vgl. **Korollar** §17.06) lässt sich somit die Loglikelihood-Quotienten-Statistik wie folgt schreiben:

$$\begin{aligned} \log(\Lambda_{\Theta^0, \Theta}(x)) &= n(\widehat{K}_n(\hat{\theta}_n^0(x), x) - \widehat{K}_n(\hat{\theta}_n(x), x)) \\ &= \sup_{\theta \in \Theta} \sum_{i \in [n]} \ell(\theta, x_i) - \sup_{\theta \in \Theta^0} \sum_{i \in [n]} \ell(\theta, x_i) =: \widehat{\lambda}_n(x) \end{aligned}$$

Unter den Annahmen von **Satz** §17.22 lässt sich nun die asymptotische Verteilung der Statistik $\widehat{\lambda}_n$ bestimmen, die es erlaubt einen Likelihood-Quotienten-Test anzugeben, der asymptotisch ein vorgegebenes Signifikanz-Niveau einhält. \square

§20.08 **Satz.** Es gelten die Annahmen und Notationen aus **Satz** §17.22.

- (i) Betrachte das Testproblem $H_0 : \theta = \theta_0$ gegen $\theta \neq \theta_0$ für ein $\theta_0 \in \text{inn}(\Theta)$. Dann gilt für die Loglikelihood-Quotienten-Statistik $\hat{\lambda}_n(x) = \sup_{\theta \in \Theta} \sum_{i \in \llbracket n \rrbracket} \ell(\theta, x_i) - \sum_{i \in \llbracket n \rrbracket} \ell(\theta_0, x_i)$ unter \mathbb{P}_{θ_0} die Verteilungskonvergenz $2\hat{\lambda}_n \xrightarrow{D} \chi_k^2$.
- (ii) Allgemeiner für $r \in \llbracket k \rrbracket$ seien $A \in \mathbb{R}^{(r,k)}$ eine Matrix mit $\text{rg}(A) = r$, $a \in \mathbb{R}^k$ und $\Theta = \mathcal{H}^0 \uplus \mathcal{H}^1 \subseteq \mathbb{R}^k$ mit $\mathcal{H}^0 := \{\theta \in \Theta : A(\theta - a) = 0\}$. Dann gilt für die Loglikelihood-Quotienten-Statistik $\hat{\lambda}_n(x) = \sup_{\theta \in \Theta} \sum_{i \in \llbracket n \rrbracket} \ell(\theta, x_i) - \sup_{\theta \in \mathcal{H}^0} \sum_{i \in \llbracket n \rrbracket} \ell(\theta, x_i)$ unter jedem \mathbb{P}_{θ_0} mit $\theta_0 \in \mathcal{H}^0 \cap \text{inn}(\Theta)$ die Verteilungskonvergenz $2\hat{\lambda}_n \xrightarrow{D} \chi_r^2$.

Insbesondere hält der Likelihood-Quotienten-Test $\varphi = \mathbb{1}_{\{\hat{\lambda}_n > \chi_{r,\alpha}^2/2\}}$ (im Fall (i) $r = k$) auf $\mathcal{H}^0 \cap \text{inn}(\Theta)$ asymptotisch das Signifikanzniveau $\alpha \in (0, 1)$ ein, wobei $\chi_{r,\alpha}^2$ das obere α -Fraktile einer χ_r^2 -Verteilung mit r Freiheitsgraden bezeichnet und wir auf Grund der stetigen asymptotischen Verteilung auf eine Randomisierung verzichten.

§20.09 **Beweis** von **Satz** §20.08. Shao [2003], Theorem 6.5, S. 432 □

§20.10 **Bemerkung.**

- (i) Der Likelihood-Quotienten-Test wird asymptotisch *verteilungsfrei* genannt, da die asymptotische Verteilung von $\hat{\lambda}_n$ unabhängig von $\theta_0 \in \Theta^0$ ist.
- (ii) Die asymptotische Verteilung von $2\hat{\lambda}_n$ ist unter lokalen Alternativen $\theta = \theta_0 + h/\sqrt{n}$ eine nicht-zentrale χ_r^2 -Verteilung (vgl. van der Vaart [1998], Satz 16.7). Für feste Alternativen $\theta \in \mathcal{H}^1$ ist der der Likelihood-Quotienten-Test $\varphi_n = \mathbb{1}_{\{\hat{\lambda}_n > \chi_{r,\alpha}^2/2\}}$ damit *konsistent*, es gilt also $\beta_{\varphi_n}(\theta) \xrightarrow{n \rightarrow \infty} 1$.
- (iii) Der *Wald-Test* und der *Score-Test* sind zwei weitere asymptotische Likelihood-Tests. Ein Wald-Test ist von der Form $\mathbb{1}_{\{W_n > \chi_{k,\alpha}^2\}}$ mit oberem α -Fraktile $\chi_{k,\alpha}^2$ einer χ_k^2 -Verteilung als kritischem Wert und Wald-Test-Statistik $W_n = n \langle \mathcal{I}_{\hat{\theta}_n}(\hat{\theta}_n - \theta_0), (\hat{\theta}_n - \theta_0) \rangle$ unter Verwendung des MLS $\hat{\theta}_n$, da unter den Bedingungen von **Satz** §17.22 W_n asymptotisch eine χ_k^2 -Verteilung besitzt. Ein Score-Test ist von der Form $\mathbb{1}_{\{R_n > \chi_{k,\alpha}^2\}}$ mit Score-Test-Statistik $R_n = \frac{1}{n} \langle \mathcal{I}_{\hat{\theta}_n}^{-1} \sum_{i \in \llbracket n \rrbracket} \dot{\ell}(\theta_0, X_i), \sum_{i \in \llbracket n \rrbracket} \dot{\ell}(\theta_0, X_i) \rangle$, die unter den Bedingungen von **Satz** §17.22 ebenfalls asymptotisch eine χ_k^2 -Verteilung besitzt (Shao [2003], Theorem 6.6, S. 434). □

§20.11 **Beispiel.** Wir führen $n \in \mathbb{N}$ unabhängige Experimente durch, deren Ergebnisse in einer endlichen Menge $\llbracket k \rrbracket$, $k \in \mathbb{N}$, liegen, und zählen die Häufigkeiten $n_i \in \llbracket n \rrbracket$ der einzelnen Ergebnisse $i \in \llbracket k \rrbracket$, also $n = \sum_{i \in \llbracket k \rrbracket} n_i$. Wir möchten anhand der beobachteten Häufigkeiten $(n_i)_{i \in \llbracket k \rrbracket} \in \llbracket n \rrbracket^k$, Hypothesen über die unbekannte Wahrscheinlichkeit $p_i \in [0, 1]$ des Auftretens der einzelnen Ergebnisse $i \in \llbracket k \rrbracket$ treffen. Der zufällige Vektor $N = (N_i)_{i \in \llbracket k \rrbracket}$ der Häufigkeiten besitzt dann eine Multinomial-Verteilung mit Parametern $n \in \mathbb{N}$ und Parametervektor $p = (p_i)_{i \in \llbracket k-1 \rrbracket} \in \Theta := \{x \in [0, 1]^{k-1} : \sum_{i \in \llbracket k-1 \rrbracket} x_i \leq 1\}$ mit der Konvention $p_k = 1 - \sum_{i \in \llbracket k-1 \rrbracket} p_i$. Wir schreiben kurz $N \odot (\text{MN}_{(n,p)})_{p \in \Theta}$. Betrachten wir für $n = 1$ das entsprechende statistische Experiment $(\mathcal{X} = \{0, 1\}^k, 2^{\mathcal{X}}, (\text{MN}_{(1,p)})_{p \in \Theta})$, so ist die Statistik N erschöpfend für das Produktexperiment $(\mathcal{X}^n, 2^{\mathcal{X}^n}, (\text{MN}_{(1,p)}^n)_{p \in \Theta})$ und somit ist die obige Asymptotik für $n \rightarrow \infty$ anwendbar. Der MLS für p ist $\hat{p}_n = \frac{1}{n} N$, so dass für $\Theta^0 = \{p^0\}$ für ein $p^0 \in \Theta$ die Loglikelihood-Quotienten-Statistik erfüllt

$$\hat{\lambda}((n_i)_{i \in \llbracket k \rrbracket}) = \log \left(\frac{\binom{n}{n_1 \dots n_k} \prod_{i \in \llbracket k \rrbracket} (n_i/n)^{n_i}}{\binom{n}{n_1 \dots n_k} \prod_{i \in \llbracket k \rrbracket} (p_i^0)^{n_i}} \right) = \sum_{i \in \llbracket k \rrbracket} n_i \log(n_i / (n p_i^0)). \tag{20.03}$$

Bezeichnet $\mathbb{E}_{p^o}^n$ die Erwartung bzgl. $\mathbb{P}_{p^o}^n := \text{MN}_{(1,p^o)}^n$, so gilt $\mathbb{E}_{p^o}^n(N) = np^o$ sowie für $i \in \llbracket k \rrbracket$ auch $\mathbb{E}_{p^o}^n((N_i - np_i^o)^2) = np_i^o(1 - p_i^o)$, also $\mathbb{E}_{p^o}^n((N_i - np_i^o)^2/(np_i^o)) = 1 - p_i^o \leq 1$. Unter Verwendung der Entwicklung $(x+h) \log((x+h)/x) = h + h^2/(2x) + o(h^2/x)$ sowie $\sum_{i \in \llbracket k \rrbracket} N_i = n$ gilt für $\hat{\lambda} = \hat{\lambda}(N)$ asymptotisch damit

$$\begin{aligned} 2\hat{\lambda} &= 2 \sum_{i \in \llbracket k \rrbracket} \left((N_i - np_i^o) + \frac{(N_i - np_i^o)^2}{2np_i^o} + o((N_i - np_i^o)^2/(np_i^o)) \right) \\ &= \sum_{i \in \llbracket k \rrbracket} \frac{(N_i - np_i^o)^2}{np_i^o} + o_{\mathbb{P}^n}(1). \end{aligned}$$

Betrachten wir das Testproblem $H_0 : p = p^o$ gegen $H_1 : p \neq p^o$, so konvergiert unter der Nullhypothese $\mathcal{H}^0 = \{p^o\}$ auch $\sum_{i \in \llbracket k \rrbracket} \frac{(N_i - np_i^o)^2}{np_i^o}$ in Verteilung gegen eine χ_{k-1}^2 -Verteilung mit $k - 1$ Freiheitsgraden. \square

§20.12 **Definition.** Sei $N \odot (\text{MN}_{(n,p)})_{p \in \Theta}$ multinomial-verteilt mit bekanntem Parameter $n \in \mathbb{N}$ und unbekanntem Parameter $p \in \Theta$. Dann heißt $Q_n := \sum_{i \in \llbracket k \rrbracket} \frac{(N_i - np_i^o)^2}{np_i^o}$ *Pearson's χ^2 -Statistik* und für das Testproblem $H_0 : p = p^o$ gegen $H_1 : p \neq p^o$ wird $\varphi = \mathbb{1}_{\{Q > \chi_{k-1,\alpha}^2\}}$ *χ^2 -Test* genannt.

§20.13 **Korollar.** Der *χ^2 -Test* hält unter der Nullhypothese $H_0 : p = p^o$ asymptotisch das Signifikanzniveau $\alpha \in (0, 1)$ ein.

§20.14 **Beweis** von **Korollar** §20.13. Die Behauptung folgt direkt aus **Satz** §20.08. \square

§20.15 **Bemerkung.** Der *χ^2 -Test* dient häufig als *Goodness-of-fit* Test. So kann zum Beispiel getestet werden, ob jede Ziffer einer Zufallszahl mit der gleichen Wahrscheinlichkeit auftritt. Dies entspricht dem Fall $k = 10$, Anzahl der Ziffern n und $p_0^o = \dots = p_9^o = \frac{1}{10}$. \square

§21 Rangtest

§21.01 **Vorbemerkung.** Auf dem Stichprobenraum $(\mathbb{R}^n, \mathcal{B}^n)$ betrachte die Statistik $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit $x \mapsto T(x) := (T_i(x))_{i \in [n]}$ und $T_i(x) := \min\{c \in \mathbb{R} : \sum_{j \in [n]} \mathbb{1}_{\{x_j \leq c\}} \geq i\}$ für $i \in [n]$. Da $T_1(x) \leq T_2(x) \leq \dots \leq T_n(x)$ für alle $x \in \mathbb{R}^n$ gilt, wird die Statistik T (und auch jede andere Statistik mit dieser Eigenschaft) *Ordnungsstatistik* genannt (vgl. **Beispiel** §09.01). Weiterhin bezeichne \mathcal{S}_n die symmetrische Gruppe vom Grad n , also die Menge aller Permutationen der Zahlen $[n]$, wobei wie üblich die Abbildung $s : i \mapsto s(i) =: s_i$ mit dem Vektor $s = (s_i)_{i \in [n]}$ identifiziert sei. Wir bezeichnen mit $s^- \in \mathcal{S}_n$ die zu $s \in \mathcal{S}_n$ inverse Permutation, also $\text{id}_{\mathcal{S}_n} = s \circ s^- = s^- \circ s$. Für eine Permutation $s = (s_i)_{i \in [n]} \in \mathcal{S}_n$ und einen Vektor $x = (x_i)_{i \in [n]} \in \mathbb{R}^n$ schreiben wir kurz $x_s := (x_{s_i})_{i \in [n]}$. Versehen wir die endliche Menge \mathcal{S}_n mit ihrer Potenzmenge $2^{\mathcal{S}_n}$ als σ -Algebra, so wird jede \mathcal{B}^n - $2^{\mathcal{S}_n}$ -messbare (Borel-messbare) Abbildung $S := (S_i)_{i \in [n]} : \mathbb{R}^n \rightarrow \mathcal{S}_n$, also $S^{-1}(\{s\}) \in \mathcal{B}^n$ für alle $s \in \mathcal{S}_n$, *zufällige Permutation* auf $(\mathbb{R}^n, \mathcal{B}^n)$ genannt. Die entsprechende *zufällige inverse Permutation* $S^- : \mathbb{R}^n \rightarrow \mathcal{S}_n$ ist festgelegt durch $\text{id}_{\mathcal{S}_n} = S^-(x) \circ S(x) = S(x) \circ S^-(x)$ für alle $x \in \mathbb{R}^n$, und natürlich auch Borel-messbar. Der zufällige Vektor $X_S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit $x \mapsto X_S(x) := (x_{S_i(x)})_{i \in [n]} = x_{S(x)} = \sum_{s \in \mathcal{S}_n} x_s \mathbb{1}_s(S(x))$ (eine endliche Summe Borel-messbarer Abbildungen $x \mapsto x_s \mathbb{1}_{S^{-1}(\{s\})}(x)$) wird *zufällige Anordnung* genannt. \square

§21.02 **Definition.** Eine zufällige Permutation $O = (O_i)_{i \in [n]}$ auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *Ordnungspermutation*, wenn die entsprechende zufällige Anordnung $X_O : \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit $x \mapsto x_{O(x)}$ eine Ordnungsstatistik ist, als $x_{O_1(x)} \leq x_{O_2(x)} \leq \dots \leq x_{O_n(x)}$ für alle $x \in \mathbb{R}^n$ gilt. Eine zufällige Permutation $R = (R_i)_{i \in [n]}$ auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *Rangpermutation*, wenn die zugehörige zufällige inverse Permutation $O := R^-$ eine Ordnungspermutation ist. Für jedes $i \in [n]$ heißt die i -te Koordinate $R_i(x)$ von $R(x)$ *Rang* der i -ten Koordinate von $x \in \mathbb{R}^n$. \square

§21.03 **Bemerkung.** Auf der Borel-Menge $\{x_i \neq x_j\} := \{(x_i)_{i \in [n]} \in \mathbb{R}^n : x_i \neq x_j, \forall j \in [n] \setminus \{i\}, \forall i \in [n]\}$ ist eine Ordnungspermutation O nur eindeutig festgelegt. Für $x \in \mathbb{R}^n$, Permutation $o = O(x) \in \mathcal{S}_n$ und $i \in [n]$ entspricht der Wert an der i -ten Position im geordneten Vektor x_o gerade dem Wert an der o_i -ten Stelle in x . Andererseits, für die Permutation $r := R(x) \in \mathcal{S}_n$ der Rangpermutation $R := O^-$ ist der Wert an der r_i -ten Position im geordneten Vektor x_o gerade gleich dem Wert an der i -ten Position im Vektor x . \square

§21.04 **Anmerkung.** Die Abbildung $R^* = (R_i^*)_{i \in [n]} : \mathbb{R}^n \rightarrow \mathcal{S}_n$ mit $x \mapsto R_i^*(x) := \sum_{j \in [i]} \mathbb{1}_{\{x_i = x_j\}} + \sum_{j \in [n]} \mathbb{1}_{\{x_i > x_j\}}$ für jedes $i \in [n]$ ist eine Rangpermutation. In der Tat, für jedes $x \in \mathbb{R}^n$ ist $r := R^*(x) \in \mathcal{S}_n$ ($r : [n] \rightarrow [n]$ ist injektiv, und damit auch bijektiv) und die inverse Permutation $o := r^-$ erfüllt $x_{o_1} \leq x_{o_2} \leq \dots \leq x_{o_n}$. Da jede Koordinate von R^* eine \mathcal{B} - $2^{[n]}$ -messbare Abbildung ist, ist also R eine Rangpermutation. Auf der Borel-Menge $\{x_i \neq x_j\}$ ist jede Rangpermutation $R = (R_i)_{i \in [n]}$ eindeutig durch $R_i(x) = \sum_{j \in [n]} \mathbb{1}_{\{x_j \leq x_i\}} = R_i^*(x)$, $i \in [n]$, festgelegt. Wir bezeichnen für $y \in \mathbb{R}$ wie bisher mit $\widehat{F}_n(y) := \widehat{\mathbb{P}}(\mathbb{1}_{(-\infty, y]})$ mit $\widehat{F}_n(y, x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{x_i \leq y\}} \in [0, 1]$ für alle $x \in \mathbb{R}^n$ die *empirische Verteilungsfunktion*. Für $r := R(x)$ und $o := r^-$ mit $x \in \{x_i \neq x_j\}$ sind die beiden Gleichungen $i = n \widehat{F}_n(x_{o_i}, x)$ und $r_i = n \widehat{F}_n(x_i, x)$ für jedes $i \in [n]$ erfüllt. \square

§21.05 **Vorbemerkung.** Wir nehmen im Folgenden an, dass $\mathbb{P}^n := \bigotimes_{i \in [n]} \mathbb{P}_i \in \mathcal{W}(\mathcal{B}^n)$ ein Produktmaß auf dem Stichprobenraum $(\mathbb{R}^n, \mathcal{B}^n)$ ist, wobei für jedes $i \in [n]$ die Randverteilung $\mathbb{P}_i \in \mathcal{W}(\mathcal{B})$ eine Lebesgue-Dichte $f_i = d\mathbb{P}_i/d\lambda \in \mathcal{B}^+$ und somit $\mathbb{P}^n \ll \lambda^n$ auch eine Lebesgue-Dichte $\prod_{i \in [n]} f_i = d\mathbb{P}^n/d\lambda^n \in (\mathcal{B}^+)^n$ besitzt. Da das Komplement $\{x_i = x_j\} := \{x_i \neq x_j\}^c$ der Borel-Menge $\{x_i \neq x_j\}$ eine Lebesgue-Nullmenge ist, ist es auch eine \mathbb{P}^n -Nullmenge. Für jede

Rangpermutation R auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit entsprechender Ordnungspermutation $O := R^-$ gilt somit $x_{O_1(x)} < x_{O_2(x)} < \dots < x_{O_n(x)}$ für \mathbb{P}^n -f.a. $x \in \mathbb{R}^n$. Insbesondere ist für \mathbb{P}^n -f.a. $x \in \mathbb{R}^n$ der Rangvektor $R(x) = (R_i(x))_{i \in \llbracket n \rrbracket}$ (und damit die Rangpermutation R) eindeutig durch $R_i(x) = \sum_{j \in \llbracket n \rrbracket} \mathbb{1}_{\{x_j \leq x_i\}} = n \widehat{F}_n(x_i, x)$ für jedes $i \in \llbracket n \rrbracket$ festgelegt. \square

§21.06 **Lemma.** Es seien $\mathbb{P}^n \in \mathcal{W}(\mathcal{B}^n)$ ein Produktmaß auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit identischer Randverteilung $\mathbb{P} \in \mathcal{W}(\mathcal{B})$, Verteilungsfunktion $F(y) := \mathbb{P}((-\infty, y])$, $y \in \mathbb{R}$, und Lebesgue-Dichte $f = d\mathbb{P}/d\lambda \in \mathcal{B}^+$, sowie R eine Rangpermutation auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit entsprechender Ordnungsstatistik X_O für $O := R^-$. Dann gelten die folgenden Aussage:

- (i) R ist unter \mathbb{P}^n gleichförmigverteilt (Laplace-verteilt) auf der symmetrischen Gruppe \mathcal{S}_n , also $(\mathbb{P}^n)^R(\{s\}) = (\mathbb{P}^n \circ R)(\{s\}) = \mathbb{P}^n(R = s) = \frac{1}{n!}$, $s \in \mathcal{S}_n$, kurz $R \sim (\mathbb{P}^n)^R = U_{\mathcal{S}_n}$.
- (ii) R und der geordnete Vektor X_O sind unabhängig unter \mathbb{P}^n .
- (iii) Die Verteilung $(\mathbb{P}^n)^{X_O}$ von X_O besitzt eine Lebesgue-Dichte $f^{X_O}(x) = n! \mathbb{1}_B(x) \prod_{i=1}^n f(x_i)$, $x \in \mathbb{R}^n$, mit $B := \{(x_i)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n, x_1 < \dots < x_n\}$.
- (iv) Für jedes $i \in \llbracket n \rrbracket$ besitzt die Verteilung der i -te Koordinate von X_O unter \mathbb{P}^n eine Lebesgue-Dichte $f^{X_i}(x) = i \binom{n}{i} |F(x)|^{i-1} |1 - F(x)|^{n-i} f(x)$, $x \in \mathbb{R}$.

§21.07 **Beweis von Lemma** §21.06. In der Vorlesung. \square

§21.08 **Definition.** Für $\mathbb{P}_0, \mathbb{P} \in \mathcal{W}(\mathcal{B})$ heißt \mathbb{P}_0 *stochastisch kleiner* als \mathbb{P} , kurz $\mathbb{P}_0 \preceq \mathbb{P}$, wenn für alle $c \in \mathbb{R}$ gilt $\mathbb{P}_0((c, \infty)) \leq \mathbb{P}((c, \infty))$. Gilt zusätzlich $\mathbb{P}_0 \neq \mathbb{P}$, dann schreiben wir $\mathbb{P}_0 \prec \mathbb{P}$. \square

§21.09 **Bemerkung.** Vereinfachend besagt $\mathbb{P}_0 \preceq \mathbb{P}$, dass Beobachtungen von \mathbb{P}_0 typischerweise kleiner sind als Beobachtungen von \mathbb{P} . \square

§21.10 **Beispiel.** Für $a, b \in \mathbb{R}$ und $\sigma \in \mathbb{R}_0^+$ betrachte zwei Normalverteilungen $N_{(a, \sigma^2)}$ und $N_{(b, \sigma^2)}$ auf $(\mathbb{R}, \mathcal{B})$. Dann gilt $N_{(a, \sigma^2)} \prec N_{(b, \sigma^2)}$ genau dann, wenn $a \leq b$. Allgemeiner sei \mathbb{P}_θ eine Lokations-Familie auf $(\mathbb{R}, \mathcal{B})$ mit Likelihood-Funktion L bzgl. des Lebesgues-Maßes, also $L(\theta, x) = L(0, x - \theta)$ für alle $x \in \mathbb{R}$ und $\theta \in \Theta$ (vgl. Definition §14.01). Ist die Lebesguedichte $L(0)$ strikt positive auf \mathbb{R} , dann gilt $\mathbb{P}_{\theta_1} \prec \mathbb{P}_{\theta_2}$ genau dann, wenn $\theta_1 \leq \theta_2$. \square

§21.11 **Motivation.** Für $\mathbb{P}_0, \mathbb{P} \in \mathcal{W}(\mathcal{B})$ möchten wir die Nullhypothese $H_0 : \mathbb{P}_0 = \mathbb{P}$ gegen die Alternative $H_1 : \mathbb{P}_0 \prec \mathbb{P}$ testen. Vereinfachend, meint dies, dass wir die Nullhypothese ablehnen möchten, wenn die Beobachtungen von \mathbb{P}_0 *signifikant* kleiner als die Beobachtungen von \mathbb{P} sind. Wir betrachten dazu eine (pooled) Stichprobe von $n = m + l$ unabhängigen reellen Zufallsvariablen $X = (X_i)_{i \in \llbracket n \rrbracket}$, also mit Werten in $(\mathbb{R}^n, \mathcal{B}^n)$, wobei die ersten m und die letzten l Zufallsvariablen identisch \mathbb{P}_0 -verteilt bzw. \mathbb{P} -verteilt sind. Die Nullhypothese möchten wir also ablehnen, wenn die Beobachtungen der ersten m Zufallsvariablen *signifikant* kleiner sind als die Beobachtungen der letzten l Zufallsvariablen. Gegeben eine Rangpermutation R auf $(\mathbb{R}^n, \mathcal{B}^n)$ und eine Beobachtung $x \in \mathbb{R}^n$ scheint es vernünftig die Nullhypothese abzulehnen, wenn die Summe der ersten m Ränge (erste Gruppe), also $W_o(x) := \sum_{i \in \llbracket m \rrbracket} R_i(x)$, *signifikant* kleinere Werte als die Summe der letzten l Ränge (zweite Gruppe), also $W(x) := \sum_{i \in \llbracket l \rrbracket} R_{i+m}(x)$, annimmt, wobei definitionsgemäß $W_o(x) + W(x) = \sum_{i \in \llbracket n \rrbracket} R_i(x) = \sum_{i \in \llbracket n \rrbracket} i = \frac{n(n+1)}{2}$ gilt. \square

§21.12 **Lemma.** Für $m, l \in \mathbb{N}$ und $n := m + l$ seien $R = (R_i)_{i \in \llbracket n \rrbracket}$ eine Rangpermutation auf $(\mathbb{R}^n, \mathcal{B}^n)$, $U_{ml} : \mathbb{R}^n \rightarrow \llbracket 0, ml \rrbracket$ mit $x \mapsto U_{ml}(x) := \sum_{i \in \llbracket m \rrbracket} \sum_{j \in \llbracket l \rrbracket} \mathbb{1}_{\{x_i > x_{j+m}\}}$, $W_o := \sum_{i \in \llbracket m \rrbracket} R_i$ und $W := \sum_{i \in \llbracket l \rrbracket} R_{i+m}$. Für alle $x \in \{x_i \neq x_j\}$ gilt dann $W_o(x) = U_{ml}(x) + \frac{m(m+1)}{2}$ und somit $W(x) = ml - U_{ml}(x) + \frac{l(l+1)}{2}$.

§21.13 **Beweis von Lemma** §21.12. In der Vorlesung. \square

§21.14 **Bemerkung.** Betrachten wir die Teststatistik W_o oder die dazu äquivalente Teststatistik U_{ml} so lehnen wir die Nullhypothese $H_0 : \mathbb{P}_o = \mathbb{P}$ gegen die Alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$ ab, wenn $U_{ml} < c$ oder dazu äquivalent $W_o < c + \frac{m(m+1)}{2}$ für einen kritischen Wert $c \in (0, kl]$ gilt. Dieser Test heißt (einseitiger) Mann-Whitney U-Test oder Wilcoxon Zweistichproben Rangsummentest. Die auf W_o basierende Version wurde von Wilcoxon [1945] vorgeschlagen, während die U_{ml} -Version unabhängig davon von Mann and Whitney [1947] untersucht wurde. Die Wahl des kritischen Wertes bezüglich eines vorgegebenen Signifikanzniveau $\alpha \in (0, 1)$ beruht wie bisher auf der Verteilung der Statistik U_{ml} oder einer asymptotischen Approximation unter der Nullhypothese. Die nächste Aussage zeigt, dass unter der Nullhypothese die Verteilung von U_{ml} in dem folgenden Sinn *verteilungsfrei* ist. Ist $\mathbb{P}_o = \mathbb{P}$ stetig-verteilt, dann ist die Verteilung von U_{ml} eindeutig festgelegt und unabhängig von der zugrundeliegenden Verteilung \mathbb{P}_o . \square

§21.15 **Lemma.** Für $m, l \in \mathbb{N}$ und $n = m + l$ sei $\mathbb{P}^n \in \mathcal{W}(\mathcal{B}^n)$ ein Produktmaß mit identischer Randverteilung $\mathbb{P} \ll \lambda$. Für $k \in \llbracket 0, ml \rrbracket$ gilt dann $\mathbb{P}^n(U_{ml} = k) = N(k; m, l) / \binom{n}{k}$, wobei $N(k; m, l)$ die Anzahl aller Zerlegungen $\sum_{i \in \llbracket m \rrbracket} k_i = k$ von k in m aufsteigend sortierte Zahlen $k_1 \leq k_2 \leq \dots \leq k_m$ aus der Menge $\llbracket 0, l \rrbracket$ ist. Insbesondere gilt $\mathbb{P}^n(U_{ml} = k) = \mathbb{P}^n(U_{ml} = ml - k)$.

§21.16 **Beweis von Lemma §21.15.** Georgii [2015], Satz 11.26, S. 342 \square

§21.17 **Bemerkung.** Für kleine Werte von n kann die Anzahl $N(k; m, l)$ mit Hilfe kombinatorischen Abzählens exakt bestimmt werden. Für große Werte von n kann die exakte aber rechenintensive Berechnung der Fraktile der Verteilung von U_{ml} unter Verwendung einer geeigneten asymptotischen Approximation umgangen werden. Im Folgenden seien m und l derart, dass $m + l$ unbeschränkt wächst. Formal betrachte wir Folgen $(m_n)_{n \in \mathbb{N}}$ und $(l_n)_{n \in \mathbb{N}}$ mit $m_n + l_n = n$ für jedes $n \in \mathbb{N}$. Wir nehmen weiterhin an, dass $m_n/n \xrightarrow{n \rightarrow \infty} \gamma \in (0, 1)$ und dementsprechend $l_n/n \xrightarrow{n \rightarrow \infty} 1 - \gamma$ gilt. Zur einfacheren Darstellung lassen wir den zusätzlichen Index n weg und schreiben weiterhin kurz $n = m + l$ mit $m/n \xrightarrow{n \rightarrow \infty} \gamma$ sowie $l/n \xrightarrow{n \rightarrow \infty} 1 - \gamma$. \square

§21.18 **Satz.** Für $m, l \in \mathbb{N}$ und $n = m + l$ sei $\mathbb{P}^n \in \mathcal{W}(\mathcal{B}^n)$ ein Produktmaß mit identischer Randverteilung $\mathbb{P} \ll \lambda$ und somit stetiger Randverteilungsfunktion \mathbb{F} . Betrachte $U_{ml} : \mathbb{R}^n \rightarrow \llbracket 0, ml \rrbracket$ und $T_{ml} : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $x \mapsto U_{ml}(x) := \sum_{i \in \llbracket m \rrbracket} \sum_{j \in \llbracket l \rrbracket} \mathbb{1}_{\{x_i > x_{j+m}\}}$ und

$$x \mapsto T_{ml}(x) := l \sum_{i \in \llbracket m \rrbracket} \mathbb{F}(x_i) - m \sum_{j \in \llbracket l \rrbracket} \mathbb{F}(x_{j+m}) = l \sum_{i \in \llbracket m \rrbracket} (\mathbb{F}(x_i) - 1/2) - m \sum_{j \in \llbracket l \rrbracket} (\mathbb{F}(x_{j+m}) - 1/2)$$

Weiterhin definiere $v_{ml} := ml(n+1)/12$, $T_{ml}^* := T_{ml}/\sqrt{v_{ml}}$ sowie $U_{ml}^* := (U_{ml} - ml/2)/\sqrt{v_{ml}}$. Falls $m/n \xrightarrow{n \rightarrow \infty} \gamma \in (0, 1)$ so gilt $U_{ml}^* - T_{ml}^* = o_{\mathbb{P}^n}(1)$, $T_{ml}^* \xrightarrow{D} N_{(0,1)}$ und somit $U_{ml}^* \xrightarrow{D} N_{(0,1)}$ für $n \rightarrow \infty$.

§21.19 **Beweis von Satz §21.18.** Georgii [2015], Satz 11.29, S. 344 \square

§21.20 **Bemerkung.** Für $n = m + l$ unabhängige reelle Zufallsvariablen $(X_i)_{i \in \llbracket m \rrbracket}$ mit identischer Randverteilung \mathbb{P}_o und $(X_{j+m})_{j \in \llbracket l \rrbracket}$ mit identischer Randverteilung \mathbb{P} betrachte den Test $\mathbb{1}_{\{U_{ml} < ml/2 + z_\alpha \sqrt{v_{ml}}\}}$, der die Nullhypothese $H_0 : \mathbb{P}_o = \mathbb{P}$ gegen die Alternative $H_1 : \mathbb{P}_o \prec \mathbb{P}$ also ablehnt, wenn $U_{ml} < ml/2 + z_\alpha \sqrt{v_{ml}}$ gilt, wobei z_α das untere α -Fraktile einer Standardnormalverteilung $N_{(0,1)}$ ist und wir auf Grund der stetigen asymptotischen Verteilung auf eine Randomisierung verzichten. Dieser Test hält asymptotisch das Signifikanzniveau ein, da nach Satz §21.18 unter der Nullhypothese gilt $\mathbb{P}_o^n(U_{ml} < ml/2 + z_\alpha \sqrt{v_{ml}}) \xrightarrow{n \rightarrow \infty} N_{(0,1)}((-\infty, z_\alpha)) = \alpha$ für $m/n \xrightarrow{n \rightarrow \infty} \gamma \in (0, 1)$. Betrachten wir dagegen das Testproblem $H_0 : \mathbb{P}_o = \mathbb{P}$ gegen $\mathbb{P}_o \succ \mathbb{P}$ so lehnen wir die Nullhypothese ab, falls $U_{ml} > ml/2 + z_\alpha \sqrt{v_{ml}}$. In der Vorlesung Statistik II wird die asymptotische Macht eines Rangtestes mit der eines t-Testes unter lokalen Alternativen verglichen. \square

6 Nonparametric estimation

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.

§22 Introduction

Nonparametric density estimation. Consider for a non-empty set of parameters Θ a family \mathbb{P}_Θ of probability measures on $(\mathbb{R}, \mathcal{B})$ which contains the distribution of an observable real random variable, $X \sim \mathbb{P}_\Theta$. The family \mathbb{P}_Θ 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 $X \sim \mathbb{P}_\theta$ for some probability measure $\mathbb{P}_\theta \in \mathcal{W}(\mathcal{B})$, means, the data generating process is known to us in advance. On the contrary, a parameter set $\Theta = \mathcal{W}(\mathcal{B})$ reflects a lack of prior knowledge. A parametric model \mathbb{P}_Θ 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 $(X_i)_{i \in \llbracket n \rrbracket} \stackrel{i.i.d.}{\sim} \mathbb{P} \in \mathcal{W}(\mathcal{B})$, that is, an independent and identically distributed sample with common probability measure $\mathbb{P} \in \mathcal{W}(\mathcal{B})$. A reasonable estimator of the associated cumulative distribution function (c.d.f.) $F(t) := \mathbb{P}((-\infty, t])$, $t \in \mathbb{R}$, is the empirical cumulative distribution function (e.c.d.f.) $\widehat{F}_n(t) := \widehat{\mathbb{P}}_n((-\infty, t])$, $t \in \mathbb{R}$. For each $t \in \mathbb{R}$, $\widehat{F}_n(t)$ is an unbiased estimator of $F(t)$ with variance $\text{Var}(\widehat{F}_n(t)) = \frac{1}{n}F(t)(1 - F(t))$. Consequently, $\widehat{F}_n(t)$ converges in probability to $F(t)$, and thus it is a consistent estimator. Moreover, by the law of large numbers (**Eigenschaft A06.05 (i)**) the convergence holds almost surely in any point and also uniformly, by Glivenko-Cantelli's theorem, i.e., $\|\widehat{F}_n - F\|_\infty = o(1)$ \mathbb{P} -a.s.. If we assume in addition that \mathbb{P} admits a Lebesgue density then F 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}_θ on $(\mathbb{R}, \mathcal{B})$ their *total variation distance* given by $\|\mathbb{P} - \mathbb{P}_\theta\|_{\text{TV}} := \sup\{|\mathbb{P}(B) - \mathbb{P}_\theta(B)|, B \in \mathcal{B}\}$. We note that for any probability measure $\mathbb{P} \in \mathcal{W}(\mathcal{B})$ admitting a Lebesgue-density we have $\|\mathbb{P} - \widehat{\mathbb{P}}_n\|_{\text{TV}} = 1$ \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, depending on the measure of accuracy (metric, topology, etc.) a different estimator of \mathbb{P} might be reasonable.

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

§22.02 **Proof of Lemma §22.01.** is given in the lecture. □

In the sequel let \mathcal{D} be the set of Lebesgue densities on $(\mathbb{R}, \mathcal{B})$, and hence $\mathcal{D} \subseteq \mathcal{L}_1 = \mathcal{L}_1(\mathcal{B}, \lambda)$. $\mathbb{P}_\mathbb{p} = \mathbb{p}\lambda$ 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 $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\mathbb{p}^n = (\mathbb{P}_\mathbb{p}^n)_{\mathbb{p} \in \mathcal{D}})$ and $(X_i)_{i \in \llbracket n \rrbracket} \sim \mathbb{P}_\mathbb{p}^n$. Typically, for $s \geq 1$ we assess the accuracy of an estimator $\widehat{\mathbb{p}}$ of \mathbb{p} either by a local measure, e.g. $\mathbb{P}_\mathbb{p}^n(|\widehat{\mathbb{p}}(t) - \mathbb{p}(t)|^s)$, for $t \in \mathbb{R}$, or by a global measure, e.g. $\mathbb{P}_\mathbb{p}^n(\|\widehat{\mathbb{p}} - \mathbb{p}\|_{\mathcal{L}_s}^s) = \mathbb{P}_\mathbb{p}^n(\lambda(|\widehat{\mathbb{p}} - \mathbb{p}|^s))$, with a focus on the special cases $s = 1$ and $s = 2$.

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 (Y, X) taking values in $(\mathbb{R}^2, \mathcal{B}^2)$. The joint distribution of (Y, X) is uniquely determined by the functional parameter of interest f , the conditional distribution of the error $\varepsilon := Y - f(X)$ 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 $(Y, X) \sim \mathbb{P}$. Thereby, the dependence on the marginal distribution \mathbb{P}^X of the regressor X and the conditional distribution of the error term ε given X is usually not made explicit. For sake of simplicity, suppose in addition that the joint distribution \mathbb{P} of (Y, X) admits a joint Lebesgue density \mathfrak{p} . Denoting by \mathfrak{p}^X the marginal density of X we use for the conditional density $\mathbb{P}_{Y|X}$ of Y given X the \mathbb{P} -a.s. identity $\mathfrak{p}^X \mathbb{P}_{Y|X} = \mathfrak{p}$ (see **Bemerkung C03.03 (iii)**) which allows for \mathbb{P} -a.e. $x \in \mathbb{R}$ to write

$$\begin{aligned} \mathfrak{q}(x) &:= f(x)\mathfrak{p}^X(x) = \mathbb{E}(Y|X = x)\mathfrak{p}^X(x) \\ &= \int_{\mathbb{R}} y\mathbb{P}_{Y|X=x}(y)dy\mathfrak{p}^X(x) = \int_{\mathbb{R}} y\mathfrak{p}(y, x)dy. \end{aligned} \quad (22.01)$$

Consequently, the function of interest is \mathbb{P} -a.s. given by $f = \mathfrak{q}/\mathfrak{p}^X$ which motivates the following estimation strategy. Given a sample of (Y, X) estimate separately \mathfrak{q} and \mathfrak{p}^X , say by $\widehat{\mathfrak{q}}$ and $\widehat{\mathfrak{p}}^X$, and then form an 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 (Y, X) and its expectation, respectively. We denote by $\mathbb{P}_{\mathcal{F}}$ the family of possible distributions of (Y, X) , but keep in mind, that the distribution \mathbb{P}_f of (Y, X) is generally not uniquely determined by $f \in \mathcal{F}$ only. If $((Y_i, X_i))_{i \in [n]}$ form an independent and identically distributed (i.i.d.) sample of $(Y, X) \sim \mathbb{P}_f$ then $\mathbb{P}_f^n = \bigotimes_{i \in [n]} \mathbb{P}_f$ denotes the joint product probability measure of the family $((Y_i, X_i))_{i \in [n]}$. We write $((Y_i, X_i))_{i \in [n]} \stackrel{i.i.d.}{\sim} \mathbb{P}_f$ or $((Y_i, X_i))_{i \in [n]} \sim \mathbb{P}_f^n$ for short. We denote by $\mathbb{P}_{\mathcal{F}}^n := (\mathbb{P}_f^n)_{f \in \mathcal{F}}$ the corresponding family of product probability measures. For $s \geq 1$ we assess also the accuracy of an estimator \widehat{f} of f either by a local measure, e.g. $\mathbb{P}_f^n(|\widehat{f}(t) - f(t)|^s)$, for $t \in \mathbb{R}$, or by a global measure, e.g. $\mathbb{P}_f^n(\|\widehat{f} - f\|_{\mathcal{L}_s}^s) = \mathbb{P}_f^n(\lambda(|\widehat{f} - f|^s))$ with a focus on the special cases $s = 1$ and $s = 2$.

§23 Kernel density estimation

Throughout this section we consider the statistical product model $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_{\mathcal{D}}^n = (\mathbb{P}_{\mathcal{D}}^n)_{\mathcal{D} \in \mathcal{D}})$ and let $(X_i)_{i \in [n]} \stackrel{i.i.d.}{\sim} \mathbb{P}_{\mathcal{D}} = \mathfrak{p}\lambda \in \mathcal{W}(\mathcal{B})$ be real-valued random variables with Lebesgue density $\mathfrak{p} \in \mathcal{D} \subseteq \mathcal{L}_1 = \mathcal{L}_1(\mathcal{B}, \lambda)$ and c.d.f. \mathbb{F} .

§23.01 **Definition.** A function $K \in \mathcal{L}_1$ with $\lambda(K) = 1$ is called a *kernel*. Given a *bandwidth* $b \in \mathbb{R}^+$ and an evaluation point $x_o \in \mathbb{R}$ define $K_b(x_o) \in \mathcal{L}_1$ with $x \mapsto K_b(x_o, x) := \frac{1}{b}K(\frac{x-x_o}{b})$. The statistic $\widehat{\mathfrak{p}}_b(x_o) := \widehat{\mathfrak{p}}_n K_b(x_o) \in \mathcal{B}^n$ satisfying

$$x^n = (x_i)_{i \in [n]} \mapsto \widehat{\mathfrak{p}}_b(x_o, x^n) = \frac{1}{n} \sum_{i \in [n]} K_b(x_o, x_i) = \frac{1}{nb} \sum_{i \in [n]} K(\frac{x_i - x_o}{b})$$

is called *kernel density estimator* of $\mathfrak{p}(x_o)$. □

§23.02 **Remark.** Since $\mathbb{F}(x+b) - \mathbb{F}(x-b) = \mathbb{p}\lambda([x-b, x+b])$ for any $b \in \mathbb{R}_0^+$ we have $\mathbb{F}(x+b) - \mathbb{F}(x-b) \approx \mathbb{p}(x)2b$ for b sufficiently small. Replacing the unknown \mathbb{F} by its empirical counterpart $\widehat{\mathbb{F}}_n$ Rosenblatt [1956] proposed for $\mathbb{p}(x)$ the estimator $\widehat{\mathbb{p}}_b(x) \in \mathcal{B}^n$ given by

$$\begin{aligned} x^n = (x_i)_{i \in \llbracket n \rrbracket} &\mapsto \widehat{\mathbb{p}}_b(x, x^n) := \frac{\widehat{\mathbb{F}}_n(x+b, x^n) - \widehat{\mathbb{F}}_n(x-b, x^n)}{2b} \\ &= \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \frac{1}{2b} \mathbb{1}_{(-1,1]} \left(\frac{x_i - x}{b} \right) = \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} \frac{1}{b} K \left(\frac{x_i - x}{b} \right) = \frac{1}{n} \sum_{i \in \llbracket n \rrbracket} K_b(x, x_i) \end{aligned}$$

setting $K(t) := \frac{1}{2} \mathbb{1}_{[-1,1]}(t)$ for $t \in \mathbb{R}$. Observe that K is a density, which in turn implies that $x \mapsto \widehat{\mathbb{p}}_b(x, x^n)$ is a density for each $b \in \mathbb{R}_0^+$ and $x^n \in \mathbb{R}^n$ as well. Parzen [1962] introduces a kernel K and a bandwidth b as in Definition §23.01 and studies the more general kernel density estimator $\widehat{\mathbb{p}}_b(x) = \widehat{\mathbb{P}}_n K_b(x)$, $x \in \mathbb{R}$. Note that $\lambda(\widehat{\mathbb{p}}_b) = 1$ since $\lambda(K) = 1$ by assumption. If the kernel K is in addition positive, i.e. $K \in \mathcal{B}^+$, then $\widehat{\mathbb{p}}_b$ is a density. An alternative motivation for a kernel density estimator provides the following lemma. □

§23.03 **Lemma (Bochner's lemma).** For $b \in \mathbb{R}_0^+$, $x_o \in \mathbb{R}$ and $K \in \mathcal{L}_1$ define $K_b(x_o) \in \mathcal{L}_1$ with $K_b(x_o, x) := \frac{1}{b} K(\frac{x-x_o}{b})$, $x \in \mathbb{R}$. If $g \in \mathcal{B}$ is continuous in x_o and bounded, i.e., $\|g\|_{\mathcal{L}_\infty} < \infty$, then $\lim_{b \rightarrow 0} \lambda(gK_b(x_o)) = g(x_o)\lambda(K)$.

§23.04 **Proof of Lemma §23.03.** is given in the lecture. □

§23.05 **Example.** Kernels typically considered are the rectangular kernel $K(u) := \frac{1}{2} \mathbb{1}_{[-1,1]}(u)$, the triangular kernel $K(u) := (1 - |u|) \mathbb{1}_{[-1,1]}(u)$, the Epanechnikov kernel $K(u) := \frac{3}{4}(1 - u^2) \mathbb{1}_{[-1,1]}(u)$ or the Gaussian kernel $K(u) := \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$. □

Local measure of accuracy.

§23.06 **Definition.** The mean squared error of a kernel density estimator $\widehat{\mathbb{p}}_b(x_o)$ satisfies

$$\text{mse}_p(x_o) := \mathbb{E}_p^n (|\widehat{\mathbb{p}}_b(x_o) - \mathbb{p}(x_o)|^2) = \text{var}_p(x_o) + |\text{bias}_p(x_o)|^2 \quad \text{at a point } x_o \in \mathbb{R}$$

by introducing a *variance* and a *bias* term, respectively,

$$\text{var}_p(x_o) := \mathbb{E}_p^n (|\widehat{\mathbb{p}}_b(x_o) - \mathbb{E}_p^n(\widehat{\mathbb{p}}_b(x_o))|^2) \quad \text{and} \quad \text{bias}_p(x_o) := \mathbb{E}_p^n(\widehat{\mathbb{p}}_b(x_o)) - \mathbb{p}(x_o). \quad \square$$

In the sequel we analyse separately the variance and bias term.

§23.07 **Lemma.** Let $\mathbb{p} \in \mathcal{L}_\infty$ and $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ with $\lambda(K) = 1$. For each $x_o \in \mathbb{R}$, $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$ we have $\text{var}_p(x_o) \leq (nb)^{-1} \|\mathbb{p}\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2$.

§23.08 **Proof of Lemma §23.07.** is given in the lecture. □

§23.09 **Remark.** Let $\mathbb{p} \in \mathcal{L}_\infty$ be continuous, and suppose that $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ satisfies $\lambda(K) = 1$. By Lemma §23.07 $\text{var}_p(x_o) \leq (nb)^{-1} \|\mathbb{p}\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2$. On the other hand, since $\text{bias}_p(x_o) = \lambda(\mathbb{p}K_b(x_o)) - \mathbb{p}(x_o)$ from Bochner's Lemma §23.03 follows $|\text{bias}_p(x_o)| = o(1)$ as $b \rightarrow 0$. By combining both results, we obtain for any sequence $(b_n)_{n \in \mathbb{N}}$ of bandwidths satisfying $1 = o(nb_n)$, i.e. $nb_n \rightarrow \infty$, and $b_n = o(1)$ that $\text{mse}_p(x_o) = o(1)$ as $n \rightarrow \infty$. As a consequence, the kernel density estimator is consistent, but its rate of convergence might be arbitrarily slow. Here and subsequently the bandwidth depends on n but we drop from now on the additional index n and write shortly $1 = o(nb)$ or $b = o(1)$ as $n \rightarrow \infty$. □

§23.10 **Lemma.** Let \mathbb{p} be twice-differentiable with second derivative $\ddot{\mathbb{p}} \in \mathcal{L}_\infty$. If $K, \text{id}_\mathbb{R}^2 K \in \mathcal{L}_1$, $\lambda(K) = 1$ and $\lambda(\text{id}_\mathbb{R} K) = 0$, then $|\text{bias}_\mathbb{p}(x_o)| \leq b^2 \frac{1}{2} \|\ddot{\mathbb{p}}\|_{\mathcal{L}_\infty} \lambda(\text{id}_\mathbb{R}^2 |K|)$ for all $x_o \in \mathbb{R}$, $b \in \mathbb{R}_0^+$.

§23.11 **Proof of Lemma §23.10.** is given in the lecture. \square

§23.12 **Remark.** Let $\mathbb{p} \in \mathcal{L}_\infty$ be twice-differentiable with second derivative $\ddot{\mathbb{p}} \in \mathcal{L}_\infty$ and suppose that $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ satisfies $\text{id}_\mathbb{R}^2 K \in \mathcal{L}_1$, $\lambda(K) = 1$ and $\lambda(\text{id}_\mathbb{R} K) = 0$. By combination of **Lemmata §23.07 und §23.10** follows for all $b \in \mathbb{R}_0^+$, $n \in \mathbb{N}$ and uniformly for all $x_o \in \mathbb{R}$

$$\text{mse}_\mathbb{p}(x_o) \leq (nb)^{-1} \|\mathbb{p}\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2 + b^4 \frac{1}{4} \|\ddot{\mathbb{p}}\|_{\mathcal{L}_\infty}^2 (\lambda(\text{id}_\mathbb{R}^2 |K|))^2.$$

The first and second term on the right hand side is increasing and decreasing, respectively, as b tends to zero. Therefore, let us minimise the right hand side as a function of b . Keep in mind that $M(b) := a(nb)^{-1} + cb^{2\beta}$, $b \in \mathbb{R}_0^+$, attains its minimum $M(b_o) = b \left(\frac{a}{2\beta c}\right)^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)}$ at

$b_o = \left(\frac{a}{2\beta c}\right)^{1/(2\beta+1)} n^{-1/(2\beta+1)}$. Thus, choosing $b_o = \left(\frac{\|\mathbb{p}\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2}{\|\ddot{\mathbb{p}}\|_{\mathcal{L}_\infty}^2 (\lambda(\text{id}_\mathbb{R}^2 |K|))^2}\right)^{1/5} n^{-1/5}$ we obtain

$$\sup_{x_o \in \mathbb{R}} \text{mse}_\mathbb{p}(x_o) \leq \frac{1}{4} \left(\|\ddot{\mathbb{p}}\|_{\mathcal{L}_\infty}^2 (\lambda(\text{id}_\mathbb{R}^2 |K|))^2\right)^{4/5} \left(\|\mathbb{p}\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2\right)^{1/5} n^{-4/5}.$$

We shall emphasise that the optimal bandwidth b_o depends not only on the kernel but also on characteristics of the unknown density \mathbb{p} , and hence, it is in general not feasible in practise. \square

§23.13 **Lemma.** Let $\mathbb{p} \in \mathcal{L}_\infty$ be continuous in x_o and $K \in \mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{L}_\infty$ with $\lambda(K) = 1$. If $1 = o(nb)$ and $b = o(1)$ as $n \rightarrow \infty$, then $\sqrt{nb}(\hat{\mathbb{p}}_\mathbb{p}(x_o) - \mathbb{E}_\mathbb{p}(\hat{\mathbb{p}}_\mathbb{p}(x_o))) \xrightarrow{D} N_{(0, \sigma^2)}$ with $\sigma^2 := \mathbb{p}(x_o) \lambda(K^2)$.

§23.14 **Proof of Lemma §23.13.** is given in the lecture. \square

§23.15 **Remark.** In addition to the assumptions of **Lemma §23.13** let \mathbb{p} be twice-differentiable with second derivative $\ddot{\mathbb{p}} \in \mathcal{L}_\infty$ continuous in x_o , and let $\text{id}_\mathbb{R}^2 K \in \mathcal{L}_1$ with $\lambda(\text{id}_\mathbb{R} K) = 0$. Then, $b^{-2} \text{bias}_\mathbb{p}(x_o) = \frac{1}{2} \ddot{\mathbb{p}}(x_o) \lambda(\text{id}_\mathbb{R}^2 K) + o(1)$ as $b \rightarrow 0$ by **Bochner's Lemma §23.03**. Therefore, setting $\mu := \frac{c^{5/2}}{2} \ddot{\mathbb{p}}(x_o) \lambda(\text{id}_\mathbb{R}^2 K)$ we have $\sqrt{nb} \text{bias}_\mathbb{p}(x_o) = \mu + o(1)$ as $bn^{1/5} \rightarrow c \in \mathbb{R}_0^+$, and thus $\sqrt{nb}(\hat{\mathbb{p}}_\mathbb{p}(x_o) - \mathbb{p}(x_o)) \xrightarrow{D} N_{(\mu, \sigma^2)}$ due **Lemma §23.13**. Moreover, we conclude similarly $\sqrt{nb}(\hat{\mathbb{p}}_\mathbb{p}(x_o) - \mathbb{p}(x_o)) \xrightarrow{D} N_{(0, \sigma^2)}$, if $bn^{1/5} = o(1)$. \square

§23.16 **Definition.** A kernel K satisfying in addition $\text{id}_\mathbb{R}^j K \in \mathcal{L}_1$ and $\lambda(\text{id}_\mathbb{R}^j K) = 0$ for each $j \in \llbracket l \rrbracket$ is called a *kernel of order $l \in \mathbb{N}$* . \square

§23.17 **Remark.** For arbitrary $l \in \mathbb{N}$ the construction of a kernel of order l and several examples are given, for instance, in Tsybakov [2009], section 1.2.2, or Comte [2015] section 3.2.4. \square

We denote by $\lfloor \beta \rfloor$ the greatest integer strictly less than the real number β .

§23.18 **Definition.** For $\beta, L \in \mathbb{R}_0^+$ the *Hölder class* $\mathcal{H}(\beta, L)$ on \mathbb{R} is a set of $l = \lfloor \beta \rfloor$ times differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ whose derivative $f^{(l)}$ satisfies $|f^{(l)}(x) - f^{(l)}(y)| \leq L|x - y|^{\beta-l}$ for all $x, y \in \mathbb{R}$. \square

§23.19 **Lemma.** Suppose that $\mathbb{p} \in \mathcal{H}(\beta, L)$ and let K be a kernel of order $l = \lfloor \beta \rfloor$ satisfying $|\text{id}_\mathbb{R} |^\beta K| \in \mathcal{L}_1$. Then, $|\text{bias}_\mathbb{p}(x_o)| \leq b^\beta \frac{L}{\Gamma} \lambda(|\text{id}_\mathbb{R} |^\beta |K|)$ for each $x_o \in \mathbb{R}$, $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$.

§23.20 **Proof of Lemma §23.19.** is given in the lecture. \square

§23.21 **Remark.** Let $\mathbb{p} \in \mathcal{H}(\beta, L)$ and suppose that $K \in \mathcal{L}_2$ is a kernel of order $l = \lfloor \beta \rfloor$ satisfying in addition $|\text{id}_\mathbb{R} |^\beta K| \in \mathcal{L}_1$. By combination of **Lemmata §23.07 und §23.19** we conclude that uniformly for all $x_o \in \mathbb{R}$

$$\text{mse}_\mathbb{p}(x_o) \leq (nb)^{-1} \|\mathbb{p}\|_{\mathcal{L}_\infty} \lambda(|K|^2) + b^{2\beta} \left(\frac{L}{\Gamma} \lambda(|\text{id}_\mathbb{R} |^\beta |K|)\right)^2.$$

Minimising the right hand side as a function of b leads to an optimal bandwidth $b_o = c n^{-1/(2\beta+1)}$ with constant given by $c^{2\beta+1} 2\beta (\frac{L}{h} \lambda(|\text{id}_{\mathbb{R}}|^{\beta}|K|))^2 = \|\mathbb{p}\|_{\mathcal{L}_2} \lambda(|K|)^2$. Consequently, by choosing the optimal bandwidth b_o we have $\sup_{x_o \in \mathbb{R}} \text{mse}_{\mathbb{p}}(x_o) = O(n^{-2\beta/(2\beta+1)})$. However, the optimal bandwidth b_o depends again on characteristics of the unknown density \mathbb{p} , and hence, it is generally not feasible in practise. \square

§23.22 **Theorem.** Suppose that $\mathbb{p} \in \mathcal{H}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ satisfying in addition $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}_1$. Fix $c \in \mathbb{R}_0^+$ and set $b_o := c n^{-1/(2\beta+1)}$ then for all $n \in \mathbb{N}$

$$\sup_{x_o \in \mathbb{R}} \sup_{\mathbb{p} \in \mathcal{H}(\beta, L) \cap \mathcal{D}} \mathbb{E}_{\mathbb{p}}^n |\widehat{\mathbb{p}}_{b_o}^n(x_o) - \mathbb{p}(x_o)|^2 \leq C n^{-2\beta/(2\beta+1)},$$

where $C \in \mathbb{R}_0^+$ is a constant depending only on β, L, c and on the kernel K .

§23.23 **Proof of Theorem §23.22.** is given in the lecture. \square

Global measure of accuracy.

§23.24 **Definition.** The *mean integrated squared error* of a kernel density estimator $\widehat{\mathbb{p}} \in \mathcal{L}_2$ satisfies

$$\text{mise}_{\mathbb{p}} := \mathbb{E}_{\mathbb{p}}^n (\|\widehat{\mathbb{p}} - \mathbb{p}\|_{\mathcal{L}_2}^2) = \lambda(\text{var}_{\mathbb{p}}) + \lambda(|\text{bias}_{\mathbb{p}}|^2) \quad \text{for a density } \mathbb{p} \in \mathcal{L}_2$$

using the variance and bias term as in **Definition §23.06.** \square

We study now separately the integrated variance and bias term.

§23.25 **Lemma.** Let $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ with $\lambda(K) = 1$. We have $\lambda(\text{var}_{\mathbb{p}}) \leq (nb)^{-1} \|K\|_{\mathcal{L}_2}^2$ for any density $\mathbb{p} \in \mathcal{D}$, $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$.

§23.26 **Proof of Lemma §23.25.** is given in the lecture. \square

§23.27 **Definition.** For $\beta, L \in \mathbb{R}_0^+$ the *Nikol'ski class* $\mathcal{N}(\beta, L)$ on \mathbb{R} is a set of $l = \lfloor \beta \rfloor$ times differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ whose derivative $f^{(l)}$ satisfies $\|f^{(l)}(\bullet + t) - f^{(l)}\|_{\mathcal{L}_2} \leq L|t|^{\beta-l}$ for all $t \in \mathbb{R}$. \square

§23.28 **Lemma.** Suppose that $\mathbb{p} \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let K be a kernel of order $l = \lfloor \beta \rfloor$ satisfying $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}_1$. Then we have $\|\text{bias}_{\mathbb{p}}\|_{\mathcal{L}_2} \leq b^{\beta} \frac{L}{h} \lambda(|\text{id}_{\mathbb{R}}|^{\beta}|K|)$ for each $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$.

§23.29 **Proof of Lemma §23.28.** is given in the lecture. \square

§23.30 **Remark.** Let $\mathbb{p} \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ satisfying $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}_1$. By combination of **Lemmata §23.25 und §23.28** follows

$$\text{mise}_{\mathbb{p}} \leq (nb)^{-1} \|K\|_{\mathcal{L}_2}^2 + b^{2\beta} \left(\frac{L}{h} \lambda(|\text{id}_{\mathbb{R}}|^{\beta}|K|)\right)^2.$$

Minimising the right hand side as a function of b leads to an optimal bandwidth $b_o = c n^{-1/(2\beta+1)}$ with constant given by $c^{2\beta+1} 2\beta (\frac{L}{h} \lambda(|\text{id}_{\mathbb{R}}|^{\beta}|K|))^2 = \lambda(K^2)$. Consequently, by choosing an optimal bandwidth b_o we have $\text{mise}_{\mathbb{p}} = O(n^{-2\beta/(2\beta+1)})$. However, the optimal bandwidth b_o depends again on characteristics of the unknown density \mathbb{p} , and hence, is in general not feasible in practise. \square

§23.31 **Theorem.** Suppose that $\mathbb{p} \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ satisfying in addition $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}_1$. Fix $c \in \mathbb{R}_0^+$ and set $b_o = c n^{-1/(2\beta+1)}$ then for all $n \in \mathbb{N}$

$$\mathbb{E}_{\mathbb{p}}^n \|\widehat{\mathbb{p}}_{b_o}^n - \mathbb{p}\|_{\mathcal{L}_2}^2 \leq C n^{-2\beta/(2\beta+1)},$$

where $C \in \mathbb{R}_0^+$ is a constant depending only on β, L, c and on the kernel K .

§23.32 **Proof of Theorem §23.31.** is given in the lecture. \square

Data-driven bandwidth selection.

§23.33 **Oracle choice.** Considering $\text{mise}_p(b) := \mathbb{E}_p \|\widehat{\mathbb{P}}_b - \mathbb{P}\|_{\mathcal{L}}^2$ of a kernel density estimator $\widehat{\mathbb{P}}_b$ the choice of the bandwidth b is crucial. For instance, an ideal value b_o of the bandwidth satisfies $\text{mise}_p(b_o) = \inf\{\text{mise}_p(b) : b \in \mathbb{R}_0^+\}$. Note that for a given density \mathbb{P} , the estimator $\widehat{\mathbb{P}}_b$, if b_o exists, has minimal $\text{mise}_p(b_o)$ within the family $\{\widehat{\mathbb{P}}_b : b \in \mathbb{R}_0^+\}$ of all kernel density estimators with fixed kernel and varying bandwidth. Unfortunately, $\text{mise}_p(b) = \mathbb{E}_p \|\widehat{\mathbb{P}}_b - \mathbb{P}\|_{\mathcal{L}}^2$ depends on unknown characteristics of the density \mathbb{P} . Therefore, both the bandwidth b_o and the kernel density estimator $\widehat{\mathbb{P}}_{b_o}$ remain purely theoretical and thus they are often called *oracle*. \square

§23.34 **Cross validation.** A common idea is to minimise an unbiased estimator rather than $\text{mise}_p(b) = J(b) + \lambda(\mathbb{P}^2)$ with $J(b) := \mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b^2) - 2\lambda(\mathbb{P}\widehat{\mathbb{P}}_b))$. We observe that $\lambda(\mathbb{P}^2)$ does not depend on the bandwidth b and hence, the oracle choice b_o , if it exists, satisfies $J(b_o) = \min\{J(b) : b \in \mathbb{R}_0^+\}$. To construct a unbiased estimator of $J(b)$ it is sufficient to estimate $\mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b^2))$ and $\mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b\mathbb{P}))$ without bias. Obviously, $\lambda(\widehat{\mathbb{P}}_b^2)$ is an unbiased estimator of $\mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b^2))$. For $x_o \in \mathbb{R}$, $i \in \llbracket n \rrbracket$ and $x^n = (x_i)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$ we consider $\widehat{\mathbb{P}}_b^{-i}(x_o, x^n) := \frac{1}{n-1} \sum_{j \in \llbracket n \rrbracket \setminus \{i\}} K_b(x_o, x_j)$, and $(\widehat{\mathbb{P}}_b^{-i})(x^n) = \frac{1}{n} \sum_{j \in \llbracket n \rrbracket} \widehat{\mathbb{P}}_b^{-i}(x_j, x^n)$, hence $\widehat{\mathbb{P}}_b^{-i}(x) \in \mathcal{B}^n$ and $\widehat{\mathbb{P}}_b^{-i} \in \mathcal{B}^n$, where the latter by construction is an unbiased estimator of $\mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b\mathbb{P}))$. Note that for each $i \in \llbracket n \rrbracket$ the i -th coordinate map $x^n \mapsto \Pi_i(x^n) := x_i$ and $\widehat{\mathbb{P}}_b^{-i}$ are independent in a statistical product experiment, which in turn implies, that $x^n \mapsto \widehat{\mathbb{P}}_b^{-i}(x_i, x^n)$ is an unbiased estimator of $\mathbb{E}_p^n(\lambda(\widehat{\mathbb{P}}_b\mathbb{P}))$. To summarise, for each $b \in \mathbb{R}_0^+$ the (*leave-one-out*) *cross-validation criterion* $\widehat{J}(b) := \lambda(\widehat{\mathbb{P}}_b^2) - \frac{1}{2} \widehat{\mathbb{P}}_b^{-i}(\widehat{\mathbb{P}}_b^{-i})$ is an unbiased estimator of $J(b)$, i.e., $J(b) = \mathbb{E}_p^n(\widehat{J}(b))$. Recall, that the oracle b_o minimises both $\text{mise}_p(b)$ and $\mathbb{E}_p^n\{\widehat{J}(b)\}$ over $b \in \mathbb{R}_0^+$. Therefore, a reasonable and feasible choice \widehat{b} of the bandwidth, if it exists, satisfies $\widehat{J}(\widehat{b}) = \min\{\widehat{J}(b), b \in \mathbb{R}_0^+\}$. Finally, we define the cross-validation estimator $\widehat{\mathbb{P}}_{\widehat{b}}$. Note that $\widehat{\mathbb{P}}_{\widehat{b}}$ is a kernel density estimator with random bandwidth \widehat{b} depending on the sample only. Under appropriate conditions the *mise* of the estimator $\widehat{\mathbb{P}}_{\widehat{b}}$ is asymptotically equivalent to that of the oracle kernel density (pseudo)-estimator $\widehat{\mathbb{P}}_{b_o}$. \square

§24 Nonparametric regression by local smoothing

Here and subsequently, we consider a statistical product experiment $(\mathbb{R}^{2n}, \mathcal{B}^{2n}, \mathbb{P}^n)$ as introduced in Abschnitt §22. Let $((Y_i, X_i))_{i \in \llbracket n \rrbracket}$ be an i.i.d. sample of $(Y, X) \sim \mathbb{P}_f$. Introducing the coordinate maps $\Pi_Y, \Pi_X \in \mathcal{B}^2$ with $(y, x) \mapsto \Pi_Y(y, x) := y$ and $(y, x) \mapsto \Pi_X(y, x) := x$ we tactically identify Y and X with Π_Y and Π_X , respectively, and thus, (Y, X) with the identity $\text{id}_{\mathbb{R}^2}$. We denote by \mathbb{P}^X the marginal distribution of the regressor X and by $\mathbb{E}_f(Y|X)$ a conditional expectation of Y given X (see Kapitel C).

§24.01 **Assumption.** The random vector $(Y, X) \in (\mathcal{B}^2)^2$ obeys \mathbb{P}^X -a.e. a nonparametric regression model $\mathbb{E}_f(Y|X) = f$ for some unknown regression function $f \in \mathcal{F}$.

(NPR1) The error term $\varepsilon := Y - f(X)$ has a finite second moment, i.e., $\varepsilon \in \mathcal{L}_2(\mathbb{P})$, and hence, $\mathbb{E}_f(\varepsilon) = 0$. We set $\sigma_\varepsilon^2 := \mathbb{E}_f(\varepsilon^2)$. The error term ε and the explanatory variable X are independent.

(NPR2) The joint distribution \mathbb{P}_f of (Y, X) admits a joint Lebesgue density $\mathbb{P} \in (\mathcal{B}^2)^+$, i.e. $\mathbb{P} = d\mathbb{P}_f/d\lambda^2$ and $\mathbb{P}_f = \mathbb{P}\lambda^2$. Denote by \mathbb{P}^X the marginal density of X . Using for the conditional density $\mathbb{P}^{Y|X}$ of Y given X the \mathbb{P}_f -a.s. identity $\mathbb{P}^X \mathbb{P}^{Y|X} = \mathbb{P}$ (see **Bemerkung C03.03 (iii)**) define $\mathbb{Q} := f\mathbb{P}^X$ as in (22.01). \square

§24.02 **Heuristics.** Given a *bandwidth* $b \in \mathbb{R}_0^+$ and *evaluation points* $y_o, x_o \in \mathbb{R}$ define $K_b(y_o)$ and $K_b(x_o)$ as in [Definition §23.01](#). The statistic $\widehat{\mathbb{p}}_b(y_o, x_o) := \widehat{\mathbb{P}}_n(K_b(y_o, Y)K_b(x_o, X)) \in \mathcal{B}^{2n}$,

$$(y, x)^n = ((y_i, x_i))_{i \in \llbracket n \rrbracket} \mapsto \widehat{\mathbb{p}}_b(y_o, x_o, (y, x)^n) = \frac{1}{n} \sum_{j \in \llbracket n \rrbracket} K_b(y_o, y_j) K_b(x_o, x_j)$$

is a kernel density estimator of the joint density $\mathbb{p} \in \mathcal{L}_1(\lambda^2)$ evaluated at (y_o, x_o) . Exploiting $\lambda(K) = 1$ the marginal $\widehat{\mathbb{p}}_b^X(x_o) := \widehat{\mathbb{P}}_n K_b(x_o, X) = \int_{\mathbb{R}} \widehat{\mathbb{p}}_b(y_o, x_o) dy_o \in \mathcal{B}^n$ is a kernel density estimator of the marginal density $\mathbb{p}^X \in \mathcal{L}_1$ evaluated at $x_o \in \mathbb{R}$. Keeping (22.01) in mind we estimate $\mathfrak{q}(x_o)$ by replacing the unknown density \mathbb{p} by its kernel estimator $\widehat{\mathbb{p}}_b$, that is, $\widehat{\mathfrak{q}}_b(x_o) := \int_{\mathbb{R}} y_o \widehat{\mathbb{p}}_b(y_o, x_o) dy_o \in \mathcal{B}^{2n}$. If the kernel K is in addition of order 1, i.e. $\lambda(\text{id}_{\mathbb{R}} K) = 0$, then we have $\widehat{\mathfrak{q}}_b(x_o) = \widehat{\mathbb{P}}_n(Y K_b(x_o, X))$ where $\widehat{\mathfrak{q}}_b(x_o, (y, x)^n) = \frac{1}{n} \sum_{j \in \llbracket n \rrbracket} y_j K_b(x_o, x_j)$ for all $(y, x)^n = ((y_i, x_i))_{i \in \llbracket n \rrbracket} \in \mathbb{R}^{2n}$. \square

§24.03 **Definition.** Given a kernel K of order 1, a bandwidth $b \in \mathbb{R}_0^+$, and an evaluation point $x_o \in \mathbb{R}$ the statistic $\widehat{f}_b(x_o) := \frac{\widehat{\mathfrak{q}}_b(x_o)}{\widehat{\mathbb{p}}_b^X(x_o)} \mathbf{1}_{\mathbb{R}_0^+}(|\widehat{\mathbb{p}}_b^X(x_o)|) \in \mathcal{B}^{2n}$ defined for all $(y, x)^n \in \mathbb{R}^{2n}$ by

$$\widehat{f}_b(x_o, (y, x)^n) = \frac{\widehat{\mathfrak{q}}_b(x_o, (y, x)^n)}{\widehat{\mathbb{p}}_b^X(x_o, (y, x)^n)} = \sum_{j \in \llbracket n \rrbracket} y_j \frac{K_b(x_o, x_j)}{\sum_{j \in \llbracket n \rrbracket} K_b(x_o, x_j)}, \quad \text{if } |\widehat{\mathbb{p}}_b^X(x_o, (y, x)^n)| \in \mathbb{R}_0^+$$

and $\widehat{f}_b(x_o, (y, x)^n) = 0$ otherwise, is called *Nadaraya–Watson estimator* of $f(x_o)$. \square

Local measure of accuracy.

§24.04 **Remark.** We make use of the properties of a kernel density estimator derived in [Abschnitt §23](#) in order to analyse the estimator $\widehat{\mathbb{p}}_b^X$ of \mathbb{p}^X . As a consequence, it remains to consider the estimator $\widehat{\mathfrak{q}}_b$ of \mathfrak{q} . We consider first its mean squared error at a given point $x_o \in \mathbb{R}$, that is,

$$\text{mse}_{\mathfrak{q}}(x_o) = \mathbb{E}_f^n |\widehat{\mathfrak{q}}_b(x_o) - \mathfrak{q}(x_o)|^2 = \text{var}_{\mathfrak{q}}(x_o) + |\text{bias}_{\mathfrak{q}}(x_o)|^2$$

by introducing a *variance* and a *bias* term, respectively,

$$\text{var}_{\mathfrak{q}}(x_o) := \mathbb{E}_f^n |\widehat{\mathfrak{q}}_b(x_o) - \mathbb{E}_f^n \widehat{\mathfrak{q}}_b(x_o)|^2 \quad \text{and} \quad \text{bias}_{\mathfrak{q}}(x_o) := \mathbb{E}_f^n \widehat{\mathfrak{q}}_b(x_o) - \mathfrak{q}(x_o).$$

In the sequel we analyse separately the variance and bias term. \square

§24.05 **Lemma.** Under [Assumption §24.01](#) let $f, \mathbb{p}^X \in \mathcal{L}_{\infty}$ and $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ with $\lambda(K) = 1$. For each $x_o \in \mathbb{R}$, $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$ we have $\text{var}_{\mathfrak{q}}(x_o) \leq (nb)^{-1} (\|f\|_{\mathcal{L}_{\infty}}^2 + \sigma_{\varepsilon}^2) \|\mathbb{p}^X\|_{\mathcal{L}_{\infty}} \|K\|_{\mathcal{L}_2}^2$.

§24.06 **Proof of Lemma §24.05.** is given in the lecture. \square

Recall the [Definitionen §23.16 und §23.18](#) of a Hölder class and a higher order kernel.

§24.07 **Corollary.** Under [Assumption §24.01](#) let $\mathfrak{q} \in \mathcal{H}(\beta, L)$ and let K be a kernel of order $l = \lfloor \beta \rfloor$ with $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}$. Then, $|\text{bias}_{\mathfrak{q}}(x_o)| \leq b^{\beta} \frac{L}{l!} \lambda(|\text{id}_{\mathbb{R}}|^{\beta} |K|)$ for each $x_o \in \mathbb{R}$, $b \in \mathbb{R}_0^+$, $n \in \mathbb{N}$.

§24.08 **Proof of Corollary §24.07.** Due to the identity $\text{bias}_{\mathfrak{q}}(x_o) = \lambda(K_b(x_o)\mathfrak{q}) - \mathfrak{q}(x_o)$ the assertion follows immediately from [Lemma §23.19](#) (replace the density \mathbb{p} by \mathfrak{q}). \square

§24.09 **Remark.** We note that $f, \mathbb{p}^X \in \mathcal{L}_{\infty}$ implies $\mathfrak{q} = f\mathbb{p}^X \in \mathcal{L}_{\infty}$. Suppose that $\mathfrak{q} \in \mathcal{H}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ with $|\text{id}_{\mathbb{R}}|^{\beta} K \in \mathcal{L}_1$. Combining [Lemma §24.05 und Corollary §24.07](#) we have

$$\sup_{x_o \in \mathbb{R}} \text{mse}_{\mathfrak{q}}(x_o) \leq (nb)^{-1} (\|f\|_{\mathcal{L}_{\infty}}^2 + \sigma_{\varepsilon}^2) \|\mathbb{p}^X\|_{\mathcal{L}_{\infty}} \|K\|_{\mathcal{L}_2}^2 + b^{2\beta} \left(\frac{L}{l!} \lambda(|\text{id}_{\mathbb{R}}|^{\beta} |K|) \right)^2. \quad (24.01)$$

Suppose further that $\mathbb{p}^X \in \mathcal{H}(\beta, L)$, then combining **Lemmata** §23.07 und §23.19 an upper bound of $\text{mse}_{\mathbb{p}^X}(x_o) := \mathbb{E}_f^n |\widehat{\mathbb{p}}^X(x_o) - \mathbb{p}^X(x_o)|^2$ is given by (see **Remark** §23.21)

$$\sup_{x_o \in \mathbb{R}} \text{mse}_{\mathbb{p}^X}(x_o) \leq (nb)^{-1} \|\mathbb{p}^X\|_{\mathcal{L}_\infty} \|K\|_{\mathcal{L}_2}^2 + b^{2\beta} \left(\frac{L}{l} \lambda(|\text{id}_{\mathbb{R}}|^\beta |K|) \right)^2. \quad (24.02)$$

Therefore minimising the right hand side in Gleichungen (24.01) und (24.02) as a function of b leads to an optimal bandwidth $b_o = cn^{-1/(2\beta+1)}$ with constant $c \in \mathbb{R}_0^+$ depending on f , \mathbb{p}^X and K . \square

§24.10 **Proposition.** Under **Assumption** §24.01 suppose that $\mathbb{q}, \mathbb{p}^X \in \mathcal{H}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ with $|\text{id}_{\mathbb{R}}|^\beta K \in \mathcal{L}$. Fix $c \in \mathbb{R}_0^+$ and set $b_o = cn^{-1/(2\beta+1)}$. If $\mathbb{p}^X(x_o) > 0$ then $|\widehat{f}_{b_o}(x_o) - f(x_o)|^2 = O_{\mathbb{P}^n}(n^{-2\beta/(2\beta+1)})$.

§24.11 **Proof** of **Proposition** §24.10. is given in the lecture. \square

§24.12 **Remark.** It is straightforward to show that under similar assumption as used in **Lemma** §23.13 the asymptotic normality of $\widehat{\mathbb{q}}_b(x_o)$ holds true, which due to Slutsky's lemma (**Eigenschaft** A06.03) allows then to establish the asymptotic normality of $\widehat{f}_b(x_o)$. \square

Global measure of accuracy.

§24.13 **Remark.** We make use of the properties of a kernel density estimator derived in Abschnitt §23 in order to analyse the mean integrated squared error of the estimator $\widehat{\mathbb{p}}_b^X$ under the additional assumption $\mathbb{p}^X \in \mathcal{L}_2$. As a consequence, it remains to study the estimator $\widehat{\mathbb{q}}_b$ of $\mathbb{q} \in \mathcal{L}_2$, where

$$\text{mise}_{\mathbb{q}} = \mathbb{E}_f^n \|\widehat{\mathbb{q}}_b - \mathbb{q}\|_{\mathcal{L}_2}^2 = \lambda(\text{var}_{\mathbb{q}}) + \lambda(|\text{bias}_{\mathbb{q}}|^2)$$

using the variance and bias term as in **Remark** §24.04. \square

We study now separately the integrated variance and bias term.

§24.14 **Lemma.** Under **Assumption** §24.01 let $f \in \mathcal{L}_2(\mathbb{P}^X)$ and $K \in \mathcal{L}_1 \cap \mathcal{L}_2$ with $\lambda(K) = 1$. For all $b \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$ we have $\lambda(\text{var}_{\mathbb{q}}) \leq (nb)^{-1} \sigma_Y^2 \|K\|_{\mathcal{L}_2}^2$ with $\sigma_Y^2 := \mathbb{E}_f(Y^2) = \mathbb{P}^X(f^2) + \sigma_\varepsilon^2$.

§24.15 **Proof** of **Lemma** §24.14. is given in the lecture. \square

Recall the **Definitionen** §23.16 und §23.27 of a Nikol'ski class and a higher order kernel.

§24.16 **Corollary.** Under **Assumption** §24.01 let $\mathbb{q} \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let K be a kernel of order $l = \lfloor \beta \rfloor$ with $|\text{id}_{\mathbb{R}}|^\beta K \in \mathcal{L}$. Then, $\|\text{bias}_{\mathbb{q}}\|_{\mathcal{L}_2} \leq b^\beta \frac{L}{l} \lambda(|\text{id}_{\mathbb{R}}|^\beta |K|)$ for all $b \in \mathbb{R}_{\nu_o}^+$, $n \in \mathbb{N}$.

§24.17 **Proof** of **Corollary** §24.16. Making use of the identity $\text{bias}_{\mathbb{q}}(x_o) = \lambda(K_b(x_o)\mathbb{q}) - \mathbb{q}(x_o)$ and replacing \mathbb{q} by the density \mathbb{p} the assertion follows immediately from **Lemma** §23.28. \square

§24.18 **Remark.** Let $f \in \mathcal{L}_2(\mathbb{P}^X)$, $\mathbb{q} \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ satisfying $|\text{id}_{\mathbb{R}}|^\beta K \in \mathcal{L}_1$. Combining **Lemma** §24.14 und **Corollary** §24.16 we have

$$\text{mise}_{\mathbb{q}} \leq (nb)^{-1} \sigma_Y^2 \|K\|_{\mathcal{L}_2}^2 + b^{2\beta} \left(\frac{L}{l} \lambda(|\text{id}_{\mathbb{R}}|^\beta |K|) \right)^2. \quad (24.03)$$

Suppose further that $\mathbb{p}^X \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$, then combining **Lemmata** §23.25 und §23.28 an upper bound of $\text{mise}_{\mathbb{p}^X} := \mathbb{E}_f^n \|\widehat{\mathbb{p}}_b^X - \mathbb{p}^X\|_{\mathcal{L}_2}^2$ is given by (see **Remark** §23.30)

$$\text{mise}_{\mathbb{p}^X} \leq (nb)^{-1} \|K\|_{\mathcal{L}_2}^2 + b^{2\beta} \left(\frac{L}{l} \lambda(|\text{id}_{\mathbb{R}}|^\beta |K|) \right)^2. \quad (24.04)$$

Therefore minimising the right hand side in Gleichungen (24.03) und (24.04) as a function of b leads to an optimal bandwidth $b_o = cn^{-1/(2\beta+1)}$ with constant $c \in \mathbb{R}_0^+$ depending on f , \mathbb{p}^X and K . \square

In order to derive an upper bound for the mise of \hat{f}_b we use in the next assertion a regularised version which makes use of a stronger assumption, that is, $\mathbb{p}^X(x) > p_o$ for all $x \in A$, for some known constant $p_o > 0$ and measurable support $A \in \mathcal{B}$.

§24.19 **Proposition.** Under *Assumption §24.01* suppose that $f \in \mathcal{L}_2(\mathbb{P}^X)$, $\mathbb{q}, \mathbb{p}^X \in \mathcal{L}_2 \cap \mathcal{N}(\beta, L)$ and let $K \in \mathcal{L}_2$ be a kernel of order $l = \lfloor \beta \rfloor$ satisfying $|\text{id}_{\mathbb{R}}|^\beta K \in \mathcal{L}_1$. Assume in addition that $\mathbb{p}^X(x) > p_o$ for all $x \in A$, for some known constant $p_o > 0$ and set $A \in \mathcal{B}$. Consider the regularised Nadaraya–Watson estimator $\hat{f}_b^o(x) := \frac{\hat{\mathbb{q}}_b(x)}{\hat{\mathbb{p}}^X(x)} \mathbb{1}_{\{\hat{\mathbb{p}}^X(x) > p_o/2\}}$ for all $x \in A$. Fix $c \in \mathbb{R}_0^+$ and set $b_o = cn^{-1/(2\beta+1)}$ then for all $n \in \mathbb{N}$

$$\mathbb{E}_f^n \|(\hat{f}_b^o - f)\mathbb{1}_A\|_{\mathcal{L}_2}^2 \leq Cn^{-2\beta/(2\beta+1)},$$

where $C \in \mathbb{R}_0^+$ is a constant depending only on β, L, c, p_o and on the kernel K .

§24.20 **Proof of Proposition §24.19.** is given in the lecture. □

Local polynomial estimators.

§24.21 **Heuristics.** Let the kernel K take only values in \mathbb{R}^+ . It is then easily verified, that the Nadaraya–Watson estimator $\hat{f}_b \in \mathcal{B}^{2n}$ as in *Definition §24.03* satisfies

$$\hat{f}_b(x_o, (y, x)^n) \in \arg \inf_{\theta \in \mathbb{R}} \sum_{j \in [n]} (y_j - \theta)^2 K_b(x_o, x_j).$$

Therefore, \hat{f}_b is obtained by a local constant least squares approximation of the responses $\{y_i\}$. The locality is determined by the kernel K that downweights all the x_i that are not close to x_o whereas θ plays the role of a local constant to be fitted. More generally, we may define a local polynomial least squares approximation, replacing the constant θ by a polynomial of a pre-specified degree. □

§24.22 **Definition.** For $m \in \mathbb{R}$ consider $U : \mathbb{R} \rightarrow \mathbb{R}^{l+1}$, $z \mapsto U(z) = (1, z, z^2/2!, \dots, z^m/m!)$. Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a kernel and $b \in \mathbb{R}_0^+$ be a bandwidth. A random vector $\hat{\theta}(x_o) \in \mathcal{B}^{m+1}$ satisfying

$$\hat{\theta}(x_o, (y, x)^n) \in \arg \inf_{\theta \in \mathbb{R}^{m+1}} \sum_{j \in [n]} (y_j - \theta^t U(\frac{x_j - x_o}{b}))^2 K_b(x_o, x_j).$$

is called a *local polynomial estimator of order m* of $\theta(x_o) = (f(x_o), b\dot{f}(x_o), \dots, b^m f^{(m)}(x_o))$. The statistic $\hat{f}_b(x_o) = U^t(0)\hat{\theta}(x_o)$ is called *local polynomial estimator of order m* of $f(x_o)$. □

§24.23 **Remark.** Note that $\hat{f}_b(x_o)$ is simply the first coordinate of the vector $\hat{\theta}(x_o)$. Obviously, the Nadaraya–Watson estimator with non-negative kernel is just a local polynomial estimator of order zero. Furthermore, properly normalised coordinates of $\hat{\theta}(x_o)$ provide estimators of the derivatives $\dot{f}(x_o), \ddot{f}(x_o), \dots, f^{(m)}(x_o)$. For theoretical properties of local polynomial estimators and their detailed discussion we refer to Tsybakov [2009], section 1.6. □

A Grundlagen der Wahrscheinlichkeitstheorie

In diesem Kapitel werden Begriffe, Notationen und Aussagen der Vorlesung *Einführung in die Wahrscheinlichkeitstheorie und Statistik (EWS)* wiederholt. Eine detaillierte Einführung findet sich zum Beispiel in *Georgii [2015], Krenzel [2005] oder Klenke [2020]*.

A01 Wahrscheinlichkeitsraum

A01.01 Definition.

- (a) Eine nicht-leere Menge Ω aller möglichen Ausgänge eines zufälligen Experiments wird *Ergebnismenge* (*Grundmenge* oder *Stichprobenraum*) genannt. Ein möglicher Versuchsausgang ω des zufälligen Experiments, also ein Element von Ω , kurz $\omega \in \Omega$ heißt *Ergebnis* (*Stichprobe*).
- (b) Ein *Ereignis* ist eine Teilmenge der Grundmenge Ω , also ein Element der Potenzmenge 2^Ω von Ω . Für einen Versuchsausgang $\omega \in \Omega$ ist ein Ereignis $A \in 2^\Omega$ eingetreten, wenn $\omega \in A$.
- (c) Ein Paar (Ω, \mathcal{A}) wird *messbarer Raum* genannt, wenn die *Menge der interessierenden Ereignisse* $\mathcal{A} \subseteq 2^\Omega$ eine σ -Algebra ist, also falls gilt: (i) $\Omega \in \mathcal{A}$; (ii) für alle $A \in \mathcal{A}$ gilt $A^c := \Omega \setminus A \in \mathcal{A}$; und (iii) für alle $A_n \in \mathcal{A}$ mit $n \in \mathbb{N}$ gilt $(\bigcup_{n \in \mathbb{N}} A_n) \in \mathcal{A}$. \square

A01.02 Beispiel.

- (a) Auf jeder nicht-leeren Ergebnismenge Ω existieren die *triviale σ -Algebra* $\{\emptyset, \Omega\}$ sowie die Potenzmenge 2^Ω als σ -Algebren.
- (b) Sei $\mathcal{E} \subseteq 2^\Omega$ ein System von Teilmengen von Ω . Dann heißt die kleinste σ -Algebra auf Ω , die \mathcal{E} enthält, $\sigma(\mathcal{E}) := \bigcap \{ \mathcal{A} \mid \mathcal{A} \text{ ist } \sigma\text{-Algebra auf } \Omega \text{ und } \mathcal{E} \subseteq \mathcal{A} \}$, die von \mathcal{E} *erzeugte σ -Algebra* auf Ω . \mathcal{E} heißt *Erzeuger* von $\sigma(\mathcal{E})$.
- (c) Für jedes nicht-leere $A \subseteq \Omega$ gilt $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$.
- (d) Für eine höchstens abzählbar unendliche (d.h. endlich oder abzählbar unendliche), kurz *abzählbare*, Indexmenge \mathcal{I} sei $\mathcal{E} = \{A_i \in 2^\Omega \setminus \{\emptyset\}, i \in \mathcal{I}, \text{ paarweise disjunkt und } \biguplus_{i \in \mathcal{I}} A_i = \Omega\}$ eine Partition von Ω . Dann ist $\sigma(\mathcal{E}) = \{ \biguplus_{j \in \mathcal{J}} A_j \mid \mathcal{J} \in 2^{\mathcal{I}} \}$.
- (e) Es sei \mathcal{S} ein metrischer (oder topologischer) Raum und \mathcal{O} das System der offenen Teilmengen von \mathcal{S} . Dann heißt die von den offenen Mengen \mathcal{O} erzeugte σ -Algebra $\mathcal{B}_{\mathcal{S}} := \sigma(\mathcal{O})$ die *Borel- σ -Algebra* über \mathcal{S} . Die Elemente von $\mathcal{B}_{\mathcal{S}}$ heißen *Borel-Mengen*.
- (f) $\mathcal{B}^n := \mathcal{B}_{\mathbb{R}^n}$ bezeichnet die Borel- σ -Algebra über \mathbb{R}^n versehen mit dem Euklidischen Abstand $d(x, y) = \|x - y\| := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ für $x = (x_i)_{i \in [n]}, y = (y_i)_{i \in [n]} \in \mathbb{R}^n$ mit $[n] := [1, n] \cap \mathbb{N}$.
- (g) Seien (Ω, \mathcal{A}) ein messbarer Raum und $B \subseteq \Omega$ eine nicht-leere Teilmenge. Dann heißt die σ -Algebra über B , $\mathcal{A}|_B := \mathcal{A} \cap B := \{A \cap B \mid A \in \mathcal{A}\}$ *Spur* von \mathcal{A} über B , wobei für $\mathcal{E} \in 2^\Omega$ gilt $\sigma(\mathcal{E})|_B = \sigma(\mathcal{E}|_B)$. \square

A01.03 Schreibweise.

- (i) Wir setzen $\mathbb{R}^+ := [0, \infty)$, $\mathbb{R}_0^+ := (0, \infty)$, $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$, $\overline{\mathbb{R}} := [-\infty, \infty]$, $\overline{\mathbb{R}}^+ := [0, \infty]$.
- (ii) Für $a, b \in \mathbb{R}^n$ schreiben wir $a < b$, wenn $a_i < b_i$ für alle $i \in \llbracket n \rrbracket$ gilt. Für $a < b$, definieren wir den offenen *Quader* als das Kartesische Produkt $(a, b) := \prod_{i=1}^n (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$. Analog, sind $[a, b]$, $[a, b)$ sowie $(a, b]$ definiert. Weiterhin, sei $(-\infty, b) := \prod_{i=1}^n (-\infty, b_i)$ und analog $(-\infty, b]$ definiert.
- (iii) Wir bezeichnen mit $\overline{\mathcal{B}} := \mathcal{B}_{\overline{\mathbb{R}}}$ die Borel- σ -Algebra über der kompaktifizierten Zahlengerade $\overline{\mathbb{R}}$, wobei die Mengen $\{-\infty\}$, $\{\infty\}$ und \mathbb{R} in $\overline{\mathbb{R}}$ abgeschlossen bzw. offen, also Borel-Mengen sind. Insbesondere, ist $\mathcal{B} := \mathcal{B}_{\mathbb{R}} = \overline{\mathcal{B}} \cap \mathbb{R}$ die Borel- σ -Algebra über \mathbb{R} . Weiterhin schreiben wir $\overline{\mathcal{B}}^+ := \overline{\mathcal{B}} \cap \overline{\mathbb{R}}^+$, $\mathcal{B}^+ := \mathcal{B} \cap \mathbb{R}^+$, $\mathcal{B}_0^+ := \mathcal{B} \cap \mathbb{R}_0^+$ und $[0, 1]_0^+ := \mathcal{B} \cap [0, 1]$.
- (iv) Ein Folge $(x_n)_{n \in \mathbb{N}}$ aus \mathbb{R} heißt *monoton wachsend* (bzw. *fallend*), wenn $x_n \leq x_{n+1}$ (bzw. $x_{n+1} \leq x_n$) für alle $n \in \mathbb{N}$ gilt. Ist eine monotone wachsende (bzw. fallende) Folge konvergent, etwa $x = \lim_{n \rightarrow \infty} x_n$, so schreiben wir kurz $x_n \uparrow x$ (bzw. $x_n \downarrow x$).
- (v) Sei $A_n \subseteq \Omega$ für $n \in \mathbb{N}$. Die Folge $(A_n)_{n \in \mathbb{N}}$ heißt *monoton wachsend* (bzw. *fallend*), wenn $A_n \subseteq A_{n+1}$ (bzw. $A_{n+1} \subseteq A_n$) für alle $n \in \mathbb{N}$ gilt. Weiterhin heißen $A_* := \liminf_{n \rightarrow \infty} A_n := \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m$ und $A^* := \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m$ *Limes inferior* bzw. *Limes superior* der Folge $(A_n)_{n \in \mathbb{N}}$. Die Folge $(A_n)_{n \in \mathbb{N}}$ heißt *konvergent*, wenn $A_* = A^* =: A$ gilt. In diesem Fall schreiben wir kurz $\lim_{n \rightarrow \infty} A_n = A$. □
- (vi) Jede monoton wachsende (bzw. fallende) Folge $(A_n)_{n \in \mathbb{N}}$ von Teilmengen von Ω ist konvergent mit $A := \lim_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} A_n$ (bzw. $A := \lim_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} A_n$). In diesem Fall schreiben wir kurz $A_n \uparrow A$ (bzw. $A_n \downarrow A$). □

A01.04 **Definition.** Sei (Ω, \mathcal{A}) ein messbarer Raum.

- (a) Eine Abbildung $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsmaß* (kurz *Verteilung*) auf \mathcal{A} , wenn sie folgenden Bedingungen genügt: (i) $\mathbb{P}(\Omega) = 1$ (*normiert*); und (ii) $\mathbb{P}(\bigsqcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ für jede Folge $(A_n)_{n \in \mathbb{N}}$ paarweise disjunkter Ereignisse aus \mathcal{A} , d.h. $A_n \cap A_m = \emptyset$ für alle $n, m \in \mathbb{N}$ mit $n \neq m$ (*σ -additiv*).
- (b) Wir bezeichnen mit $\mathcal{W} := \mathcal{W}(\mathcal{A})$ die Menge aller Wahrscheinlichkeitsmaße auf \mathcal{A} . Ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ bestehend aus einer Ergebnismenge Ω , einer σ -Algebra \mathcal{A} interessierender Ereignisse sowie einem Wahrscheinlichkeitsmaß \mathbb{P} auf \mathcal{A} wird *Wahrscheinlichkeitsraum* genannt.
- (c) Für ein Wahrscheinlichkeitsmaß \mathbb{P} auf (Ω, \mathcal{A}) heißt ein Element $\omega \in \Omega$ mit $\{\omega\} \in \mathcal{A}$ und $\mathbb{P}(\{\omega\}) > 0$ *Atom*. Die Wahrscheinlichkeit $\mathbb{P}(\{\omega\})$ wird *Masse* des Atoms ω genannt.
- (d) Für ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}, \mathcal{B})$ wird $\mathbb{F} : \mathbb{R} \rightarrow [0, 1]$ mit $x \mapsto \mathbb{F}(x) := \mathbb{P}((-\infty, x])$ die zugehörige *Verteilungsfunktion* genannt.
- (e) Wenn Ω eine abzählbare Menge ist, dann wird ein Wahrscheinlichkeitsmaß \mathbb{P} auf $\mathcal{A} = 2^\Omega$ *diskret* genannt und $(\Omega, \mathcal{A}, \mathbb{P})$ heißt *diskreter Wahrscheinlichkeitsraum*. Die Abbildung $\mathbb{p} : \Omega \rightarrow [0, 1]$ mit $\omega \mapsto \mathbb{p}(\omega) := \mathbb{P}(\{\omega\})$ wird *Zähldichte* des Wahrscheinlichkeitsmaßes \mathbb{P} genannt.
- (f) Ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *stetig*, wenn eine *Wahrscheinlichkeitsdichte* (kurz *Dichte*) \mathbb{f} auf \mathbb{R}^n , also eine *Lebesgue-integrierbare Funktion* $\mathbb{f} : \mathbb{R}^n \rightarrow \mathbb{R}^+$ mit $\mathbb{P}(B) = \int_B \mathbb{f}(x) dx$ für alle $B \in \mathcal{B}^n$ existiert. $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P})$ wird dann *stetiger Wahrscheinlichkeitsraum* genannt. □

A01.05 **Eigenschaft.**

- (i) Ist $(\mathbb{P}_n)_{n \in \mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaße in \mathcal{W} , so ist auch jede Konvexkombination $\sum_{n \in \mathbb{N}} w_n \mathbb{P}_n$ mit $w_n \geq 0$, $n \in \mathbb{N}$, und $\sum_{n \in \mathbb{N}} w_n = 1$ in \mathcal{W} . Die Diracmaße bilden Extrempunkte der konvexen Menge aller Wahrscheinlichkeitsmaße.
- (ii) Ein Wahrscheinlichkeitsmaß \mathbb{P} auf (Ω, \mathcal{A}) besitzt abzählbar viele Atome.
- (iii) Ist eine Abbildung $\mathbb{F} : \mathbb{R} \rightarrow [0, 1]$ monoton wachsend, rechtsstetig mit $\lim_{x \rightarrow -\infty} \mathbb{F}(x) = 0$ und $\lim_{x \rightarrow \infty} \mathbb{F}(x) = 1$, so existiert genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}, \mathcal{B})$ mit $\mathbb{P}((x, y]) = \mathbb{F}(y) - \mathbb{F}(x)$ für alle $x, y \in \mathbb{R}$ mit $x < y$. Insbesondere ist \mathbb{F} die Verteilungsfunktion von \mathbb{P} . \square

A01.06 **Beispiel.**

- (a) Sei (Ω, \mathcal{A}) ein messbarer Raum. Für $A \subseteq \Omega$ bezeichnet $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ mit $\mathbb{1}_A^{-1}(\{1\}) = A$ und $\mathbb{1}_A^{-1}(\{0\}) = A^c$ die *Indikatorfunktion* auf A . Für jedes $\omega \in \Omega$ ist das *Einpunkt-* oder *Diracmaß* $\delta_\omega : \mathcal{A} \rightarrow \{0, 1\}$ mit $\delta_\omega(A) := \mathbb{1}_A(\omega)$ für alle $A \in \mathcal{A}$ ein Wahrscheinlichkeitsmaß aus $\mathcal{W}(\mathcal{A})$.
- (b) Sei $\emptyset \neq \Omega \subseteq \mathbb{R}$ abzählbar und \mathbb{p} eine Zähldichte auf Ω . Das Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}, \mathcal{B})$ mit $\mathbb{P}(B) := \sum_{\omega \in \Omega} \mathbb{p}(\omega) \delta_\omega$ für $B \in \mathcal{B}$ und die zugehörige Verteilungsfunktion $\mathbb{F} : \mathbb{R} \rightarrow [0, 1]$ mit $\mathbb{F}(x) = \mathbb{P}((-\infty, x]) = \sum_{\omega \in \Omega} \mathbb{p}(\omega) \delta_\omega((-\infty, x]) =: \sum_{\omega \leq x} \mathbb{p}(\omega)$ für $x \in \mathbb{R}$ werden *diskret* genannt.
- (c) *Laplace-/Gleich-Verteilung*, kurz Lap_Ω , mit $|\Omega| < \infty$ und $\mathbb{p}_{\text{Lap}_\Omega}(\omega) = \frac{1}{|\Omega|}$ für $\omega \in \Omega$.
- (d) *Bernoulli-Schema*, kurz B_p^n , mit Erfolgswahrscheinlichkeit $p \in [0, 1]$, Länge $n \in \mathbb{N}$, $\Omega = \{0, 1\}^n$ und $\mathbb{p}_{\text{B}_p^n}(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$ für $\omega = (\omega_i)_{i \in [n]} \in \{0, 1\}^n$. *Bernoulli-Verteilung*. Im Fall $n = 1$, schreiben wir kurz B_p , also $\mathbb{p}_{\text{B}_p}(\omega) = p^\omega (1-p)^{1-\omega}$ für $\omega \in \{0, 1\}$.
- (e) *Binomial-Verteilung*, kurz $\text{Bin}_{(n,p)}$, mit Erfolgswahrscheinlichkeit $p \in [0, 1]$, Länge $n \in \mathbb{N}$, $\Omega := [0, n]$ und $\mathbb{p}_{\text{Bin}_{(n,p)}}(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$ für $\omega \in [0, n]$.
- (f) *Poissonverteilung*, kurz Poi_λ , mit Parameter $\lambda \in \mathbb{R}_0^+$, $\Omega = \mathbb{N}_0$ und $\mathbb{p}_{\text{Poi}_\lambda}(\omega) = \exp(-\lambda) \frac{\lambda^\omega}{\omega!}$ für $\omega \in \mathbb{N}_0$.
- (g) *Gleich-/Uniformverteilung*, kurz U_G , auf $G \in \mathcal{B}$ mit Lebesgue-Maß $\lambda(G) \in \mathbb{R}_0^+$ und $\mathbb{f}_{\text{U}_G}(x) = \frac{1}{\lambda(G)} \mathbb{1}_G(x)$ für $x \in \mathbb{R}$.
- (h) *Exponentialverteilung*, kurz Exp_λ , mit $\lambda \in \mathbb{R}_0^+$ und $\mathbb{f}_{\text{Exp}_\lambda}(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x)$ für $x \in \mathbb{R}$.
- (i) *Normalverteilung*, kurz $\text{N}_{(\mu, \sigma^2)}$, mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}_0^+$ sowie $\mathbb{f}_{\text{N}_{(\mu, \sigma^2)}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$ für $x \in \mathbb{R}$. Weiterhin vereinbaren wir $\text{N}_{(\mu, 0)} := \delta_\mu$. \square

A01.07 **Schreibweise.** Sei $((\mathcal{S}_i, \mathcal{S}_i))_{i \in \mathcal{I}}$ eine Familie messbarer Räume mit beliebiger, nicht-leerer Indexmenge \mathcal{I} . Die Menge $\mathcal{S}_\mathcal{I} := \times_{i \in \mathcal{I}} \mathcal{S}_i$ aller Abbildungen $(s_i)_{i \in \mathcal{I}} : \mathcal{I} \rightarrow \cup_{i \in \mathcal{I}} \mathcal{S}_i$ sodass $s_i \in \mathcal{S}_i$ für alle $i \in \mathcal{I}$ ist, heißt *Produkttraum* oder kartesisches Produkt. Sind alle \mathcal{S}_i gleich, etwa $\mathcal{S}_i = \mathcal{S}$, dann schreiben wir $\mathcal{S}^\mathcal{I} := \mathcal{S}_\mathcal{I}$, im Fall $n := |\mathcal{I}| < \infty$, auch nur kurz $\mathcal{S}^n := \mathcal{S}^\mathcal{I}$. Für jedes $j \in \mathcal{I}$ bezeichne $\Pi_{\mathcal{S}_j} : \mathcal{S}_\mathcal{I} \rightarrow \mathcal{S}_j$ mit $(s_i)_{i \in \mathcal{I}} \mapsto s_j$ die Koordinatenabbildung. \square

A01.08 **Definition.**

- (a) Für eine Familie $(\mathcal{A}_i)_{i \in \mathcal{I}}$ von Teil- σ -Algebren von \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} bezeichnet $\bigwedge_{i \in \mathcal{I}} \mathcal{A}_i := \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ und $\bigvee_{i \in \mathcal{I}} \mathcal{A}_i := \sigma(\bigcup_{i \in \mathcal{I}} \mathcal{A}_i)$ die *größte σ -Algebra*, die in allen $\mathcal{A}_i, i \in \mathcal{I}$, enthalten ist, bzw. die *kleinste σ -Algebra*, die alle $\mathcal{A}_i, i \in \mathcal{I}$, enthält.
- (b) Die Produkt- σ -Algebra $\mathcal{S}_{\mathcal{I}} := \bigotimes_{i \in \mathcal{I}} \mathcal{S}_i$ auf dem Produktraum $\mathcal{S}_{\mathcal{I}}$ ist die kleinste σ -Algebra, sodass für jedes $i \in \mathcal{I}$ die Koordinatenabbildung $\Pi_{\mathcal{S}_i} : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_i$ $\mathcal{S}_{\mathcal{I}}$ - \mathcal{S}_i -messbar ist, d.h. $\mathcal{S}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathcal{S}_i := \bigvee_{i \in \mathcal{I}} \sigma(\Pi_{\mathcal{S}_i}) = \bigvee_{j \in \mathcal{I}} \Pi_{\mathcal{S}_j}^{-1}(\mathcal{S}_j)$. Sind alle $(\mathcal{S}_i, \mathcal{S}_i)$ gleich, etwa $(\mathcal{S}_i, \mathcal{S}_i) = (\mathcal{S}, \mathcal{S})$, dann schreiben wir $\mathcal{S}^{\mathcal{I}} := \mathcal{S}_{\mathcal{I}}$, im Fall $n := |\mathcal{I}| < \infty$, auch nur $\mathcal{S}^n := \mathcal{S}^{\mathcal{I}}$.
- (c) Ist für jedes $i \in \mathcal{I}$ weiterhin \mathbb{P}_i ein Wahrscheinlichkeitsmaß auf $(\mathcal{S}_i, \mathcal{S}_i)$, dann heißt ein Wahrscheinlichkeitsmaß \mathbb{P} auf $\bigotimes_{i \in \mathcal{I}} \mathcal{S}_i$ *Produktmaß*, wenn für alle endlichen $\mathcal{J} \subseteq \mathcal{I}$ und $S_j \in \mathcal{S}_j, j \in \mathcal{J}$ gilt $\mathbb{P}(\bigcap_{j \in \mathcal{J}} \Pi_{\mathcal{S}_j}^{-1}(S_j)) = \prod_{j \in \mathcal{J}} \mathbb{P}_j(S_j)$. In dem Fall schreiben wir $\mathbb{P} = \bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$. Sind alle \mathbb{P}_i gleich, etwa $\mathbb{P}_i = \mathbb{P}$, dann schreiben wir kurz $\mathbb{P}^{\mathcal{I}} := \mathbb{P}$ und im Fall $n := |\mathcal{I}| < \infty$ auch $\mathbb{P}^n := \mathbb{P}$. □

A01.09 **Eigenschaft.**

- (i) Sei \mathcal{I} abzählbar, für jedes $i \in \mathcal{I}$ sei \mathcal{S}_i separabler und vollständiger metrischer Raum (polnisch) mit Borel- σ -Algebra \mathcal{B}_i und sei $\mathcal{B}_{\mathcal{S}_{\mathcal{I}}}$ die Borel- σ -Algebra bzgl. der Produkttopologie auf $\mathcal{S}_{\mathcal{I}} = \prod_{i \in \mathcal{I}} \mathcal{S}_i$. Dann gilt $\mathcal{B}_{\mathcal{S}_{\mathcal{I}}} = \mathcal{B}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathcal{B}_i$, also insbesondere $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}^n$ (vgl. Klenke [2012] Satz 14.8).
- (ii) Seien $\mathbb{P}_1, \dots, \mathbb{P}_n$ diskrete Wahrscheinlichkeitsmaße auf Ω mit Zähldichten p_1, \dots, p_n . Die *Produktzähldichte* $\prod_{i=1}^n p_i(\omega_i)$ für $(\omega_i)_{i \in [n]} \in \Omega^n$ ist die Zähldichte des Produktmaßes $\mathbb{P}_{[n]} = \bigotimes_{i \in [n]} \mathbb{P}_i$ auf $(\Omega^n, 2^{\Omega^n})$.
- (iii) Seien $\mathbb{P}_1, \dots, \mathbb{P}_n$ stetige Wahrscheinlichkeitsmaße auf \mathbb{R} mit Wahrscheinlichkeitsdichten f_1, \dots, f_n . Die *Produkt-dichte* $\prod_{i=1}^n f_i(x_i)$ für $(x_i)_{i \in [n]} \in \mathbb{R}^n$ ist Dichte des Produktmaßes $\mathbb{P}_{[n]} = \bigotimes_{i \in [n]} \mathbb{P}_i$ auf $(\mathbb{R}^n, \mathcal{B}^n)$. □

A02 Zufallsvariablen

Im Folgenden seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\mathcal{S}, \mathcal{S})$ ein messbarer Raum und $((\mathcal{S}_i, \mathcal{S}_i))_{i \in \mathcal{I}}$ eine Familie messbarer Räume mit beliebiger nicht-leerer Indexmenge \mathcal{I} .

A02.01 **Definition.**

- (a) Eine Funktion $X : \Omega \rightarrow \mathcal{S}$ heißt *\mathcal{A} - \mathcal{S} -messbar* (kurz *messbar*), falls $\sigma(X) := X^{-1}(\mathcal{S}) := \{X^{-1}(S) \mid S \in \mathcal{S}\} \subseteq \mathcal{A}$ gilt. Jede solche messbare Funktion wird ($(\mathcal{S}, \mathcal{S})$ -wertige) *Zufallsvariable* genannt. $\sigma(X)$ wird die *von X erzeugte σ -Algebra* genannt und ist die kleinste Teil- σ -Algebra aus 2^{Ω} , so dass f messbar ist.
- (b) Das Bildmaß $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$ von \mathbb{P} unter X auf $(\mathcal{S}, \mathcal{S})$ wird die Verteilung von X , kurz $X \sim \mathbb{P}^X$, genannt. Mit der Verteilungsfunktion F^X (Dichte f^X , Zähldichte p^X) von X werden wir stets die zu \mathbb{P}^X gehörigen Größen bezeichnen. X heißt *diskret-verteilte Zufallsvariable*, wenn \mathbb{P}^X ein diskretes Wahrscheinlichkeitsmaß auf $(\mathcal{S}, \mathcal{S})$ ist. Eine reelle Zufallsvariable (Zufallsvektor) X mit stetigem Bildmaß \mathbb{P}^X wird *stetig-verteilt* genannt.
- (c) $(\mathcal{S}, \mathcal{S})$ -wertige Zufallsvariablen $X_i, i \in \mathcal{I}$ heißen *identisch verteilt* (kurz *i.v.*), wenn für alle $i \in \mathcal{I}$ gilt $\mathbb{P}^{X_i} = \mathbb{P}$ für ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathcal{S})$.
- (d) Das Bildmaß $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$ unter $X := (X_i)_{i \in \mathcal{I}}$ auf $(\mathcal{S}_{\mathcal{I}}, \mathcal{S}_{\mathcal{I}})$ heißt *gemeinsame Verteilung* der Familie $(X_i)_{i \in \mathcal{I}}$. Für jedes $i \in \mathcal{I}$ wird das Bildmaß $\mathbb{P}^{X_i} := \mathbb{P} \circ X_i^{-1} = \mathbb{P} \circ (\Pi_i \circ X)^{-1} = \mathbb{P}^X \circ \Pi_i^{-1}$ *Randverteilung (marginale Verteilung)* von X_i bzgl. \mathbb{P}^{X_i} genannt. □

A02.02 **Eigenschaft.** Sei $X : \Omega \rightarrow \mathcal{S}$ eine Abbildung.

- (i) Für jedes Teilmengensystem \mathcal{E} aus $2^{\mathcal{S}}$ gilt $\sigma(X^{-1}(\mathcal{E})) = X^{-1}(\sigma(\mathcal{E}))$.
- (ii) Falls $X^{-1}(\mathcal{E}) \subseteq \mathcal{A}$ für einen Erzeuger \mathcal{E} von \mathcal{S} , also $\mathcal{S} = \sigma(\mathcal{E})$, gilt, so ist X eine \mathcal{A} - \mathcal{S} -messbare Funktion, also eine $(\mathcal{S}, \mathcal{S})$ -wertige Zufallsvariable.
- (iii) Für Abbildungen $X_i : \Omega \rightarrow \mathcal{S}_i$, $i \in \mathcal{I}$ ist $\sigma((X_i)_{i \in \mathcal{I}}) := \bigvee_{i \in \mathcal{I}} \sigma(X_i) = \sigma(\bigcup_{i \in \mathcal{I}} X_i^{-1}(\mathcal{S}_i))$ die kleinste Teil- σ -Algebra, so dass alle X_i , $i \in \mathcal{I}$, messbar sind.
- (iv) Jede stetige Funktion $g : \mathcal{S} \rightarrow \mathcal{T}$ zwischen metrischen Räumen \mathcal{S} und \mathcal{T} ist $\mathcal{B}_{\mathcal{S}}$ - $\mathcal{B}_{\mathcal{T}}$ -messbar, kurz **Borel-messbar**. □

A02.03 **Definition.** Sei $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ eine Zufallsvariable.

- (a) Falls $(\mathcal{S}, \mathcal{S}) = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, so heißt X **numerische Zufallsvariable**, kurz $X \in \overline{\mathcal{A}}$.
Falls $(\mathcal{S}, \mathcal{S}) = (\overline{\mathbb{R}}^+, \overline{\mathcal{B}}^+)$, so heißt X **positive numerische Zufallsvariable**, kurz $X \in \overline{\mathcal{A}}^+$.
- (b) Falls $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$, so heißt X **reelle Zufallsvariable**, kurz $X \in \mathcal{A}$.
Falls $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}^+, \mathcal{B}^+)$, so heißt X **positive reelle Zufallsvariable**, kurz $X \in \mathcal{A}^+$.
- (c) Falls $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}^n)$, so heißt X **Zufallsvektor**, kurz $X \in \mathcal{A}^n$. □

A02.04 **Beispiel.**

- (a) Seien (Ω, \mathcal{A}) ein messbarer Raum und $\mathbb{1}_A$ mit $A \subseteq \Omega$ die Indikatorfunktion. Dann gilt $\mathbb{1}_A \in \mathcal{A}$, d.h. $\mathbb{1}_A$ ist genau dann eine reelle Zufallsvariable, wenn $A \in \mathcal{A}$ gilt.
- (b) Für jedes $n \in \mathbb{N}$ ist die identische Abbildung $\text{id}_{\mathbb{R}^n} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit $x \mapsto \text{id}_{\mathbb{R}^n}(x) := x$ ist ein Zufallsvektor auf $(\mathbb{R}^n, \mathcal{B}^n)$, also $\text{id}_{\mathbb{R}^n} \in \mathcal{B}^n$.
- (c) Eine reelle Zufallsvariable $X \in \mathcal{A}$ heißt **einfach** oder **elementar**, falls sie nur endlich viele reelle Werte annimmt, d.h. für $X(\Omega) = \{x_i, i \in \llbracket n \rrbracket\} \subseteq \mathbb{R}$ mit $n \in \mathbb{N}$ besitzt X eine Darstellung der Form

$$X = \sum_{i \in \llbracket n \rrbracket} x_i \mathbb{1}_{A_i} \quad \text{mit } A_i := X^{-1}(\{x_i\}) = \{X = x_i\} \in \mathcal{A}.$$

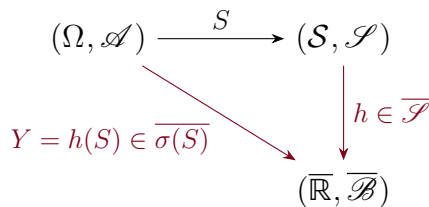
- (d) Sei $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P})$ ein stetiger Wahrscheinlichkeitsraum mit einer Wahrscheinlichkeitsdichte $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$. Dann ist f eine positive reelle Zufallsvariable. □

A02.05 **Schreibweise.**

- (i) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge numerischer (bzw. reeller) Zufallsvariablen auf (Ω, \mathcal{A}) . Die Folge $(X_n)_{n \in \mathbb{N}}$ heißt **(punktweise) monoton wachsend** (kurz **isoton**) bzw. **fallend** (kurz **antiton**), wenn $X_n(\omega) \leq X_{n+1}(\omega)$ bzw. $X_{n+1}(\omega) \leq X_n(\omega)$ für alle $\omega \in \Omega$ und für alle $n \in \mathbb{N}$ gilt. Für jedes $\omega \in \Omega$ definiere $[\liminf_{n \rightarrow \infty} X_n](\omega) := \sup_{n \geq 1} \inf_{m \geq n} X_m(\omega)$ und $[\limsup_{n \rightarrow \infty} X_n](\omega) := \inf_{n \geq 1} \sup_{m \geq n} X_m(\omega)$. Dann heißen $X_\star := \liminf_{n \rightarrow \infty} X_n$ und $X^\star := \limsup_{n \rightarrow \infty} X_n$ **Limes inferior** bzw. **Limes superior** der Folge $(X_n)_{n \in \mathbb{N}}$. Die Folge $(X_n)_{n \in \mathbb{N}}$ heißt **konvergent**, wenn $X_\star = X^\star =: X$ gilt, d.h. der punktweise Grenzwert existiert überall. In diesem Fall schreiben wir $\lim_{n \rightarrow \infty} X_n = X$.
- (ii) Jede **monoton wachsende** (bzw. **fallende**) Folge $(X_n)_{n \in \mathbb{N}}$ von numerischen oder reellen Zufallsvariablen ist konvergent mit $X := \lim_{n \rightarrow \infty} X_n = \sup_{n \in \mathbb{N}} X_n$ (bzw. $X := \lim_{n \rightarrow \infty} X_n = \inf_{n \in \mathbb{N}} X_n$). In diesem Fall schreiben wir kurz $X_n \uparrow X$ (bzw. $X_n \downarrow X$).
- (iii) Sei $(A_n)_{n \in \mathbb{N}}$ Folge von Teilmengen aus Ω . Für $A_\star := \liminf_{n \rightarrow \infty} A_n := \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k$ und $A^\star := \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$ gilt dann $\mathbb{1}_{A_\star} = (\mathbb{1}_{A_n})_\star = \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n}$ und $\mathbb{1}_{A^\star} = (\mathbb{1}_{A_n})^\star = \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}$. □

A02.06 **Eigenschaft.**

- (i) Für Zufallsvariablen $X, Y \in \overline{\mathcal{A}}$ und $a \in \mathbb{R}$ gelten: $aX \in \overline{\mathcal{A}}$ (mit der Konvention $0 \times \infty = 0$); $X \vee Y := \max(X, Y), X \wedge Y := \min(X, Y) \in \overline{\mathcal{A}}$ und insbesondere $X^+ := X \vee 0, X^- := (-X)^+ \in \overline{\mathcal{A}^+}$ und $|X| \in \overline{\mathcal{A}^+}$; $\{X < Y\}, \{X \leq Y\}, \{X = Y\} \in \mathcal{A}$. Für $x \in \overline{\mathbb{R}^+}$ sei $\lfloor x \rfloor := \sup\{z \in \mathbb{Z} : z \leq x\}$. Dann gilt $\lfloor X \rfloor \in \overline{\mathcal{A}^+}$.
- (ii) Es seien $X_1, \dots, X_n \in \mathcal{A}$ und $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Borel-messbar. Dann ist $h(X_1, \dots, X_n) \in \overline{\mathcal{A}^m}$. Insbesondere gilt also $(X_1, \dots, X_n) \in \overline{\mathcal{A}^n}$ und $X_1 + X_2, X_1 - X_2, X_1 X_2 \in \mathcal{A}$ sowie, falls überall wohldefiniert, $X_1/X_2 \in \mathcal{A}$.
- (iii) Sei $(X_n)_{n \in \mathbb{N}}$ aus $\overline{\mathcal{A}}$. Dann gilt $\sup_{n \in \mathbb{N}} X_n \in \overline{\mathcal{A}}, \inf_{n \in \mathbb{N}} X_n \in \overline{\mathcal{A}}, X_* = \liminf_{n \rightarrow \infty} X_n \in \overline{\mathcal{A}}$ und $X^* = \limsup_{n \rightarrow \infty} X_n \in \overline{\mathcal{A}}$. Falls $X := \lim_{n \rightarrow \infty} X_n$ existiert, dann ist $X \in \overline{\mathcal{A}}$.
- (iv) Sei $S : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ messbar und $Y : \Omega \rightarrow \overline{\mathbb{R}}$. Dann sind äquivalent: (a) Y ist messbar bzgl. $\sigma(S)$, kurz $Y \in \overline{\sigma(S)}$; (b) Es existiert $h : (\mathcal{S}, \mathcal{S}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ messbar, kurz $h \in \overline{\mathcal{S}}$, mit $Y = h(X)$. Falls Y reell, beschränkt oder positiv ist, so erbt h diese Eigenschaft.



- (v) Für jedes $X \in \overline{\mathcal{A}^+}$ ist $(X_n)_{n \in \mathbb{N}}$ mit $X_n := (2^{-n} \lfloor 2^n X \rfloor) \wedge n$ für $n \in \mathbb{N}$ eine Folge einfacher Zufallsvariablen aus \mathcal{A}^+ , derart dass gilt (a) $X_n \uparrow X$; (b) $X_n \leq X \wedge n$; (c) Für jedes $c \in \mathbb{R}^+$ gilt $\lim_{n \rightarrow \infty} X_n = X$ gleichmäßig auf $\{X \leq c\}$. □

A03 Unabhängigkeit

A03.01 **Definition.**

- (a) Zwei Ereignisse $A, B \in \mathcal{A}$ heißen (stochastisch) unabhängig (unter \mathbb{P}), kurz $A \perp\!\!\!\perp B$, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ gilt.
- (b) Eine Familie $(A_i)_{i \in \mathcal{I}}$ von Ereignissen aus \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} heißt (stochastisch) unabhängig (unter \mathbb{P}), kurz $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$, wenn für jede endliche Teilmenge $\mathcal{J} \subseteq \mathcal{I}$ gilt $\mathbb{P}(\bigcap_{j \in \mathcal{J}} A_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(A_j)$.
- (c) Eine Familie $(\mathcal{E}_i)_{i \in \mathcal{I}}$ von Teilmengensystemen aus \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} , d.h. $\mathcal{E}_i \subseteq \mathcal{A}$ für alle $i \in \mathcal{I}$, heißt (stochastisch) unabhängig (unter \mathbb{P}), kurz $\perp\!\!\!\perp_{i \in \mathcal{I}} \mathcal{E}_i$, wenn $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$ für alle $A_i \in \mathcal{E}_i$ und $i \in \mathcal{I}$ gilt.
- (d) Sei $(\mathcal{A}_n)_{n \in \mathbb{N}}$ eine Folge von Teil- σ -Algebren aus \mathcal{A} , dann heißt $\mathcal{A}_\infty := \bigwedge_{n \geq 1} \bigvee_{m \geq n} \mathcal{A}_m$ die asymptotische (terminale) σ -Algebra. Ein Ereignis $A \in \mathcal{A}_\infty$ wird asymptotisch (terminal) bzgl. $(\mathcal{A}_n)_{n \in \mathbb{N}}$ genannt.
- (e) Die Familie $(X_i)_{i \in \mathcal{I}}$ von Zufallsvariablen heißt unabhängig, kurz $\perp\!\!\!\perp_{i \in \mathcal{I}} X_i$, wenn die Familie $(\sigma(X_i))_{i \in \mathcal{I}}$ von Teil- σ -Algebren von \mathcal{A} unabhängig ist
- (f) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von $(\mathcal{S}_n, \mathcal{S}_n)$ -wertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann heißt $\mathcal{A}_X := \bigwedge_{n \geq 1} \bigvee_{m \geq n} \sigma(X_m)$ die asymptotische σ -Algebra. Ein Ereignis $A \in \mathcal{A}_X$ wird asymptotisch bzgl. $(X_n)_{n \in \mathbb{N}}$ genannt, d.h. A hängt für alle $n \geq 1$ nur von $(X_m)_{m \geq n}$ ab. □

A03.02 **Beispiel.**

- (a) Ist $(A_i)_{i \in \mathcal{I}}$ eine Familie von Ereignissen, so gilt $\sigma(\{A_i\}) = \sigma(X_i)$ für die Bernoulli-Zufallsvariable $X_i := \mathbb{1}_{A_i}$, $i \in \mathcal{I}$. Weiterhin gilt $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$ genau dann, wenn $\perp\!\!\!\perp_{i \in \mathcal{I}} X_i$.
- (b) Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Ereignissen aus \mathcal{A} , dann sind $A_* = \liminf_{n \rightarrow \infty} A_n$ und $A^* = \limsup_{n \rightarrow \infty} A_n$ asymptotische Ereignisse bzgl. $(\sigma(\{A_n\}))_{n \in \mathbb{N}}$.
- (c) Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von **unabhängigen** Ereignissen aus \mathcal{A} . Da $A_* = \liminf_{n \rightarrow \infty} A_n$ und $A^* = \limsup_{n \rightarrow \infty} A_n$ asymptotische Ereignisse bzgl. der Folge unabhängiger σ -Algebren $(\sigma(A_n))_{n \in \mathbb{N}}$ sind, folgt aus §A03.03 (iii) $\mathbb{P}(A_*) \in \{0, 1\}$ und $\mathbb{P}(A^*) \in \{0, 1\}$.
- (d) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge numerischer Zufallsvariablen, so sind die Abbildungen $X_* := \liminf_{n \rightarrow \infty} X_n$ und $X^* := \limsup_{n \rightarrow \infty} X_n$ messbar bzgl. \mathcal{A}_X .
- (e) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge **unabhängiger** numerischer Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so sind $X_* := \liminf_{n \rightarrow \infty} X_n$ und $X^* := \limsup_{n \rightarrow \infty} X_n$ fast sicher konstant, das heißt, es gibt $x_*, x^* \in \overline{\mathbb{R}}$ mit $\mathbb{P}(X_* = x_*) = 1$ und $\mathbb{P}(X^* = x^*) = 1$. \square

A03.03 Eigenschaft.

- (i) (**Satz von Borel-Cantelli**) Für $(A_n)_{n \in \mathbb{N}}$ aus \mathcal{A} gilt: (a) $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$, falls $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$; (b) Ist zusätzlich $\perp\!\!\!\perp_{n \in \mathbb{N}} A_n$, so auch $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$, falls $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty$.
- (ii) Für jede unabhängige Familie $(\mathcal{A}_i)_{i \in \mathcal{I}}$ von Teil- σ -Algebren aus \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} , also $\perp\!\!\!\perp_{i \in \mathcal{I}} \mathcal{A}_i$, und jede Partition $\{\mathcal{J}_k : k \in \mathcal{K}\}$ von \mathcal{I} ist die Familie $(\bigvee_{i \in \mathcal{I}_k} \mathcal{A}_i)_{k \in \mathcal{K}}$ von Teil- σ -Algebren aus \mathcal{A} unabhängig, also $\perp\!\!\!\perp_{k \in \mathcal{K}} \bigvee_{i \in \mathcal{I}_k} \mathcal{A}_i$.
- (iii) (**0-1-Gesetz von Kolmogorov**) Sei $(\mathcal{A}_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Teil- σ -Algebren aus \mathcal{A} . Dann ist die Wahrscheinlichkeit für jedes bzgl. $(\mathcal{A}_n)_{n \in \mathbb{N}}$ asymptotischen Ereignisses entweder 0 oder 1, also $\mathcal{A}_\infty \subseteq \overline{\mathcal{T}} := \{A \in \mathcal{A} \mid \mathbb{P}(A) \in \{0, 1\}\}$. $\overline{\mathcal{T}}$ wird **\mathbb{P} -triviale σ -Algebra** genannt.
- (iv) Seien $(\mathcal{T}_i, \mathcal{I}_i)_{i \in \mathcal{I}}$ eine Familie messbarer Räume und für jedes $i \in \mathcal{I}$, $h_i : \mathcal{S}_i \rightarrow \mathcal{T}_i$ eine \mathcal{I}_i - \mathcal{I}_i -messbare Funktion. Ist $\perp\!\!\!\perp_{i \in \mathcal{I}} X_i$, so gilt auch $\perp\!\!\!\perp_{i \in \mathcal{I}} h_i(X_i)$. Für jede Partition $\{\mathcal{J}_k : k \in \mathcal{K}\}$ von \mathcal{I} ist $((h_i(X_i))_{i \in \mathcal{J}_k})_{k \in \mathcal{K}}$ eine Familie von unabhängigen Zufallsvariablen, also $\perp\!\!\!\perp_{k \in \mathcal{K}} (h_i(X_i))_{i \in \mathcal{J}_k}$.
- (v) Für jede Familie $(\mathcal{S}_i, \mathcal{I}_i, \mathbb{P}_i)_{i \in \mathcal{I}}$ von Wahrscheinlichkeitsräumen mit beliebiger nicht-leerer Indexmenge \mathcal{I} existiert eine Familie $(X_i)_{i \in \mathcal{I}}$ unabhängiger $(\mathcal{S}_i, \mathcal{I}_i)$ -wertiger Zufallsvariablen definiert auf einem gemeinsamen Wahrscheinlichkeitsraum mit Randverteilung \mathbb{P}_i und dem Produktmaß $\bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$ als gemeinsamer Verteilung auf dem Produktraum $(\mathcal{S}_{\mathcal{I}}, \mathcal{I}_{\mathcal{I}})$.
- (vi) (**0-1-Gesetz von Kolmogorov**) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann ist die Wahrscheinlichkeit für jedes bzgl. $(X_n)_{n \in \mathbb{N}}$ asymptotischen Ereignisses entweder 0 oder 1, also $\mathcal{A}_X \subseteq \overline{\mathcal{T}}$. \square

A04 Erwartung

A04.01 **Definition.** Für jedes Wahrscheinlichkeitsmaß \mathbb{P} auf einem messbarem Raum (Ω, \mathcal{A}) heißt das eindeutig bestimmte Funktional $\mathbb{E} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$, das die folgenden Bedingungen erfüllt:

- (E1) Für alle $X, Y \in \overline{\mathcal{A}}^+$ und $a, b \in \overline{\mathbb{R}}^+$ gilt $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$; (linear)
- (E2) Für alle $(X_n)_{n \in \mathbb{N}} \uparrow X$ in $\overline{\mathcal{A}}^+$ gilt $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$; (monoton konvergent)
- (E3) Für jedes $A \in \mathcal{A}$ gilt $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$. (normiert)

Erwartung bzgl. \mathbb{P} und für jedes $X \in \overline{\mathcal{A}}^+$ heißt $\mathbb{E}(X)$ der **Erwartungswert** von X , wobei für $\mathcal{P} := \{\mathcal{P} \subseteq \mathcal{A} \mid \mathcal{P} \text{ endliche Partition von } \Omega\}$ gilt $\mathbb{E}(X) = \sup_{\mathcal{P} \in \mathcal{P}} \left\{ \sum_{A \in \mathcal{P}} \left(\inf_{\omega \in A} X(\omega) \right) \mathbb{P}(A) \right\}$. \square

A04.02 **Eigenschaft.**

- (i) Für $X \in \overline{\mathcal{A}}^+$ ist $\mathbb{P}(X = 0) = 1$ genau dann wenn $\mathbb{E}(X) = 0$. Insbesondere für $\mathbb{E}(X) < \infty$ gilt $\mathbb{P}(X = +\infty) = 0$.
- (ii) Für $X, Y \in \overline{\mathcal{A}}^+$ gilt $\mathbb{P}(X \leq Y) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) \leq \mathbb{E}(Y\mathbb{1}_A)$ für alle $A \in \mathcal{A}$ gilt. Insbesondere, aus $\mathbb{P}(X \leq Y) = 1$ folgt $\mathbb{E}(X) \leq \mathbb{E}(Y)$. Weiterhin gilt $\mathbb{P}(X = Y) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$ für alle $A \in \mathcal{A}$ gilt.
- (iii) (**Lemma von Fatou**) Für $(X_n)_{n \in \mathbb{N}}$ aus $\overline{\mathcal{A}}^+$ gilt $\mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$.
- (iv) Eine Familie $(\mathcal{A}_i)_{i \in \mathcal{I}}$ von Teil- σ -Algebren aus \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} ist unabhängig, also $\perp_{i \in \mathcal{I}} \mathcal{A}_i$, genau dann, wenn für jede Familie $(X_i)_{i \in \mathcal{I}}$ positiver numerischer Zufallsvariablen mit $X_i \in \overline{\mathcal{A}}^+_{i, i \in \mathcal{I}}$, und jede endliche nicht-leere Teilmenge $\mathcal{J} \subseteq \mathcal{I}$ gilt $\mathbb{E}\left(\prod_{j \in \mathcal{J}} X_j\right) = \prod_{j \in \mathcal{J}} \mathbb{E}(X_j)$. \square

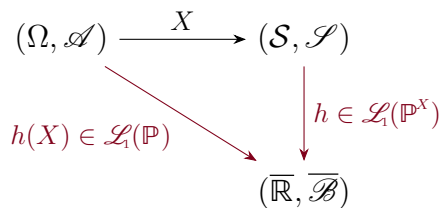
A04.03 **Definition.** Sei $X \in \overline{\mathcal{A}}$ eine numerische Zufallsvariable.

- (a) Ist höchstens einer der beiden Erwartungswerte $\mathbb{E}(X^+)$ und $\mathbb{E}(X^-)$ nicht endlich, dass heißt, $\mathbb{E}(X^+) \wedge \mathbb{E}(X^-) < \infty$, so definiert $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$ den **Erwartungswert** von X mit den üblichen Konventionen $\infty + x = \infty$ und $-\infty + x = -\infty$ für alle $x \in \mathbb{R}$. Der Erwartungswert von X ist nicht definiert, wenn $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$ gilt.
- (b) Falls $\mathbb{E}(|X|) < \infty$, also falls $\mathbb{E}(X^+) < \infty$ und $\mathbb{E}(X^-) < \infty$, gilt, dann heißt X **integrierbar**. Die Menge aller integrierbaren numerischen Zufallsvariablen bezeichnen wir mit $\mathcal{L}_1 := \mathcal{L}_1(\mathbb{P}) := \mathcal{L}(\mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}} : \mathbb{E}(|X|) < \infty\}$.
- (c) Für $p \in \mathbb{R}_0^+$ definiere $\|X\|_{\mathcal{L}_p} := (\mathbb{E}(|X|^p))^{1/p}$ und $\|X\|_{\mathcal{L}_\infty} := \inf \{x \in \mathbb{R}^+ : \mathbb{P}(|X| > x) = 0\}$. Für $p \in \overline{\mathbb{R}}_0^+$ heißt X **\mathcal{L}_p -integrierbar**, wenn $\|X\|_{\mathcal{L}_p} < \infty$. Die Menge aller \mathcal{L}_p -integrierbaren Zufallsvariablen bezeichnen wir mit $\mathcal{L}_p := \mathcal{L}_p(\mathbb{P}) := \mathcal{L}_p(\mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}} : \|X\|_{\mathcal{L}_p} < \infty\}$.
- (d) Für $X \in \mathcal{L}_p$ und $p \in \mathbb{N}$ heißt $\mathbb{E}(X^p)$ das **p -te Moment** von X ; für $X \in \mathcal{L}_p$ und $p \in \mathbb{R}_0^+$ heißt $\mathbb{E}(|X|^p)$ das **p -te absolute Moment** von X .
- (e) Für Zufallsvariablen $X, Y \in \mathcal{L}_2$ bezeichnet $\text{Cov}(X, Y) := \mathbb{E}(\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ die **Kovarianz** zwischen X und Y . Mit $\text{Var}(X) := \text{Cov}(X, X) = \mathbb{E}(\{X - \mathbb{E}(X)\}^2) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2$ bzw. $\text{std}(X) := \sqrt{\text{Var}(X)}$ bezeichnet man die **Varianz** bzw. **Standardabweichung** (standard deviation) von X .
- (f) Für $X, Y \in \mathcal{L}_2$ mit $\text{std}(X), \text{std}(Y) \in \mathbb{R}_0^+$ heißt $\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\text{std}(X)\text{std}(Y)} \in [-1, 1]$ **Korrelation** zwischen X und Y . Falls $\text{Cov}(X, Y) = 0$ gilt, heißen X und Y **unkorreliert**. \square

A04.04 **Schreibweise.** Für $k \in \mathbb{N}$ bezeichnen wir mit $\mathcal{W}_k := \mathcal{W}_k(\mathcal{B})$ die Menge aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ mit endlichem k -ten absolutem Moment, also für alle $\mathbb{P} \in \mathcal{W}_k$ gilt $\text{id}_{\mathbb{R}} \in \mathcal{L}_k(\mathbb{P})$ (vgl. **Beispiel A02.04 (b)**). Weiterhin, bezeichnet \mathbb{E} die Erwartung bzgl. \mathbb{P} , so schreiben wir für $Y \sim \mathbb{P}$ zum Beispiel auch $\mathbb{E}_p(Y) := \mathbb{E}(\text{id}_{\mathbb{R}}) = \int_{\mathbb{R}} y \mathbb{P}(dy)$. \square

A04.05 **Eigenschaft.** Für $p \in [1, \infty]$ ist \mathcal{L}_p ein Vektorraum. Für das Funktional $\mathbb{E} : \mathcal{L}_p \rightarrow \mathbb{R}$ mit $X \mapsto \mathbb{E}(X)$ gilt:

- (i) Für alle $X, Y \in \mathcal{L}_p$ und $a, b \in \mathbb{R}$ gilt $aX + bY \in \mathcal{L}_p$ und $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, falls $X \geq 0$ so auch $\mathbb{E}(X) \geq 0$; $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$; falls $X \leq Y$ so auch $\mathbb{E}(X) \leq \mathbb{E}(Y)$;
- (ii) Seien $X \in \mathcal{L}_1$ und $Y \in \overline{\mathcal{A}}$ mit $\mathbb{P}(X = Y) = 1$, dann gilt $Y \in \mathcal{L}_1$ und $\mathbb{E}(X) = \mathbb{E}(Y)$.
- (iii) Falls $X, Y \in \mathcal{L}_1$ unabhängig sind, so gilt $XY \in \mathcal{L}_1$ und $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Somit sind unabhängige Zufallsvariablen unkorreliert. Die Umkehrung gilt nicht. Für unkorrelierte Zufallsvariablen $X, Y \in \mathcal{L}_2$ gilt $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
- (iv) Für $X \in \mathcal{L}_1$ ist $\mathbb{P}(X = 0) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) = 0$ für alle $A \in \mathcal{A}$ gilt.
- (v) (**Ungleichung von Jensen**) Seien $X \in \mathcal{L}_1$ und $\phi : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion mit $\phi(X) \in \mathcal{L}$. Dann gilt $\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$.
- (vi) (**Monotone Konvergenz**) Sei $(X_n)_{n \in \mathbb{N}}$ aus \mathcal{L}_1 mit $X_n \uparrow X$ (bzw. $X_n \downarrow X$) für ein $X \in \overline{\mathcal{A}}$ und $\sup_{n \in \mathbb{N}} |\mathbb{E}(X_n)| < \infty$. Dann gilt $X \in \mathcal{L}_1$ und $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ (bzw. $\mathbb{E}(X_n) \downarrow \mathbb{E}(X)$).
- (vii) (**Dominierte Konvergenz**) Sei $(X_n)_{n \in \mathbb{N}}$ aus \mathcal{L}_1 mit $\lim_{n \rightarrow \infty} X_n = X$ für ein $X \in \overline{\mathcal{A}}$ und existiert $Y \in \mathcal{L}_1$ mit $\sup_{n \in \mathbb{N}} |X_n| \leq Y$, also $\sup_{n \in \mathbb{N}} |X_n| \in \mathcal{L}_1$, dann gilt $X \in \mathcal{L}_1$, $\lim_{n \rightarrow \infty} \mathbb{E}(|X - X_n|) = 0$ und $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.
- (viii) Sei X eine $(\mathcal{S}, \mathcal{S})$ -wertige Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, sei $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$ die Verteilung von X auf $(\mathcal{S}, \mathcal{S})$ und sei $h \in \overline{\mathcal{F}}$ eine numerische Zufallsvariable auf $(\mathcal{S}, \mathcal{S})$. Für $h \geq 0$, also $h \in \overline{\mathcal{F}}^+$, gilt $\mathbb{E}_{\mathbb{P}}(h(X)) = \mathbb{E}_{\mathbb{P}^X}(h)$. Weiterhin gilt $h(X) \in \mathcal{L}_1(\Omega, \mathcal{A}, \mathbb{P})$ genau dann, wenn $h \in \mathcal{L}_1(\mathcal{S}, \mathcal{S}, \mathbb{P}^X)$.



- (ix) Für $p \in \mathbb{R}_0^+$ gilt $X \in \mathcal{L}_p$ genau dann, wenn $|X|^p \in \mathcal{L}_1$ gilt. Für $p = \infty$ gilt $\mathbb{P}(|X| > \|X\|_{\mathcal{L}_\infty}) = 0$.
- (x) Für $p, q \in \mathbb{R}_0^+$ mit $p \leq q$ und $X \in \mathcal{L}_q$ gilt $\|X\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_q}$ und somit $\mathcal{L}_q \subseteq \mathcal{L}_p$.
- (xi) Sei $p \in \mathbb{R}_0^+$. Für $X \in \overline{\mathcal{A}}$ gilt $\|X\|_{\mathcal{L}_p} = 0$ genau dann, wenn $\mathbb{P}(X = 0) = 1$. Für $a \in \mathbb{R}$ gilt $\|aX\|_{\mathcal{L}_p} = |a| \|X\|_{\mathcal{L}_p}$. Für $X \in \mathcal{L}_p$ gilt $\mathbb{P}(|X| < \infty) = 1$. Für $Y \in \overline{\mathcal{A}}$ mit $\mathbb{P}(X = Y) = 1$ gilt $\|X\|_{\mathcal{L}_p} = \|Y\|_{\mathcal{L}_p}$.
- (xii) (**Hölder Ungleichung**) Seien $X, Y \in \overline{\mathcal{A}}$ und $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt: $\mathbb{E}(|XY|) \leq \|X\|_{\mathcal{L}_p} \|Y\|_{\mathcal{L}_q}$.
(Cauchy-Schwarz Ungleichung) Für $X, Y \in \mathcal{L}_2$ gilt $XY \in \mathcal{L}_1$ und $|\mathbb{E}(XY)|^2 \leq \|X\|_{\mathcal{L}_2}^2 \|Y\|_{\mathcal{L}_2}^2 = \mathbb{E}(|X|^2) \mathbb{E}(|Y|^2)$.
- (xiii) (**Minkowski Ungleichung**) Für $X, Y \in \mathcal{L}_p$ mit $p \in [1, \infty]$ gilt $X + Y \in \mathcal{L}_p$ und $\|X + Y\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_p} + \|Y\|_{\mathcal{L}_p}$. Insbesondere ist $\|\cdot\|_{\mathcal{L}_p}$ für $p \in [1, \infty]$ eine Pseudonorm auf \mathcal{L}_p , das heißt für $X, Y \in \mathcal{L}_p$ und $a \in \mathbb{R}$ gilt: (a) $\|X\|_{\mathcal{L}_p} \geq 0$ für alle X und $\|X\|_{\mathcal{L}_p} = 0$, falls $X = 0$ \mathbb{P} -f.s.; (b) $\|aX\|_{\mathcal{L}_p} = |a| \|X\|_{\mathcal{L}_p}$ und (c) $\|X + Y\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_p} + \|Y\|_{\mathcal{L}_p}$.
- (xiv) (**Markov Ungleichung**) Für $Y \in \overline{\mathcal{A}}^+$ und $p, \varepsilon \in \mathbb{R}_0^+$ gilt $\varepsilon^p \mathbb{1}_{\{Y > \varepsilon\}} \leq Y^p$, sodass $\mathbb{P}(Y > \varepsilon) \leq \varepsilon^{-p} \mathbb{E}(Y^p)$.
(Tschebischeff Ungleichung) Für $X \in \mathcal{L}_2$ und $\varepsilon \in \mathbb{R}_0^+$ gilt $\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \varepsilon^{-2} \text{Var}(X)$. □

A04.06 **Schreibweise.** Im Folgenden fassen wir Vektoren $a = (a_1 \cdots a_n)^t \in \mathbb{R}^n$ als Spaltenvektoren auf. Wir bezeichnen mit $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt auf \mathbb{R}^n , das heißt, $\langle a, b \rangle = \sum_{i=1}^n a_i b_i = b^t a$ für alle $a, b \in \mathbb{R}^n$. Für $p \in \overline{\mathbb{R}}_0^+$ sei $\|\cdot\|_p$ die übliche \mathcal{L}_p -Norm auf \mathbb{R}^n , das heißt, für alle $a \in \mathbb{R}^n$ ist $\|a\|_p = (\sum_{i=1}^n |a_i|^p)^{1/p}$ für $p \in \mathbb{R}_0^+$ sowie $\|a\|_\infty = \max_{i \in \llbracket n \rrbracket} |a_i|$. Für die vom Standardskalarprodukt $\langle \cdot, \cdot \rangle$ auf \mathbb{R}^n induzierte Norm schreiben wir kurz $\|\cdot\| := \|\cdot\|_2$, also $\|a\| = \sqrt{\langle a, a \rangle} = \sqrt{a^t a} = (\sum_{i=1}^n a_i^2)^{1/2}$ für alle $a \in \mathbb{R}^n$. Für eine symmetrische Matrix $\Sigma = (\Sigma_{ij}) \in \mathbb{R}^{(p,p)}$ schreiben wir $\Sigma \geq 0$ oder $\Sigma \in \mathbb{R}_{\geq}^{(p,p)}$, falls sie positiv semi-definit ist, sowie $\Sigma > 0$ oder $\Sigma \in \mathbb{R}_{>}^{(p,p)}$, falls sie strikt positiv definit ist. \square

A04.07 **Definition.** Sei $X = (X_i)_{i \in \llbracket n \rrbracket} \in \overline{\mathcal{A}}^n$ ein numerischer Zufallsvektor (aufgefasst als Spaltenvektor) auf $(\Omega, \mathcal{A}, \mathbb{P})$, also $X_i \in \overline{\mathcal{A}}, i \in \llbracket n \rrbracket$, sowie $\|X\|_p \in \overline{\mathcal{A}}$ für $p \in \overline{\mathbb{R}}_0^+$.

- (a) Für $p \in \overline{\mathbb{R}}_0^+$ heißt X \mathcal{L}_p -integrierbar, falls $\|X\|_p \in \mathcal{L}_p(\mathcal{A}, \mathbb{P})$ oder dazu äquivalent $X_i \in \mathcal{L}_p$ für alle $i \in \llbracket n \rrbracket$ gilt. Wir definieren $\|X\|_{\mathcal{L}_p^n} := \|\|X\|_p\|_{\mathcal{L}_p}$ sowie $\mathcal{L}_p^n := \mathcal{L}_p^n(\mathbb{P}) := \mathcal{L}_p^n(\mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}}^n : \|X\|_{\mathcal{L}_p^n} \in \mathbb{R}^+\}$.
- (b) Für $X \in \mathcal{L}_1^n$ heißt $\mathbb{E}(X) := (\mathbb{E}(X_i))_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$ *Erwartungswertvektor* von X .
- (c) Falls $X \in \mathcal{L}_2^n$, dann heißt

$$\begin{aligned} \text{Cov}(X) &:= (\text{Cov}(X_i, X_j))_{i,j \in \llbracket n \rrbracket} := \left(\mathbb{E}(\{X_i - \mathbb{E}(X_i)\}\{X_j - \mathbb{E}(X_j)\}) \right)_{i,j \in \llbracket n \rrbracket} \\ &=: \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t) = \mathbb{E}(XX^t) - \mathbb{E}(X)\mathbb{E}(X)^t \in \mathbb{R}^{(n,n)} \end{aligned}$$

Kovarianzmatrix von X . \square

A04.08 **Schreibweise.** Für $k \in \mathbb{N}$ bezeichnen wir mit $\mathcal{W}_k := \mathcal{W}_k(\mathcal{B}^n)$ die Menge aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit endlichem k -ten absolutem Moment, also für alle $\mathbb{P} \in \mathcal{W}_k(\mathcal{B}^n)$ gilt $\text{id}_{\mathbb{R}^n} \in \mathcal{L}_k(\mathcal{B}^n, \mathbb{P})$ (vgl. **Beispiel** A02.04 (b)). Weiterhin, bezeichnet \mathbb{E} die Erwartung bzgl. \mathbb{P} , so schreiben wir für $Y \sim \mathbb{P}$ auch $\mathbb{E}(Y) = \mathbb{P}(Y) := \mathbb{E}(\text{id}_{\mathbb{R}^n}) = \left(\int_{\mathbb{R}^n} y_i \mathbb{P}(dy) \right)_{i \in \llbracket n \rrbracket}$. \square

A04.09 **Eigenschaft.**

- (i) (*Ungleichung von Jensen*) Es seien $\mathcal{E} \subseteq \mathbb{R}^n$ konvex und $X = (X_i)_{i \in \llbracket n \rrbracket} \in \mathcal{L}_1^n(\mathbb{P})$ mit $\mathbb{P}(X \in \mathcal{E}) = 1$, so dass $\mathbb{E}(X) = (\mathbb{E}(X_i))_{i \in \llbracket n \rrbracket} \in \mathcal{E}$. Ist $\psi : \mathcal{E} \rightarrow \mathbb{R}$ eine konvexe Funktion dann gilt $\mathbb{E}(\psi(X)) \geq \psi(\mathbb{E}(X))$.
- (ii) Für die Kovarianzmatrix $\Sigma := \text{Cov}(X)$ eines Zufallsvektors $X \in \mathcal{L}_2^n$ gilt für alle $a, b \in \mathbb{R}^n$, $\text{Cov}(\langle a, X \rangle, \langle X, b \rangle) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}(X_i, X_j) = b^t \Sigma a = \langle \Sigma a, b \rangle$. Insbesondere gilt $\Sigma_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = \Sigma_{ji}$ für alle $i, j \in \llbracket n \rrbracket$ und $\langle \Sigma c, c \rangle = \text{Var}(\langle X, c \rangle) \geq 0$ für alle $c \in \mathbb{R}^n$, also $\Sigma \in \mathbb{R}_{\geq}^{(n,n)}$.
- (iii) Sei $X \in \mathcal{A}^m$ in \mathcal{L}_2 . Für alle $b \in \mathbb{R}^n$ und $A \in \mathbb{R}^{(n,m)}$ ist dann $Y = AX + b \in \mathcal{L}_2^n$. Bezeichnen wir weiterhin mit $\mu := \mathbb{E}(X) \in \mathbb{R}^m$ und $\Sigma := \text{Cov}(X) \in \mathbb{R}^{(m,m)}$ den Erwartungswertvektor und die Kovarianzmatrix von X , dann gilt $\mathbb{E}(Y) = A\mu + b$ und $\text{Cov}(Y) = A\Sigma A^t$. \square

A04.10 **Definition.** Für $X \in \mathcal{L}_2^k$ und $Y \in \mathcal{L}_2^n$ heißt $Z^* = A^*X + b^*$ mit $A^* \in \mathbb{R}^{(n,k)}$ und $b^* \in \mathbb{R}^n$ eine *lineare Vorhersage* von Y durch X . Z^* wird *beste lineare Vorhersage* von Y durch X genannt, wenn $\mathbb{E}\|Y - Z^*\|^2 \leq \mathbb{E}\|Y - (AX + b)\|^2$ für alle $A \in \mathbb{R}^{(n,k)}$ und $b \in \mathbb{R}^n$ gilt. \square

A04.11 **Bemerkung.** Für eine Matrix $A \in \mathbb{R}^{(n,k)}$ heißt eine Matrix $A^+ \in \mathbb{R}^{(k,n)}$ *Moore-Penrose Inverse* von A , wenn $AA^+A = A$, $A^+AA^+ = A^+$ und AA^+ sowie A^+A symmetrisch sind. Die Moore-Penrose Inverse ist eindeutig festgelegt. Ist $B \in \mathbb{R}^n$ symmetrisch, so dass $U^t B U = \text{Diag}(\lambda)$

für eine orthogonale Matrix U und eine Diagonalmatrix $\text{Diag}(\lambda)$ mit reellen Diagonaleinträgen $\lambda = (\lambda_i)_{i \in \llbracket n \rrbracket}$. Setzen wir $\lambda^+ = (\lambda_i^+)_{i \in \llbracket n \rrbracket}$ mit $\lambda_i^+ = \lambda_i^{-1} \mathbb{1}_{\mathbb{R}_0^+}(\lambda_i)$, das heißt, für $\lambda_i \in \mathbb{R}_0^+$ ist $\lambda_i^+ = \lambda_i^{-1}$ und ansonsten $\lambda_i^+ = 0$. Dann ist $B^+ = U \text{Diag}(\lambda^+) U^t$ die Moore-Penrose Inverse von B . Allgemein ist $A^+ = (A^t A)^+ A^t \in \mathbb{R}^{(k,n)}$ die Moore-Penrose Inverse von $A \in \mathbb{R}^{(n,k)}$. \square

A04.12 **Eigenschaft.** Für $X \in \mathcal{L}_2^k$ und $Y \in \mathcal{L}_2^n$ mit $\text{Cov}(Y, X) := \mathbb{E}((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))^t)$ ist $Z^* = \mathbb{E}(Y) + \text{Cov}(Y, X) \text{Cov}(X)^+(X - \mathbb{E}(X))$ die *beste lineare Vorhersage* von Y durch X . Der Fehler $\varepsilon := Y - Z^*$ und AX für beliebiges $A \in \mathbb{R}^{(n,k)}$ sind unkorreliert, also $\text{Cov}(\varepsilon, AX) = 0$. Es gilt $\text{Cov}(\varepsilon) = \text{Cov}(Y) - \text{Cov}(Y, X) \text{Cov}(X)^+ \text{Cov}(X, Y)$ und $\mathbb{E}(\varepsilon) = 0$. \square

A05 Multivariate Normalverteilung

Nicht degenerierte multivariate Normalverteilungen können direkt über ihre Dichte definiert werden. Eine Normalverteilung heißt degeneriert, falls ihre Kovarianzmatrix nicht strikt positiv definit ist (nicht vollen Rang hat). In der Vorlesung werden wir auch Zufallsvariablen mit degenerierten Normalverteilungen betrachten. Beispiele für solche Zufallsvariablen sind Projektionen von nicht degenerierten normalverteilten Zufallsvariablen auf lineare Teilräume.

A05.01 **Satz von Cramér-Wold.** Die Verteilung eines \mathbb{R}^n -wertigen Zufallsvektors X ist eindeutig festgelegt durch die Verteilungen der linearen Formen $\langle X, c \rangle$ für alle $c \in \mathbb{R}^n$. \square

A05.02 **Definition.** Ein \mathbb{R}^n -wertiger Zufallsvektor X besitzt eine *multivariate Normalverteilung* $N_{(\mu, \Sigma)}$ mit $\mu \in \mathbb{R}^n$ und positiv semi-definiter Matrix $\Sigma \in \mathbb{R}^{(n,n)}$, falls für alle $c \in \mathbb{R}^n$ die reelle Zufallsvariable $\langle X, c \rangle$ eine $N_{(\langle \mu, c \rangle, \langle \Sigma c, c \rangle)}$ -Verteilung besitzt. Das Produktmaß $N_{(0, E_n)} = \bigotimes_{i=1}^n N_{(0,1)} = N_{(0,1)}^n$ heißt insbesondere (*n-dimensionale*) *Standardnormalverteilung*, wobei E_n die n -dimensionale Einheitsmatrix ist. \square

A05.03 **Vorbemerkung.** Für eine Matrix $A \in \mathbb{R}^{(n,m)}$ mit Spaltenvektoren $a_{\bullet 1}, \dots, a_{\bullet m}$ bezeichnet $\text{Bild}(A) = \langle a_{\bullet 1}, \dots, a_{\bullet m} \rangle \subseteq \mathbb{R}^n$ die lineare Hülle der Spaltenvektoren, also das Bild der linearen Abbildung $\mathbb{R}^m \rightarrow \mathbb{R}^n$ mit $x \mapsto Ax$. Für einen linearen Unterraum $\mathcal{U} \subseteq \mathbb{R}^n$ bezeichnet $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$ die direkte orthogonale Summe, das heißt, \mathcal{U} und \mathcal{U}^\perp sind orthogonal, also für alle $u \in \mathcal{U}$ und $v \in \mathcal{U}^\perp$ gilt $\langle u, v \rangle = 0$, und jedes Element $x \in \mathbb{R}^n$ hat eine eindeutige Darstellung $x = u + v$ mit $u \in \mathcal{U}$ und $v \in \mathcal{U}^\perp$. Wir bezeichnen mit $\Pi_{\mathcal{U}}$ die Darstellungsmatrix der orthogonalen Projektion von \mathbb{R}^n in \mathcal{U} , also $\mathcal{U} \oplus \mathcal{U}^\perp \rightarrow \mathcal{U}$ mit $x = u + v \mapsto u = \Pi_{\mathcal{U}} x$. Eine Matrix $U \in \mathbb{R}^{(n,m)}$ heißt partielle Isometrie, falls $UU^t = \Pi_{\text{Bild}(U)}$ und $U^t U = \Pi_{\text{Bild}(U^t)}$. \square

A05.04 **Eigenschaft.** Seien $Z \sim N_{(0, E_m)}$ und $Y \sim N_{(0, E_k)}$, dann gelten die folgenden Aussagen:

- (i) Falls $A \in \mathbb{R}^{(n,m)}$ und $B \in \mathbb{R}^{(n,k)}$ mit $AA^t = BB^t$ gilt, dann sind die \mathbb{R}^n -wertigen Zufallsvektoren AZ und BY identisch verteilt.
- (ii) Falls $U \in \mathbb{R}^{(n,m)}$ eine partielle Isometrie ist, dann gilt $UZ \sim N_{(0, \Pi_{\text{Bild}(U)})}$.
- (iii) Falls $A \in \mathbb{R}^{(m,n)}$ und $B \in \mathbb{R}^{(m,k)}$ mit $A^t B = 0$. Dann sind $\Pi_{\text{Bild}(A)} Z \sim N_{(0, \Pi_{\text{Bild}(A)})}$ und $\Pi_{\text{Bild}(B)} Z \sim N_{(0, \Pi_{\text{Bild}(B)})}$ unabhängig.

Sei $X \sim N_{(\mu, \Sigma)}$ mit $\mu \in \mathbb{R}^n$ und $\Sigma \in \mathbb{R}_{\geq}^{(n,n)}$, dann gelten die folgenden Aussagen:

- (iv) Für alle $i \in \llbracket n \rrbracket$ gilt $X_i \sim N_{(\mu_i, \Sigma_{ii})}$.
- (v) Für alle $i, j \in \llbracket n \rrbracket$ mit $i \neq j$ sind die Koordinaten X_i und X_j von X genau dann unabhängig, wenn $\Sigma_{ij} = 0$ gilt.
- (vi) Für $A \in \mathbb{R}^{(m,n)}$ und $b \in \mathbb{R}^m$ gilt $Y = AX + b \sim N_{(A\mu + b, A\Sigma A^t)}$.

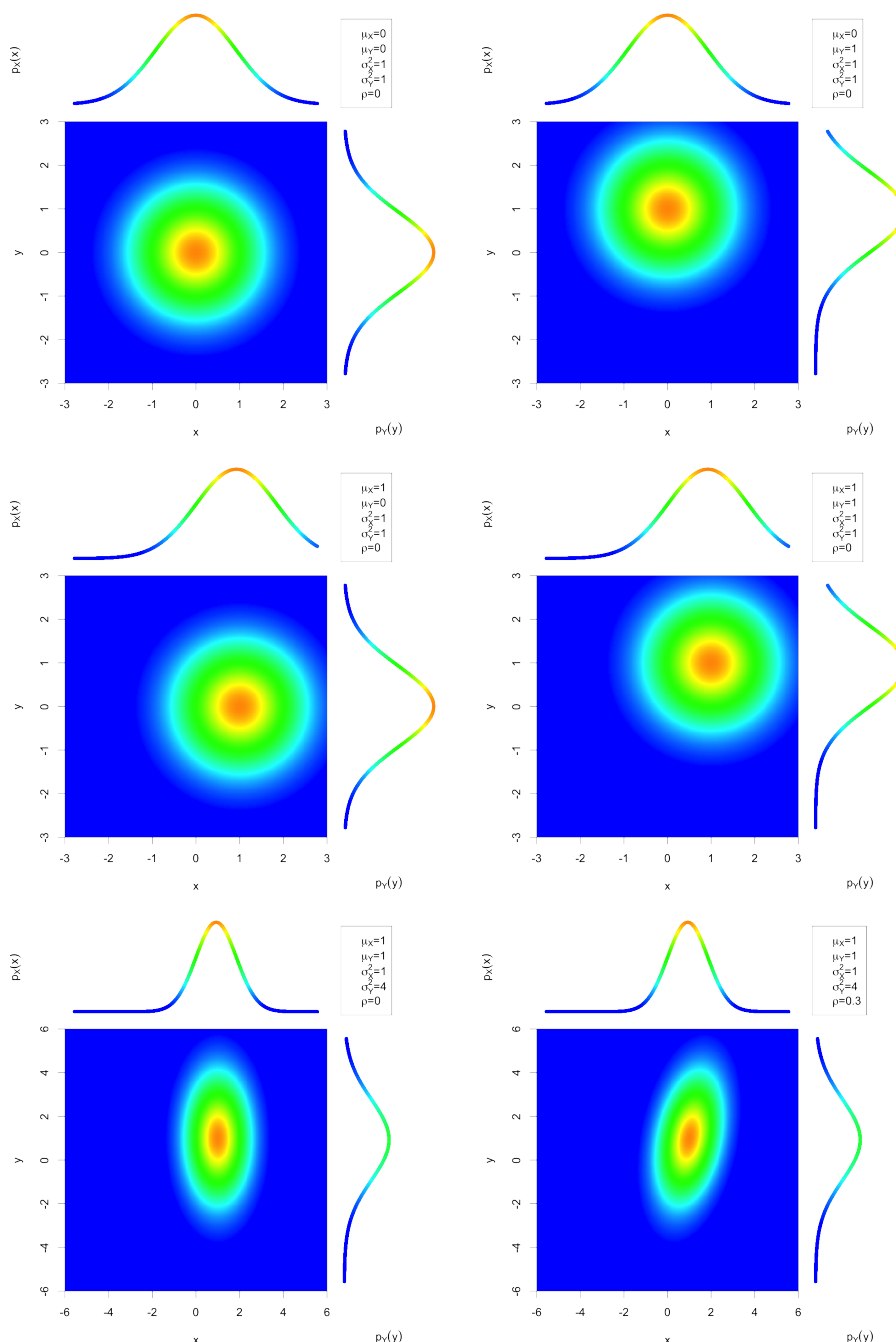
(vii) Ist Σ positiv definit, dann ist X stetig verteilt mit Lebesgue-Dichte

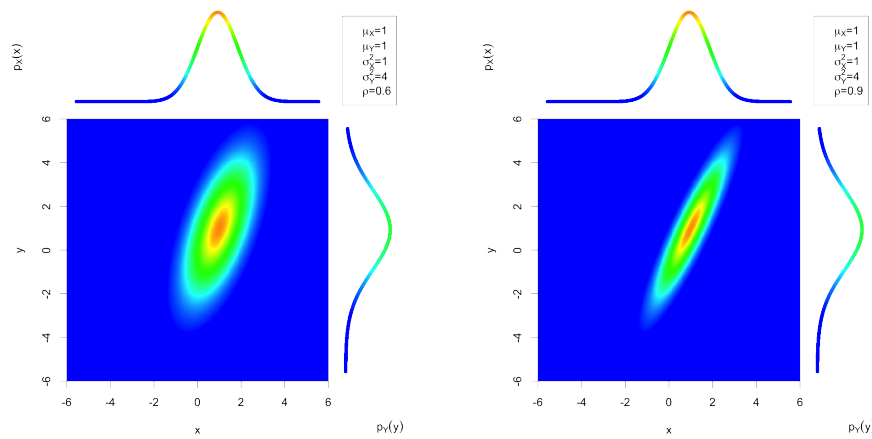
$$f(x) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} \langle \Sigma^{-1}(x - \mu), (x - \mu) \rangle \right\}, x \in \mathbb{R}^n. \quad \square$$

A05.05 **Beispiel.** Der stetig-verteilte Zufallsvektor (X, Y) wird *bivariat normalverteilt* mit Parametern $\mu_X, \mu_Y \in \mathbb{R}, \sigma_X, \sigma_Y \in \mathbb{R}_0^+$ und $\rho \in (-1, 1)$ genannt, wenn die gemeinsame Dichte durch

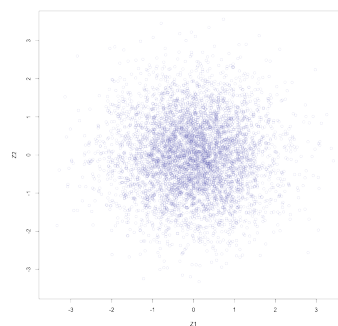
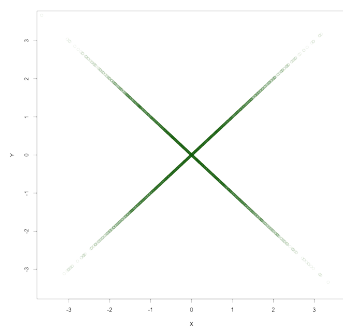
$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{(x-\mu_X)^2}{2(1-\rho^2)\sigma_X^2}\right) \exp\left(\frac{2\rho(x-\mu_X)(y-\mu_Y)}{2(1-\rho^2)\sigma_X\sigma_Y}\right) \exp\left(-\frac{(y-\mu_Y)^2}{2(1-\rho^2)\sigma_Y^2}\right)$$

gegeben ist, wobei $\mu_X = \mathbb{E}(X), \mu_Y = \mathbb{E}(Y), \sigma_X^2 = \text{Var}(X), \sigma_Y^2 = \text{Var}(Y)$ und $\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$. Die nächsten Graphiken stellen die gemeinsame und die marginalen Dichten für verschiedene Werte der Parameter dar.





X und Y sind genau dann unabhängig, wenn sie unkorreliert sind, also $\rho = 0$ gilt (vgl. §A05.04 (v)). **Achtung**, es ist natürlich möglich, dass $X \sim N_{(\mu_X, \sigma_X^2)}$ und $Y \sim N_{(\mu_Y, \sigma_Y^2)}$ unkorreliert sind, aber der Vektor (X, Y) nicht bivariat normalverteilt ist. Betrachte dazu zwei unabhängige Zufallsvariablen X und V , wobei $X \sim N_{(0,1)}$ und V ist eine Rademacher-Zufallsvariable, d.h. $V \in \{-1, 1\}$ mit $P(V = -1) = 1/2 = P(V = 1)$. Es ist nun leicht zu zeigen, dass die Zufallsvariablen $Y := VX$ und X unkorreliert sind und dass $Y \sim N_{(0,1)}$ (Nachrechnen!). Die Zufallsvariablen X und Y sind somit standardnormalverteilt und unkorreliert, aber ihre gemeinsame Verteilung ist keine Normalverteilung (warum?). Die nächsten Graphiken zeigen 5000 Realisierungen von (X, Y) (in grün) und zum Vergleich 5000 Realisierungen einer bivariaten Standardnormalverteilung.



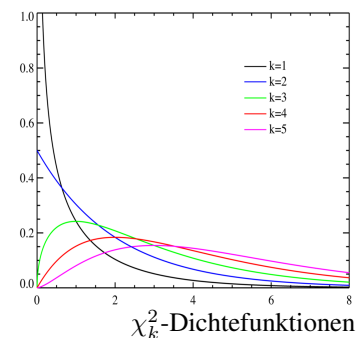
□

Im Folgenden sind $(Z_i)_{i \in [0, m+k]}$ unabhängige und identisch $N_{(0,1)}$ -verteilte Zufallsvariablen, also $(Z_i)_{i \in [0, m+k]} \sim N_{(0,1)}^{1+m+k}$.

A05.06 **χ^2 -Verteilung**. Die Verteilung der Zufallsvariable

$$Q := \sum_{i=1}^k Z_i^2$$

heißt (zentrale) χ^2 -Verteilung mit k Freiheitsgraden, kurz $Q \sim \chi_k^2$. Für $\alpha \in (0, 1)$ bezeichnen wir weiterhin den Wert $\chi_{k, \alpha}^2 \in \mathbb{R}_0^+$ als α -Quantil einer (zentralen) χ^2 -Verteilung mit k Freiheitsgraden, falls $\mathbb{P}(Q \leq \chi_{k, \alpha}^2) = \alpha$ gilt.



Für $\delta \in \mathbb{R}$ heißt die Verteilung der Zufallsvariable

$$Q := (Z_1 + \delta)^2 + \sum_{i=2}^k Z_i^2$$

nicht-zentrale χ_k^2 -Verteilung mit k Freiheitsgraden und Nichtzentralitätsparameter δ^2 , kurz

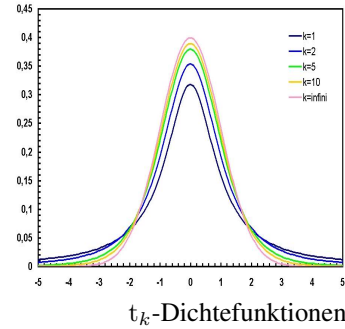
$Q \sim \chi_k^2(\delta^2)$. $\chi_{k,\alpha}^2(\delta^2) \in \mathbb{R}_0^+$ bezeichnet das α -Quantil einer nicht-zentralen χ^2 -Verteilung mit k Freiheitsgraden und Nichtzentralitätsparameter δ^2 , d.h. $\mathbb{P}(Q \leq \chi_{k,\alpha}^2(\delta^2)) = \alpha$. □

A05.07 **Eigenschaft.** Sei $Q \sim \chi_k^2$ und $W \sim \chi_k^2(\delta^2)$, dann gilt $\mathbb{E}(Q) = k$, $\text{Var}(Q) = 2k$ und $\mathbb{E}(W) = \delta^2 + k$. Für $Z \sim N_{(0,1)}^m$, $v \in \mathbb{R}^m$ und $A \in \mathbb{R}^{m \times p}$ mit $\text{rg}(A) = p$ gelten außerdem: (i) $\|\Pi_{\text{Bild}(A)}Z\|^2 \sim \chi_p^2$ und (ii) $\|Z + v\|^2 \sim \chi_m^2(\|v\|^2)$. □

A05.08 **(Student-) t-Verteilung.** Die Verteilung der Zufallsvariable

$$T := \frac{Z_0}{\sqrt{\frac{1}{k} \sum_{i=1}^k Z_i^2}}$$

heißt **(Student-) t-Verteilung** mit k **Freiheitsgraden**, kurz $T \sim t_k$, und $t_{k,\alpha} \in \mathbb{R}$ bezeichnet das α -Quantil einer (Student-) t-Verteilung mit k -Freiheitsgraden, d.h. $\mathbb{P}(T \leq t_{k,\alpha}) = \alpha$.



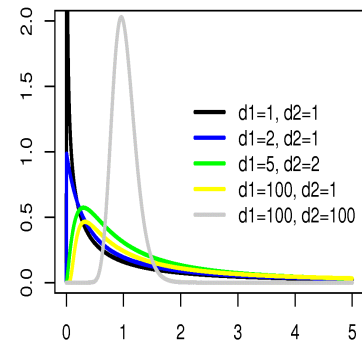
A05.09 **Eigenschaft.**

- (i) Die (Student-) t_1 -Verteilung mit einem ($k = 1$) Freiheitsgrad entspricht gerade der Cauchy-Verteilung.
- (ii) Für jedes $k \in \mathbb{N}$ besitzt die t_k -Verteilung endliche Momente nur bis zur Ordnung $p < k$ (sie ist heavy-tailed). Insbesondere, ist $T \sim t_k$ so gilt $\mathbb{E}(T) = 0$ für $k > 1$, sowie $\text{Var}(T) = k/(k - 2)$ für $k > 2$.
- (iii) Für $(X_i)_{i \in [n]} \sim N_{(\mu, \sigma^2)}^n$, $\bar{X}_n := \sum_{i=1}^n X_i$ und $\hat{S}_n^{(2)} := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sind $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \sim N_{(0,1)}$ und $\frac{(n-1)}{\sigma^2} \hat{S}_n^{(2)} \sim \chi_{n-1}^2$ unabhängig, so dass $\hat{T}_n = \frac{\sqrt{n}}{\hat{S}_n}(\bar{X}_n - \mu) \sim t_{n-1}$ mit $\hat{S}_n := \sqrt{\hat{S}_n^{(2)}}$ gilt. □

A05.10 **(Fisher-) F-Verteilung.** Die Verteilung der Zufallsvariable

$$F := \frac{\frac{1}{m} \sum_{i=1}^m Z_i^2}{\frac{1}{k} \sum_{i=m+1}^{m+k} Z_i^2}$$

heißt **zentrale (Fisher-) F-Verteilung** mit m und k **Freiheitsgraden**, kurz $F \sim F_{m,k}$. $F_{m,k,\alpha}$ bezeichnet das α -Quantil einer zentralen Fisher-F $_{m,k}$ -Verteilung mit m und k Freiheitsgraden, d.h. $\mathbb{P}(F \leq F_{m,k,\alpha}) = \alpha$.



Für $\delta \in \mathbb{R}$ heißt die Verteilung der Zufallsvariable

$$F := \frac{\frac{1}{m} \{(Z_1 + \delta)^2 + \sum_{i=2}^m Z_i^2\}}{\frac{1}{k} \sum_{i=m+1}^{m+k} Z_i^2}$$

nicht-zentrale (Fisher-) F-Verteilung mit m und k **Freiheitsgraden** und **Nichtzentralitätsparameter** δ^2 , kurz $F \sim F_{m,k}(\delta^2)$. $F_{m,k,\alpha}(\delta^2) \in \mathbb{R}^+$ bezeichnet das α -Quantil einer nicht-zentralen

$F_{n,k}(\delta^2)$ -Verteilung mit m und k Freiheitsgraden und Nichtzentralitätsparameter δ^2 , das heißt $P(F \leq F_{m,k,\alpha}(\delta^2)) = \alpha$. \square

A05.11 Eigenschaft.

- (i) Sei $F \sim F_{n,k}$ mit $k > 1$, dann ist F^{-1} eine $F_{k,m}$ -verteilte Zufallsvariable. Für $T \sim t_k$ ist T^2 eine $F_{1,k}$ -verteilte Zufallsvariable.
- (ii) Sei $F_k \sim F_{m,k}$, $k \in \mathbb{N}$, dann konvergiert die Folge von Zufallsvariablen $(mF_k)_{k \in \mathbb{N}}$ für $k \rightarrow \infty$ in Verteilung gegen ein χ_m^2 -verteilte Zufallsvariable. \square

A06 Grenzwertsätze

A06.01 **Schreibweise.** Im Folgenden bezeichnen wir mit $\mathcal{C}_b := \mathcal{C}_b(\mathbb{R}^k)$ die Menge aller beschränkten, stetigen, reellen Funktionen auf \mathbb{R}^k . Für $h \in \mathcal{C}_b$ ist somit $\|h\|_\infty := \sup_{x \in \mathbb{R}^k} |h(x)| < \infty$, so dass für jedes Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}^k, \mathcal{B}^k)$, kurz $\mathbb{P} \in \mathcal{W}(\mathcal{B}^k)$, gilt $\mathcal{C}_b \subseteq \mathcal{L}_\infty(\mathcal{B}^k, \mathbb{P})$. \square

A06.02 Definition.

(a) Eine Folge $(X_n)_{n \in \mathbb{N}}$ reeller Zufallsvariablen in \mathcal{A} konvergiert

- **\mathbb{P} -fast sicher** (\mathbb{P} -f.s.) gegen die numerische Zufallsvariable $X \in \overline{\mathcal{A}}$, kurz $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$, wenn $\limsup_{n \rightarrow \infty} |X_n - X| = 0$ \mathbb{P} -f.s., das heißt $\mathbb{P}(\limsup_{n \rightarrow \infty} |X_n - X| = 0) = 1$ gilt.
- **\mathbb{P} -fast vollständig** (\mathbb{P} -f.v.) gegen die numerische Zufallsvariable $X \in \overline{\mathcal{A}}$, kurz $X_n \xrightarrow{\mathbb{P}\text{-f.v.}} X$, wenn für alle $\varepsilon \in \mathbb{R}_0^+$ gilt $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$.
- **stochastisch** gegen die numerische Zufallsvariable $X \in \overline{\mathcal{A}}$, kurz $X_n \xrightarrow{\mathbb{P}} X$, wenn für alle $\varepsilon \in \mathbb{R}_0^+$ gilt $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

(b) Eine Folge $(X_n)_{n \in \mathbb{N}}$ in \mathcal{L}_p , $p \in \overline{\mathbb{R}}_0^+$, **konvergiert in \mathcal{L}_p** gegen $X \in \mathcal{L}_p$, kurz $X_n \xrightarrow{\mathcal{L}_p} X$, wenn gilt $\lim_{n \rightarrow \infty} \|X_n - X\|_{\mathcal{L}_p} = 0$.

(c) Eine Folge $(\mathbb{P}_n)_{n \in \mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^k, \mathcal{B}^k)$ **konvergiert schwach** gegen ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}^k, \mathcal{B}^k)$, wenn $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n}(h) = \mathbb{E}_{\mathbb{P}}(h)$ für alle $h \in \mathcal{C}_b$ gilt. Wir schreiben kurz $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ oder $\mathbb{P} = w\text{-}\lim_{n \rightarrow \infty} \mathbb{P}_n$.

(d) Eine Folge $(X_n)_{n \in \mathbb{N}}$ von \mathbb{R}^k -wertigen Zufallsvektoren **konvergiert in Verteilung** gegen einen \mathbb{R}^k -wertigen Zufallsvektor X , kurz $X_n \xrightarrow{D} X$, wenn $\lim_{n \rightarrow \infty} \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$ für alle $h \in \mathcal{C}_b$, also $\mathbb{P}^X = w\text{-}\lim_{n \rightarrow \infty} \mathbb{P}^{X_n}$ gilt.

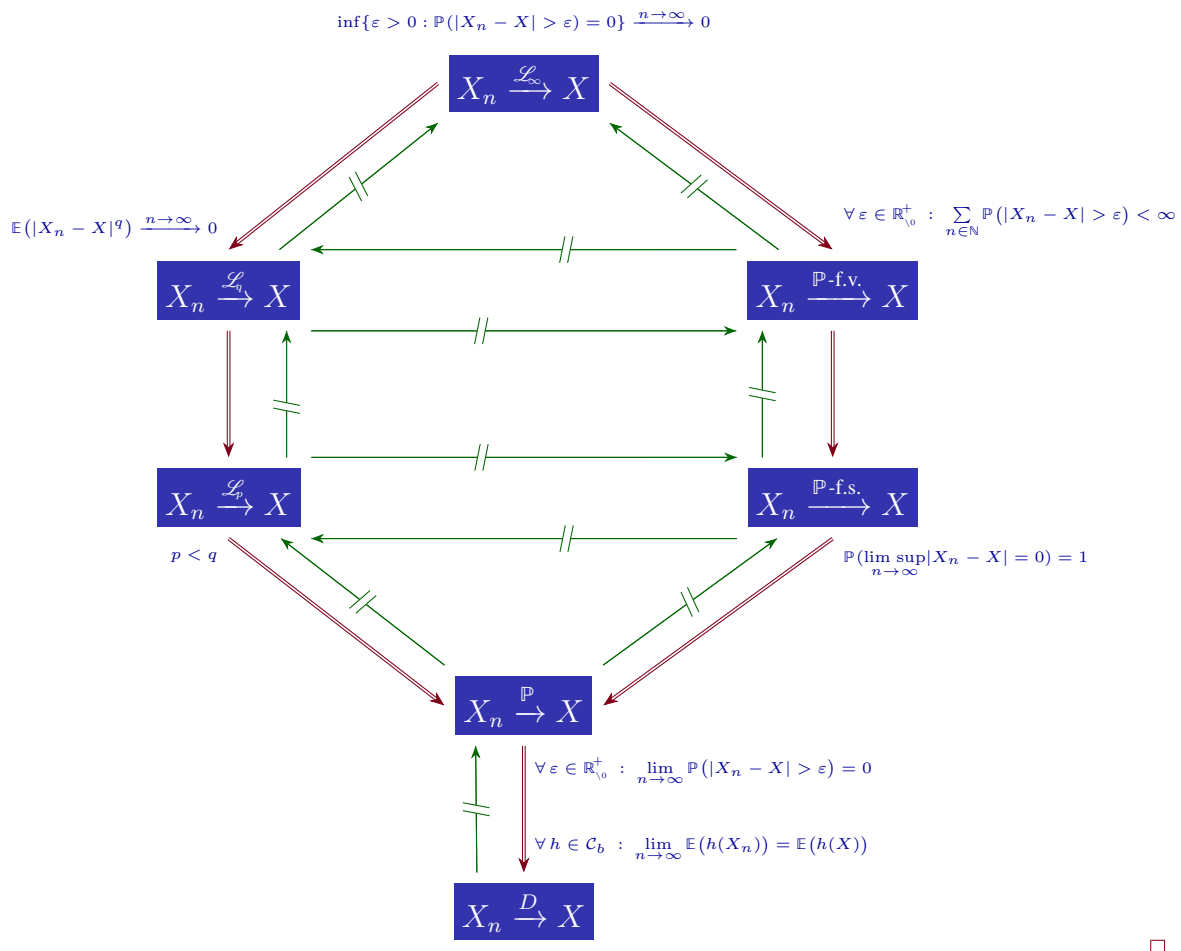
(e) Für eine Folge $(X_n)_{n \in \mathbb{N}}$ definieren wir **Konvergenz in Verteilung** mit einem Wahrscheinlichkeitsmaß \mathbb{P} als Grenzwert, kurz $X_n \xrightarrow{D} \mathbb{P}$, allgemein durch $\mathbb{P}^{X_n} \xrightarrow{w} \mathbb{P}$. \square

A06.03 **Eigenschaft.** Es seien X, X_n und Y_n , $n \in \mathbb{N}$, reelle Zufallsvariablen.

- (i) Sei $h : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion und $(X_n)_{n \in \mathbb{N}}$ konvergiere \mathbb{P} -f.s. (stochastisch bzw. in Verteilung) gegen X . Dann konvergiert $(h(X_n))_{n \in \mathbb{N}}$ auch gegen $h(X)$ \mathbb{P} -f.s. (stochastisch bzw. in Verteilung).
- (ii) Konvergiert $(Y_n)_{n \in \mathbb{N}}$ in Verteilung gegen X , also $Y_n \xrightarrow{D} X$, und konvergiert $(X_n - Y_n)_{n \in \mathbb{N}}$ stochastisch gegen Null, also $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$, dann konvergiert $(X_n)_{n \in \mathbb{N}}$ in Verteilung auch

gegen X , also $X_n \xrightarrow{D} X$. Falls $X_n \xrightarrow{D} X$ und $Y_n \xrightarrow{\mathbb{P}} a$, so gilt auch $X_n + Y_n \xrightarrow{D} X + a$ sowie $Y_n X_n \xrightarrow{D} aX$.

- (iii) $(X_n)_{n \in \mathbb{N}}$ konvergiert in Verteilung gegen eine Konstante $a \in \mathbb{R}$, also $X_n \xrightarrow{D} a$, genau dann, wenn sie stochastisch gegen a konvergiert, also $X_n \xrightarrow{\mathbb{P}} a$.
- (iv) (**Delta Methode**) Seien $x \in \mathbb{R}$ und $a_n \uparrow \infty$ mit $a_n(X_n - x) \xrightarrow{D} X$. Dann gilt $X_n \xrightarrow{\mathbb{P}} x$. Ist weiterhin $f \in \mathcal{B}$ differenzierbar in x , so gilt $a_n(f(X_n) - f(x)) - a_n(X_n - x)f'(x) \xrightarrow{\mathbb{P}} 0$ und somit auch $a_n(f(X_n) - f(x)) \xrightarrow{D} f'(x)Y$.
- (v) (**Cramér-Wold device**) Für eine Folge \mathbb{R}^d -wertiger Zufallsvektoren $(X_n)_{n \in \mathbb{N}}$ sind äquivalent:
 - (a) Es gibt einen Zufallsvektor X mit $X_n \xrightarrow{D} X$.
 - (b) Für jedes $v \in \mathbb{R}^d$ existiert ein X^v mit $\langle v, X_n \rangle \xrightarrow{D} X^v$. Falls (a) und (b) gelten, so sind X^v und $\langle v, X \rangle$ identisch verteilt.
- (vi) Gegenbeispiele zeigen, dass die Umkehrungen (in grün) der folgenden direkten Implikationen (in rot) nicht gelten.



A06.04 **Definition.** Eine Familie $(X_{n,j})_{j \in [n], n \in \mathbb{N}}$ reeller Zufallsvariablen in \mathcal{L}_2 heißt **standardisiertes Dreiecksschema**, wenn (i) $(X_{n,j})_{j \in [n]}$ sind unabhängig; sowie (ii) $\mathbb{E}(X_{n,j}) = 0, j \in [n]$, und $\sum_{j \in [n]} \text{Var}(X_{n,j}) = 1$ für jedes $n \in \mathbb{N}$ gilt.

Ein standardisiertes Dreiecksschema $(X_{n,j})_{j \in [n], n \in \mathbb{N}}$ erfüllt

- (a) die **Lindeberg-Bedingung**, wenn $\lim_{n \rightarrow \infty} \sum_{j \in [n]} \mathbb{E}(X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}| \geq \delta\}}) = 0$ für jedes $\delta \in \mathbb{R}_0^+$ gilt;
- (b) die **Lyapunov-Bedingung**, wenn $\lim_{n \rightarrow \infty} \sum_{j \in [n]} \mathbb{E}(|X_{n,j}|^{2+\delta}) = 0$ für ein $\delta \in \mathbb{R}_0^+$ gilt. □

A06.05 **Eigenschaft.** Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen.

- (i) (*Starkes Gesetz der großen Zahlen*) Seien X_n , $n \in \mathbb{N}$, identisch-verteilt. $X_1 \in \mathcal{L}_1$ gilt genau dann, wenn $\lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X_1)$ \mathbb{P} -f.s. (und dann auch in \mathcal{L}_1).
- (ii) (*Lévy's Äquivalenzsatz*) Die Folge der Partialsummen $(S_n)_{n \in \mathbb{N}}$ mit $S_n := \sum_{i \in [n]} X_i$, $n \in \mathbb{N}$, konvergiert \mathbb{P} -f.s. genau dann, wenn sie stochastisch konvergiert. Andernfalls ist $(S_n)_{n \in \mathbb{N}}$ divergent mit Wahrscheinlichkeit Eins.
- (*Dreireihensatz von Kolmogorov*) $(S_n)_{n \in \mathbb{N}}$ konvergiert \mathbb{P} -f.s. genau dann, wenn die folgenden drei Bedingungen für ein $\varepsilon \in \mathbb{R}_0^+$ gelten: (a) $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) < \infty$; (b) $\sum_{n \in \mathbb{N}} \mathbb{E}(X_n \mathbb{1}_{\{|X_n| \leq \varepsilon\}})$ konvergiert; und (c) $\sum_{n \in \mathbb{N}} \text{Var}(X_n \mathbb{1}_{\{|X_n| \leq \varepsilon\}}) < \infty$.
- Sei $(X_{n,j})_{j \in [n], n \in \mathbb{N}}$ ein standardisiertes Dreiecksschema.
- (iii) Erfüllt $(X_{n,j})_{j \in [n], n \in \mathbb{N}}$ die Lyapunov-Bedingung, so auch die Lindeberg-Bedingung.
- (iv) (*Zentraler Grenzwertsatz nach Lindeberg (1922)*) Erfüllt $(X_{n,j})_{j \in [n], n \in \mathbb{N}}$ die Lindeberg-Bedingung, so gilt für (die Zeilensumme) $S_n^* = \sum_{j \in [n]} X_{nj} \xrightarrow{D} N_{(0,1)}$. \square

B Elementare Maß- und Integrationstheorie

In diesem Kapitel werden Begriffe, Notationen und Aussagen der Vorlesung *Wahrscheinlichkeitstheorie I* (WT1) wiederholt. Eine detaillierte Darstellung findet sich zum Beispiel in Bauer [1992], Elstrodt [2011] oder Klenke [2020].

B01 Maßraum

B01.01 **Definition.** Sei (Ω, \mathcal{A}) ein messbarer Raum und $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}^+$ eine Mengenfunktion. μ heißt

(a) *additiv*, falls für je endlich viele paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{A}$ gilt, dass

$$\mu\left(\biguplus_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i).$$

(b) *σ -additiv*, falls für je abzählbar vieler paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt,

$$\text{dass } \mu\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Eine σ -additive Mengenfunktion μ mit $\mu(\emptyset) = 0$ heißt *Maß*. Ein Maß μ heißt

(d) *endlich*, falls $\mu(\Omega) < \infty$,

(e) *σ -endlich*, falls es eine Folge $(A_n)_{n \in \mathbb{N}}$ in \mathcal{A} gibt mit $\Omega = \bigcup_{n=1}^{\infty} A_n$ und $\mu(A_n) < \infty$ für jedes $n \in \mathbb{N}$. □

B01.02 **Definition.**

(a) Wir bezeichnen mit $\mathcal{M}(\cdot) = \mathcal{M}(\cdot, \mathcal{A})$ die Menge aller Maße auf \mathcal{A} . Ein Tripel $(\Omega, \mathcal{A}, \mu)$ bestehend aus einer Ergebnismenge Ω , einer σ -Algebra \mathcal{A} sowie einem Maß μ auf \mathcal{A} wird *Maßraum* genannt.

(b) Ist Ω abzählbar, so wir der Maßraum $(\Omega, 2^\Omega, \mu)$ *diskret* genannt.

(c) Die Menge aller σ -endliche Maße auf \mathcal{A} bezeichnen wir mit $\mathcal{M}_\sigma(\mathcal{A})$. Die Teilmenge aller endliche Maße auf \mathcal{A} bezeichnen wir mit $\mathcal{M}_e(\mathcal{A})$ und zur Erinnerung $\mathcal{W}(\mathcal{A})$ bezeichnet die Teilmenge der Wahrscheinlichkeitsmaße (Definition A01.04 (b)).

(d) Das *Bildmaß* von $\mu \in \mathcal{M}(\mathcal{A})$ auf $(\mathcal{S}, \mathcal{S})$ unter einer messbaren Abbildung $f : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ bezeichnen wir weiterhin mit $\mu^f := \mu \circ f^{-1}$. □

B01.03 **Beispiel.**

(a) Für $A \in 2^\Omega$ bezeichnet $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ mit $\mathbb{1}_A^{-1}(\{1\}) = A$ und $\mathbb{1}_A^{-1}(\{0\}) = A^c$ die *Indikatorfunktion* auf A . Für jede σ -Algebra $\mathcal{A} \subseteq 2^\Omega$ und jedes $\omega \in \Omega$ ist das *Einpunkt-* oder *Diracmaß* $\delta_\omega : \mathcal{A} \rightarrow \{0, 1\}$ mit $\delta_\omega(A) := \mathbb{1}_A(\omega)$ ein Wahrscheinlichkeitsmaß, also $\delta_\omega \in \mathcal{M}_1(\mathcal{A})$.

(b) Sei $\Omega \neq \emptyset$ abzählbar und $\mathbb{p} : \Omega \rightarrow \mathbb{R}^+$. Dann ist $\mu : 2^\Omega \rightarrow \overline{\mathbb{R}}^+$ mit $A \mapsto \mu(A) := \sum_{\omega \in \Omega} \mathbb{p}(\omega) \delta_\omega(A)$ ein σ -endliches Maß auf 2^Ω , also $\mu \in \mathcal{M}_\sigma(2^\Omega)$. Wir bezeichnen \mathbb{p} als *Zähl-dichte* von μ . Ist $\Omega \subseteq \mathbb{R}^k$ so definiert $B \mapsto \mu(B)$ ein σ -endliches Maß auf $(\mathbb{R}^k, \mathcal{B}^k)$. Zur Erinnerung, erfüllt \mathbb{p} zusätzlich $\sum_{\omega \in \Omega} \mathbb{p}(\omega) = 1$ dann ist $\mu \in \mathcal{W}(2^\Omega) = \mathcal{M}_1(2^\Omega)$ ein diskretes Wahrscheinlichkeitsmaß. Im Spezialfall $\mathbb{p}(\omega) = 1$ für jedes $\omega \in \Omega$ heißt $\zeta_\Omega := \mu$ das *Zählmaß* auf Ω .

- (c) Es existiert ein eindeutig bestimmtes Maß λ^n auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit der Eigenschaft $\lambda^n((a, b]) = \prod_{i \in [1, n]} (b_i - a_i)$ für alle $a, b \in \mathbb{R}^n$ mit $a < b$. λ^n heißt **Lebesgue-Maß** auf $(\mathbb{R}^n, \mathcal{B}^n)$ (vgl. Vorlesung Analysis 3). Für das Lebesgue-Maß auf $(\mathbb{R}, \mathcal{B})$ schreiben wir kurz $\lambda := \lambda^1$. \square

B01.04 **Definition.** Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

- (a) Eine Menge $N \in \mathcal{A}$ heißt **μ -Nullmenge**, oder kurz Nullmenge, falls $\mu(N) = 0$. Mit \mathcal{N}_μ bezeichnen wir das System aller Teilmengen von μ -Nullmengen.
- (b) Eine Aussage gilt **μ -fast überall (μ -f.ü.)**, wenn es eine μ -Nullmenge $N \in \mathcal{N}_\mu$ gibt, so dass die Aussage für alle $\omega \in \Omega \setminus N = N^c$ gilt. Ist $A \in \mathcal{A}$, so sagen wir, eine Aussage gilt **μ -fast überall auf A** , falls $N \in \mathcal{N}_\mu$ gibt, so dass die Aussage für alle $\omega \in A \setminus N$ gilt. Ist $\mu = \mathbb{P} \in \mathcal{W}(\mathcal{A})$ ein Wahrscheinlichkeitsmaß, so sagen wir dann auch, dass die Aussage **\mathbb{P} -fast sicher (\mathbb{P} -f.s.)** gilt, beziehungsweise \mathbb{P} -fast sicher auf A .
- (c) Der Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt **vollständig**, falls $\mathcal{N}_\mu \subseteq \mathcal{A}$.
- (d) Für $B \in \mathcal{A}$ wird durch $\mu|_B(A) := \mu(A)$ für $A \in \mathcal{A}$ mit $A \subseteq B$ ein Maß auf der Spur $\mathcal{A}|_B := \mathcal{A} \cap B$ von \mathcal{A} über B definiert. Dieses Maß wird **Einschränkung** von μ auf B genannt. \square

B02 Integral

B02.01 **Definition.** Für jedes Maß μ auf einem Messraum (Ω, \mathcal{A}) heißt **Integral** bezüglich μ das eindeutig bestimmte Funktional $\mathbb{I}_\mu : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$, das die folgenden Bedingungen erfüllt:

- (I1) Für alle $f, g \in \overline{\mathcal{A}}^+$ und $a, b \in \mathbb{R}^+$ gilt $\mathbb{I}_\mu(af + bg) = a\mathbb{I}_\mu(f) + b\mathbb{I}_\mu(g)$; *(linear)*
- (I2) Für alle $(f_n)_{n \in \mathbb{N}} \uparrow f$ in $\overline{\mathcal{A}}^+$ gilt $\mathbb{I}_\mu(f_n) \uparrow \mathbb{I}_\mu(f)$; *(monoton konvergent)*
- (I3) Für jedes $A \in \mathcal{A}$ gilt $\mathbb{I}_\mu(\mathbb{1}_A) = \mu(A)$. *(normiert)*

Für jedes $f \in \overline{\mathcal{A}}^+$ heißt $\int f d\mu := \mathbb{I}_\mu(f)$ das **Integral** von f bezüglich μ , wobei $\mathbb{I}_\mu(f) = \sup_{\mathcal{P} \in \mathcal{P}} \left\{ \sum_{A \in \mathcal{P}} \left(\inf_{\omega \in A} f(\omega) \right) \mu(A) \right\}$ für $\mathcal{P} := \{ \mathcal{P} \subseteq \mathcal{A} \mid \mathcal{P} \text{ endliche Partition von } \Omega \}$. Für $A \in \mathcal{A}$ schreiben wir kurz $\int_A f d\mu := \int (f\mathbb{1}_A) d\mu$. \square

B02.02 **Schreibweise.** Für $f \in \overline{\mathcal{A}}^+$ und $A \in \mathcal{A}$ schreiben wir auch $\mu(f) = \int f d\mu = \int_\Omega f(\omega) \mu(d\omega)$ sowie $\mu(f\mathbb{1}_A) = \int_A f d\mu = \int_A f(\omega) \mu(d\omega)$. \square

B02.03 **Eigenschaft.** Sei $(\Omega, \mathcal{A}, \mu)$ ein beliebiger Maßraum und $(f_n)_{n \in \mathbb{N}}$ eine Folge aus $\overline{\mathcal{A}}^+$.

- (i) (**Lemma von Fatou**) Dann gilt $\mu\left(\liminf_{n \rightarrow \infty} f_n\right) = \int \left(\liminf_{n \rightarrow \infty} f_n\right) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu$. Insbesondere für jede Folge $(A_n)_{n \in \mathbb{N}}$ von Mengen aus \mathcal{A} gilt $\mu\left(\liminf_{n \rightarrow \infty} A_n\right) \leq \liminf_{n \rightarrow \infty} \mu(A_n)$. Ist $\mu \in \mathcal{M}_e(\mathcal{A})$ endlich, so gilt außerdem $\limsup_{n \rightarrow \infty} \mu(A_n) \leq \mu\left(\limsup_{n \rightarrow \infty} A_n\right)$.
- (ii) Es gilt $\sum_{n \in \mathbb{N}} f_n \in \overline{\mathcal{A}}^+$ und $\mu\left(\sum_{n \in \mathbb{N}} f_n\right) = \int \left(\sum_{n \in \mathbb{N}} f_n\right) d\mu = \sum_{n \in \mathbb{N}} \mu(f_n)$.
Seien weiterhin $f, g \in \overline{\mathcal{A}}^+$.
- (iii) Genau dann ist $f = 0$ μ -f.ü., wenn $\mu(f) = \int f d\mu = 0$ gilt. Ist $\mu(f) \in \mathbb{R}^+$, so gilt $f \in \mathbb{R}^+$ μ -f.ü. und die Einschränkung von μ auf $\{f \neq 0\}$ ist ein σ -endliches Maß.
- (iv) Die Mengenfunktion $f\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}^+$ mit $A \mapsto f\mu(A) := \mu(\mathbb{1}_A f) = \int (\mathbb{1}_A f) d\mu$ ist ein Maß auf (Ω, \mathcal{A}) . Für alle $A \in \mathcal{A}$ mit $\mu(A) = 0$ gilt $f\mu(A) = 0$.

- (v) Ist $f \leq g$ (bzw. $f = g$) μ -f.ü., so gilt $f\mu \leq g\mu$ (bzw. $f\mu = g\mu$).
Ist (c1) f μ -integrierbar, oder (c2) $\mu \in \mathcal{M}_\sigma(\mathcal{A})$, oder (c3) $g\mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, so gilt auch die Umkehrung.
Insbesondere, gilt $\mu(f) = \int f d\mu \leq \int g d\mu = \mu(g)$ (bzw. $\mu(f) = \mu(g)$).
- (vi) Genau dann ist $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, wenn es $h \in \mathcal{A}_0^+$ mit $\mu(h) \in \mathbb{R}^+$ (μ -integrierbar) gibt. Insbesondere, existiert zu jedem σ -endlichen $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ ein $h \in \mathcal{A}_0^+$, derart dass $h\mu \in \mathcal{M}_\sigma(\mathcal{A})$ endlich ist und $h\mu$ besitzt die gleichen Nullmengen wie μ .
- (vii) Für $f \in \mathcal{A}$ mit $f \geq 0$ μ -f.ü. gilt $\sum_{n \in \mathbb{N}} \mu(\{f \geq n\}) \leq \int f d\mu \leq \sum_{n \in \mathbb{N}_0} \mu(\{f > n\})$ und $\int f d\mu = \int_0^\infty \mu(\{f \geq t\}) dt$. \square

B02.04 **Definition.** Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f \in \overline{\mathcal{A}}^+$. Wir sagen, dass das durch $\nu(A) := \int (\mathbb{1}_A f) d\mu$ für $A \in \mathcal{A}$ definierte Maß $f\mu := \nu$ die *Dichte* $d\nu/d\mu := \frac{d\nu}{d\mu} := f$ bezüglich μ besitzt.

B02.05 **Eigenschaft.** Seien $\mu \in \mathcal{M}(\mathcal{A})$ ein Maß auf (Ω, \mathcal{A}) und $\nu = f\mu$ mit $\frac{d\nu}{d\mu} = f \in \overline{\mathcal{A}}^+$.

- (i) Für jede Funktion $g \in \overline{\mathcal{A}}^+$ gilt $\int g d\nu = \int (gf) d\mu = \mu(gf) = f\mu(g) = \nu(g)$.
- (ii) Sei $\rho := q\nu$ mit $q \in \overline{\mathcal{A}}^+$. Dann gilt $\rho = q\nu = q(f\mu) = (qf)\mu$.
- (iii) Die μ -Dichte $d\nu/d\mu$ von ν ist eindeutig bis auf Gleichheit μ -fast überall, wenn $\nu \in \mathcal{M}_\sigma(\mathcal{A})$ oder $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich ist.
- (iv) Ist $\nu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, so ist $\frac{d\nu}{d\mu} = f \in \mathbb{R}^+$ μ -f.ü.. Die Umkehrung gilt, wenn $\mu \in \mathcal{M}_\sigma(\mathcal{A})$. \square

B02.06 **Schreibweise.** Ist $f \in \overline{\mathcal{A}}$ eine numerische Funktion, so sind $f^+ := f \vee 0$, $f^- := (-f)^+$, $f^+ + f^- = |f| \in \overline{\mathcal{A}}^+$ positive numerische Funktionen (vgl. **Eigenschaft** A02.06). \square

B02.07 **Definition.** Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f \in \overline{\mathcal{A}}$ eine numerische Funktion.

- (a) Ist höchstens eines der beiden Integrale $\int f^+ d\mu$ und $\int f^- d\mu$ nicht endlich, dass heißt, $\mu(f^+) \wedge \mu(f^-) \in \mathbb{R}^+$, so definiert $\int f d\mu := \int f^+ d\mu - \int f^- d\mu$ das *Integral* von f bezüglich μ mit den üblichen Konventionen $\infty + x = \infty$ und $-\infty + x = -\infty$ für alle $x \in \mathbb{R}$. In diesem Fall wird f *quasiintegrierbar* genannt. Das Integral von f ist nicht definiert, wenn $\int f^+ d\mu = \infty = \int f^- d\mu$ gilt.
- (b) Falls $\int |f| d\mu \in \mathbb{R}^+$, also falls $\int f^+ d\mu \in \mathbb{R}^+$ und $\int f^- d\mu \in \mathbb{R}^+$, gilt, dann heißt f μ -*integrierbar*. Die Menge aller μ -integrierbaren numerischen Funktionen bezeichnen wir mit $\mathcal{L}_1 := \mathcal{L}_1(\mu) := \mathcal{L}_1(\mathcal{A}, \mu) := \{f \in \overline{\mathcal{A}} : \int |f| d\mu \in \mathbb{R}^+\}$.
- (c) Für $p \in \mathbb{R}_0^+$ definiere $\|f\|_{\mathcal{L}_p} := (\int (|f|^p) d\mu)^{1/p}$ und $\|f\|_{\mathcal{L}_\infty} := \inf \{x \in \mathbb{R}^+ : \mu(\{|f| > x\}) = 0\}$. Für $p \in \overline{\mathbb{R}}_0^+$ heißt f \mathcal{L}_p -*integrierbar*, wenn $\|f\|_{\mathcal{L}_p} \in \mathbb{R}^+$. Die Menge aller \mathcal{L}_p -integrierbaren Funktionen bezeichnen wir mit $\mathcal{L}_p := \mathcal{L}_p(\mu) := \mathcal{L}_p(\mathcal{A}, \mu) := \{f \in \overline{\mathcal{A}} : \|f\|_{\mathcal{L}_p} \in \mathbb{R}^+\}$. Für $p \in [1, \infty]$ ist $\|\cdot\|_{\mathcal{L}_p}$ eine Pseudonorm auf dem Vektorraum \mathcal{L}_p .
- (d) $\langle \cdot, \cdot \rangle_{\mathcal{L}_2} : \mathcal{L}_2(\mu) \times \mathcal{L}_2(\mu) \rightarrow \mathbb{R}$ mit $(f, g) \mapsto \langle f, g \rangle_{\mathcal{L}_2} := \int fg d\mu$ ist eine positive semidefinite symmetrische Bilinearform. \square

B02.08 **Bemerkung.** Für $X \in \mathcal{L}_1(\mathbb{P})$ entspricht damit der Erwartungswert von X dem Lebesgue-Integral von X bzgl. des Wahrscheinlichkeitsmaßes \mathbb{P} , dass heißt, $E(X) = \int_\Omega X(\omega) \mathbb{P}(d\omega)$ und wir erhalten die bekannten Darstellungen:

- (i) Sei \mathbb{P} ein *stetiges* Wahrscheinlichkeitsmaß auf \mathbb{R} mit Dichte f bezüglich des Lebesgue-Maßes λ . Dann gilt $X \in \mathcal{L}_1(\mathbb{P})$ genau dann, wenn $\int_{\mathbb{R}^n} |X(\omega)| f(\omega) \lambda(d\omega) < \infty$, wobei wir

auch weiterhin die Schreibweise $\int_{\mathbb{R}^n} |X(\omega)| f(\omega) d\omega$ benutzen. In diesem Fall ist

$$\mathbb{E}(X) = \int_{\mathbb{R}} X(\omega) f(\omega) d\omega.$$

- (ii) Sei \mathbb{P} ein *diskretes* Wahrscheinlichkeitsmaß mit Zähldichte \mathbb{p} . Dann gilt $X \in \mathcal{L}_1(\mathbb{P})$ genau dann, wenn $\int_{\Omega} |X(\omega)| \mathbb{P}(d\omega) = \sum_{\omega \in \Omega} |X(\omega)| \mathbb{p}(\omega) < \infty$. In diesem Fall ist

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{p}(\omega).$$

□

B02.09 **Lemma (Eigenschaften).** Seien $f, g \in \mathcal{L}_1(\mathcal{A}, \mu)$.

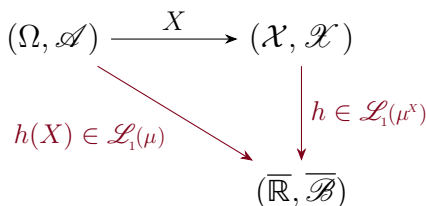
- (i) Für alle $a, b \in \mathbb{R}$ gilt $af + bg \in \mathcal{L}_1(\mu)$ und $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$; **(linear)**
- (ii) Sei $h \in \overline{\mathcal{A}}$. Ist $h = f$ μ -f.ü., dann gilt $h \in \mathcal{L}_1(\mu)$ und $\int h d\mu = \int f d\mu$.
Ist $|h| \leq g$ μ -f.ü., dann gilt $h \in \mathcal{L}_1(\mu)$.
- (iii) Ist $f \leq g$ μ -f.ü., so ist $\int f d\mu \leq \int g d\mu$. **(monoton)**
Insbesondere aus $f \in \mathbb{R}^+$ μ -f.ü. folgt $\int f d\mu \in \mathbb{R}^+$. **(positiv)**
- (iv) Es gilt $|\int f d\mu| \leq \int |f| d\mu$. **(Dreiecksungleichung)**
- (v) Es gilt $f = 0$ μ -f.ü. genau dann, wenn $\int_A f d\mu = 0$ für alle $A \in \mathcal{A}$ gilt.
- (vi) Ist $\mu \in \mathcal{M}_e(\mathcal{A})$ endlich und $h \in \mathcal{A}$ **beschränkt**, also $\sup_{\omega \in \Omega} |h(\omega)| < \infty$, so gilt $h \in \mathcal{L}_1(\mu)$.
- (vii) Für $\mu, \nu \in \mathcal{M}(\mathcal{A})$ ist $h \in \mathcal{L}_1(\mu + \nu)$ genau dann wenn $h \in \mathcal{L}_1(\mu) \cap \mathcal{L}_1(\nu)$ gilt. In diesem Fall ist $\int h d(\mu + \nu) = \int h d\mu + \int h d\nu$.
- (viii) Sei $\nu = f\mu$ mit $\frac{d\nu}{d\mu} = f \in \overline{\mathcal{A}}^+$. Eine Funktion $g \in \overline{\mathcal{A}}$ ist genau dann ν -integrierbar, wenn gf μ -integrierbar ist. In diesem Fall gilt $\nu(g) = \mu(gf) = \int (gf) d\mu = \int g d(f\mu) = \int g d\nu$.

Seien nun $f, g \in \overline{\mathcal{A}}$.

- (ix) Sei $p \in \mathbb{R}_0^+$. Genau dann ist $f \in \mathcal{L}_p(\mu)$, wenn $|f|^p \in \mathcal{L}_1(\mu)$. Es gilt $\mu(\{|f| > \|f\|_{\mathcal{L}_\infty}\}) = 0$.
- (x) Sei $p \in \mathbb{R}_0^+$. Genau dann ist $\|f\|_{\mathcal{L}_p} = 0$, wenn $f = 0$ μ -f.ü.. Für $a \in \mathbb{R}$ gilt $\|af\|_{\mathcal{L}_p} = |a| \|f\|_{\mathcal{L}_p}$. Ist $f \in \mathcal{L}_p(\mu)$ und $f = g$ μ -f.ü., so gilt $|f| < \infty$ μ -f.ü. sowie $\|f\|_{\mathcal{L}_p} = \|g\|_{\mathcal{L}_p}$.

Seien weiterhin $(\mathcal{X}, \mathcal{X})$ ein Messraum, $X : \Omega \rightarrow \mathcal{X}$ eine \mathcal{A} - \mathcal{X} -messbare Funktion und $\mu^X := \mu \circ X^{-1} \in \mathcal{M}(\mathcal{X})$ das Bildmaß von μ unter X auf $(\mathcal{X}, \mathcal{X})$.

- (xi) Für $h \in \mathcal{X}^+$ gilt dann $\int h(X) d\mu = \int h d\mu^X$. Genau dann ist $h \in \overline{\mathcal{X}}$ integrierbar bezüglich μ^X , wenn $h(X) \in \overline{\mathcal{A}}$ integrierbar bezüglich μ ist, und dann gilt ebenfalls $\int h(X) d\mu = \int h d\mu^X$.



Ist speziell X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, so gilt

$$\int h(x) \mathbb{P}^X(dx) = \int h d\mathbb{P}^X = \mathbb{P}^X(h) = \mathbb{P}(h(X)) = \int h(X) d\mathbb{P} = \int h(X(\omega)) \mathbb{P}(d\omega). \quad \square$$

B03 Konvergenzarten

B03.01 Definition. Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Eine Folge $(f_n)_{n \in \mathbb{N}}$ aus $\overline{\mathcal{A}}$ konvergiert gegen $f \in \overline{\mathcal{A}}$

μ -fast überall (μ -f.ü.), kurz $f_n \xrightarrow{\mu\text{-f.ü.}} f$, wenn $\limsup_{n \rightarrow \infty} |f_n - f| = 0$ μ -f.ü. gilt. Es also eine μ -Nullmenge $N \in \mathcal{A}$ gibt, sodass für jedes $\omega \in N^c := \Omega \setminus N$ gilt, dass $\lim_{n \rightarrow \infty} |f_n(\omega) - f(\omega)| = 0$.

μ -fast vollständig (μ -f.v.), kurz $f_n \xrightarrow{\mu\text{-f.v.}} f$, wenn für jedes $A \in \mathcal{A}$ mit $\mu(A) \in \mathbb{R}^+$ und für jedes $\varepsilon \in \mathbb{R}_0^+$ gilt $\sum_{n \in \mathbb{N}} \mu(\{|f_n - f| > \varepsilon\} \cap A) \in \mathbb{R}^+$.

μ -stochastisch (oder dem Maße μ nach), kurz $f_n \xrightarrow{\mu} f$, wenn für jedes $A \in \mathcal{A}$ mit $\mu(A) \in \mathbb{R}^+$ und für jedes $\varepsilon \in \mathbb{R}_0^+$ gilt $\lim_{n \rightarrow \infty} \mu(\{|f_n - f| > \varepsilon\} \cap A) = 0$.

im p -ten Mittel (oder in $\mathcal{L}_p(\mu)$), $p \in \overline{\mathbb{R}}_0^+$, kurz $f_n \xrightarrow{\mathcal{L}_p(\mu)} f$, wenn $(f_n)_{n \in \mathbb{N}}$ und f aus $\mathcal{L}_p(\mu)$ sind und es gilt $\lim_{n \rightarrow \infty} \|f_n - f\|_{\mathcal{L}_p} = 0$.

Gelegentlich verwenden wir abkürzend $f_n \xrightarrow{\text{f.ü.}} f$ oder auch $f_n \xrightarrow{\mathcal{L}_p} f$ wenn das zu Grund liegende Maß μ aus dem Kontext hervorgeht. □

B03.02 Eigenschaft. Sei $(\Omega, \mathcal{A}, \mu)$ ein beliebiger Maßraum.

(i) **(Monotone Konvergenz)** Sei $(f_n)_{n \in \mathbb{N}}$ aus $\mathcal{L}_1(\mu)$ und $f \in \overline{\mathcal{A}}$ mit $f_n \uparrow f$ μ -f.ü.. Dann gilt $\int f_n d\mu \uparrow \int f d\mu$.

(ii) **(Dominierte Konvergenz)** Sei $(f_n)_{n \in \mathbb{N}}$ aus $\overline{\mathcal{A}}$ μ -f.ü. konvergent und es gebe $g \in \mathcal{L}_1(\mu)$ mit $\sup_{n \in \mathbb{N}} |f_n| \leq g$ μ -f.ü.. Dann existiert ein $f \in \overline{\mathcal{A}}$ mit $\lim_{n \rightarrow \infty} f_n = f$ μ -f.ü., $(f_n)_{n \in \mathbb{N}}$ und f sind aus $\mathcal{L}_1(\mu)$ und es gilt $\lim_{n \rightarrow \infty} \int (|f - f_n|) d\mu = 0$ sowie $\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$. Ist $g \in \mathcal{L}_p(\mu)$ für $p \in [1, \infty)$ dann sind auch $(f_n)_{n \in \mathbb{N}}$ und f aus $\mathcal{L}_p(\mu)$ und es gilt $\lim_{n \rightarrow \infty} \|f_n - f\|_{\mathcal{L}_p} = 0$ sowie $\lim_{n \rightarrow \infty} \|f_n\|_{\mathcal{L}_p} = \|f\|_{\mathcal{L}_p}$.

(iii) **(Satz von Scheffé)** Seien $(f_n)_{n \in \mathbb{N}}$ und f aus $\overline{\mathcal{A}}^+$ integrierbar und es gelte $f_n \xrightarrow{\mu\text{-f.ü.}} f$ sowie $\int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f d\mu$. Dann gilt $f_n \xrightarrow{\mathcal{L}_1(\mu)} f$.

(iv) **(Satz von Riesz)** Seien $(f_n)_{n \in \mathbb{N}}$ und f aus $\mathcal{L}_p(\mu)$ mit $p \in [1, \infty)$ und es gelte $f_n \xrightarrow{\mu\text{-f.ü.}} f$. Dann gilt $\int |f_n|^p d\mu \xrightarrow{n \rightarrow \infty} \int |f|^p d\mu$ genau dann, wenn $f_n \xrightarrow{\mathcal{L}_p(\mu)} f$.

(v) Für f und $(f_n)_{n \in \mathbb{N}}$ aus \mathcal{A} gilt $f_n \xrightarrow{\mu\text{-f.ü.}} f \implies f_n \xrightarrow{\mu} f \iff f_n \xrightarrow{\mathcal{L}_p(\mu)} f$. Ist $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, dann gilt auch $f_n \xrightarrow{\mu\text{-f.v.}} f \implies f_n \xrightarrow{\mu\text{-f.ü.}} f$. Weiterhin gilt $f_n \xrightarrow{\mu} f$ genau dann, wenn zu jeder Teilfolge von $(f_n)_{n \in \mathbb{N}}$ eine gegen f μ -fast überall konvergente Teilfolge existiert. □

B03.03 Definition. Eine Familie \mathcal{F} aus $\mathcal{L}_1(\mu)$ heißt **gleichgradig μ -integrierbar**, wenn

$$\inf \left\{ \sup_{f \in \mathcal{F}} \mu(|f| \mathbb{1}_{\{|f| \geq g\}}) : g \in \mathcal{L}_1(\mu) \cap \overline{\mathcal{A}}^+ \right\} = 0.$$

Ist $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ endlich, so ist die gleichgradige μ -Integrierbarkeit äquivalent zu

$$\inf \left\{ \sup_{f \in \mathcal{F}} \mu(|f| \mathbb{1}_{\{|f| \geq a\}}) : a \in \mathbb{R}^+ \right\} = 0. \quad \square$$

B03.04 Bemerkung.

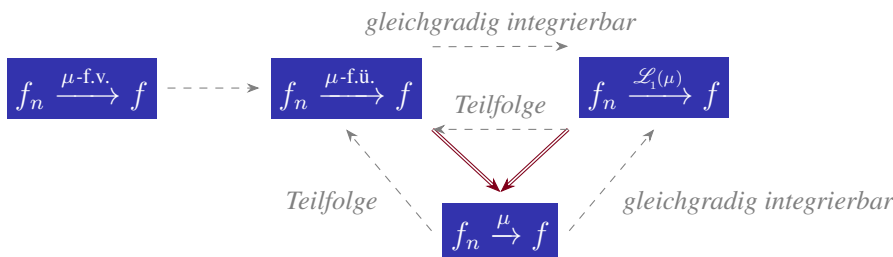
(i) Ist $\mathcal{F} \subseteq \overline{\mathcal{A}}$ und $g \in \mathcal{L}_p(\mu) \cap \overline{\mathcal{A}}^+$ mit $|f| \leq g$ μ -f.ü. für alle $f \in \mathcal{F}$, dann ist $\mathcal{F}^p := \{|f|^p : f \in \mathcal{F}\}$ gleichgradig μ -integrierbar. □

B03.05 **Eigenschaft.** Sei $(\Omega, \mathcal{A}, \mu)$ ein beliebiger Maßraum. Sei $(f_n)_{n \in \mathbb{N}}$ reellwertig aus $\mathcal{L}_p(\mu)$, $p \in [1, \infty)$. Dann sind äquivalent:

- (i) $(f_n)_{n \in \mathbb{N}}$ konvergiert in $\mathcal{L}_p(\mu)$;
- (ii) $(|f_n|^p)_{n \in \mathbb{N}}$ ist gleichgradig μ -integrierbar und $(f_n)_{n \in \mathbb{N}}$ ist konvergent dem Maße μ nach. \square

B03.06 **Zusammenfassung.** Seien $(\Omega, \mathcal{A}, \mu)$ ein beliebiger Maßraum, $p \in [1, \infty]$ und $(f_n)_{n \in \mathbb{N}}$ aus $\mathcal{L}_p(\mu)$. Dann sind die folgenden Aussagen äquivalent:

- (i) Es gibt ein $f \in \mathcal{L}_p(\mu)$ mit $f_n \xrightarrow{\mathcal{L}_p(\mu)} f$.
- (ii) $(f_n)_{n \in \mathbb{N}}$ ist eine $\mathcal{L}_p(\mu)$ -Cauchy Folge, also $\lim_{n,m \rightarrow \infty} \|f_n - f_m\|_{\mathcal{L}_p} = 0$.
Ist $p < \infty$ und $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, so sind (i) und (ii) zudem äquivalent zu
- (iii) $(|f_n|^p)_{n \in \mathbb{N}}$ ist gleichgradig integrierbar, und es gibt $f \in \mathcal{A}$ mit $f_n \xrightarrow{\mu} f$.
Die Limiten in (i) und (iii) stimmen überein.



Die Abbildung wurde auf der Grundlage von Klenke [2020, Abb.6.1, S.159] erstellt. \square

B03.07 **Bemerkung.** Zur Erinnerung für $p \in \overline{\mathbb{R}}_0^+$ und $f, g \in \overline{\mathcal{A}}$ gilt $\|f - g\|_{\mathcal{L}_p} = 0$ genau dann, wenn $f = g$ μ -f.ü.. In diesem Fall sehen wir f und g als äquivalent an. Sei nun $\mathcal{N} = \{f \in \overline{\mathcal{A}} : f = 0 \text{ } \mu\text{-f.ü.}\}$, dann ist \mathcal{N} ein Untervektorraum von $\mathcal{L}_p(\mu)$, so dass wir formal den Quotientenraum $\mathbb{L}_p := \mathbb{L}_p(\mu) := \mathbb{L}_p(\mathcal{A}, \mu) := \{[f] := f + \mathcal{N} : f \in \mathcal{L}_p(\mu)\}$ bilden können. Für $[f] \in \mathbb{L}_p(\mu)$ setzen wir $\|[f]\|_{\mathbb{L}_p} := \|f\|_{\mathcal{L}_p}$ für ein $f \in [f]$ und $\int [f] d\mu := \int f d\mu$, falls der Ausdruck für f definiert ist. Dabei hängt $\|[f]\|_{\mathbb{L}_p}$ nicht von der Wahl des Repräsentanten $f \in [f]$ ab. Analog für $[f], [g] \in \mathbb{L}_2(\mu)$ definieren wir $\langle [f], [g] \rangle_{\mathbb{L}_2} := \langle f, g \rangle_{\mathcal{L}_2}$, wobei $f \in [f]$ und $g \in [g]$. \square

B03.08 **Eigenschaft.** Sei $(\Omega, \mathcal{A}, \mu)$ ein beliebiger Maßraum.

- (i) **(Hölder Ungleichung)** Seien $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt $\|fg\|_{\mathcal{L}_1} \leq \|f\|_{\mathcal{L}_p} \|g\|_{\mathcal{L}_q}$.
(Cauchy-Schwarz Ungleichung) Insbesondere gilt $|\langle f, g \rangle_{\mathcal{L}_2}| \leq \|f\|_{\mathcal{L}_2} \|g\|_{\mathcal{L}_2}$.
- (ii) Seien $\mu \in \mathcal{M}_e(\mathcal{A})$ endlich, $q \in \overline{\mathbb{R}}_0^+$ und $p \in (0, q)$. Dann gilt $\mu(\Omega)^{1/q} \|f\|_{\mathcal{L}_p} \leq \mu(\Omega)^{1/p} \|f\|_{\mathcal{L}_q}$ und somit $\mathcal{L}_q \subseteq \mathcal{L}_p$.
- (iii) **(Minkowski Ungleichung)** Für $p \in [1, \infty]$ gilt $\|f + g\|_{\mathcal{L}_p} \leq \|f\|_{\mathcal{L}_p} + \|g\|_{\mathcal{L}_p}$.
- (iv) **(Fischer-Riesz)** Für $p \in [1, \infty]$ ist $(\mathbb{L}_p(\mu), \|\cdot\|_{\mathbb{L}_p})$ ein Banachraum.
Insbesondere ist $(\mathbb{L}_2(\mu), \langle \cdot, \cdot \rangle_{\mathbb{L}_2})$ ein reeller Hilbertraum. \square

B04 Maße mit Dichten - Satz von Radon-Nikodym

B04.01 **Definition.** Seien $\nu, \mu \in \mathcal{M}(\mathcal{A})$ Maße auf (Ω, \mathcal{A}) .

$\nu \ll \mu$: ν heißt **(absolut-)stetig** bezüglich μ , kurz μ -stetig, oder **dominiert** durch μ , wenn jede μ -Nullmenge aus \mathcal{A} auch eine ν -Nullmenge ist, also für jedes $A \in \mathcal{A}$ mit $\mu(A) = 0$ gilt auch $\nu(A) = 0$. Die Maße μ und ν heißen **äquivalent** (kurz $\mu \ll \nu$), falls $\nu \ll \mu$ und $\mu \ll \nu$.

$\mu \perp \nu$: μ heißt *singulär* zu ν , kurz *ν -singulär*, wenn es eine μ -Nullmenge $N \in \mathcal{A}$ gibt mit $\nu = \mathbb{1}_N \nu$, also $\nu(A) = \nu(A \cap N)$ für alle $A \in \mathcal{A}$, oder äquivalent dazu $N^c := \Omega \setminus N$ ist eine ν -Nullmenge. \square

B04.02 Eigenschaft.

- (i) Sind $\nu, \mu \in \mathcal{M}_e(\mathcal{A})$ endliche Maße mit $\nu \leq \mu$, so gibt es eine meßbare Funktion $h : \Omega \rightarrow [0, 1]$ mit $\nu = h\mu$.
- (ii) (*Satz von Radon-Nikodym*) Seien $\mu \in \mathcal{M}_\sigma(\mathcal{A})$ ein σ -endliches Maß und $\nu \in \mathcal{M}(\mathcal{A})$ ein μ -stetiges Maß auf (Ω, \mathcal{A}) , also $\nu \ll \mu$. Dann hat ν eine Dichte $f = d\nu/d\mu \in \overline{\mathcal{A}}^+$ bezüglich μ , d.h. es gilt $\nu = f\mu$. \square

B04.03 **Bemerkung.** Es seien $\mu, \nu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endliche Maße und $f = d\nu/d\mu \in \overline{\mathcal{A}}^+$ eine μ -Dichte von $\nu \ll \mu$. Dann folgen aus **Eigenschaft** B02.05 direkt die üblichen *Kettenregeln*:

- (i) Ist $g \in \overline{\mathcal{A}}$ quasiintegrierbar, so gilt $\int_A g d\nu = \int_A gf d\mu$ für alle $A \in \mathcal{A}$.
- (ii) Ist $\rho \in \mathcal{M}_\sigma(\mathcal{A})$ ein σ -endliches Maß mit $\rho \ll \nu \ll \mu$ so gilt $\frac{d\rho}{d\mu} = \frac{d\rho}{d\nu} \frac{d\nu}{d\mu}$ μ -f.ü..
- (iii) Ist $h : \Omega \rightarrow [0, 1]$ messbar mit $h = \frac{d\nu}{d(\nu+\mu)}$ μ -f.ü., so gilt $\frac{d\nu}{d\mu} = \frac{h}{1-h}$ μ -f.ü.. \square

B04.04 Beispiel.

- (a) *Stetige Wahrscheinlichkeitsmaße* auf $(\mathbb{R}^k, \mathcal{B}^k)$ (vgl. **Definition** A01.04 (f)) sind somit gerade die bezüglich des Lebesgue-Maßes λ^k absolutstetigen Wahrscheinlichkeitsmaße mit entsprechender (Radon-Nikodym-) Dichte.
- (b) Bezeichnen wir mit ζ_Ω das Zählmaß auf einer abzählbaren Menge Ω , so sind *diskrete Wahrscheinlichkeitsmaße* auf $(\Omega, 2^\Omega)$ (vgl. **Definition** A01.04 (e)) absolutstetig bezüglich ζ_Ω und ihre Zähldichte entspricht gerade der (Radon-Nikodym-) Dichte bezüglich ζ_Ω . Ist $\Omega \subseteq \mathbb{R}$ und \mathbb{P} ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit Zähldichte \mathfrak{p} wie in **Beispiel** A01.06 (b), dann ist \mathbb{P} absolutstetig bezüglich des Zählmaßes ζ_Ω auf Ω aufgefasst als Maß auf $(\mathbb{R}, \mathcal{B})$ und \mathfrak{p} gerade die (Radon-Nikodym-) Dichte von \mathbb{P} bezüglich ζ_Ω . \square

B04.05 **Lebesgue'scher Zerlegungssatz.** Seien $\mu, \nu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endliche Maße auf (Ω, \mathcal{A}) . Dann gibt es genau eine Zerlegung $\nu = \nu_a + \nu_s$ von ν in zwei Maße $\nu_a, \nu_s \in \mathcal{M}(\mathcal{A})$, derart dass $\nu_a \ll \mu$ der μ -stetige Anteil und $\nu_s \perp \mu$ der μ -singuläre Anteil von ν bezüglich μ ist. Dabei sind $\nu_a, \nu_s \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, und genau dann $\nu_a, \nu_s \in \mathcal{M}_e(\mathcal{A})$ endlich, wenn $\nu \in \mathcal{M}_e(\mathcal{A})$ endlich ist. ν_a besitzt eine μ -Dichte $d\nu_a/d\mu \in \overline{\mathcal{A}}^+$, die μ -fast überall rellwertig ist. \square

B04.06 **Bemerkung.** Wir können in **Satz** B04.05 die numerische Funktion $f = d\nu_a/d\mu \in \overline{\mathcal{A}}^+$ durch die reelle Funktion $\tilde{f} := f \mathbb{1}_{\{f \in \mathbb{R}^+\}} \in \mathcal{A}^+$ ersetzen, da $f = \tilde{f}$ μ -f.ü.. Mit anderen Worten $\tilde{f} \in \mathcal{A}^+$ ist ebenfalls eine Festlegung der Radon-Nikodym Dichte von ν_a bezüglich μ . Sind $\nu, \mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endlich, so werden wir stets $d\nu_a/d\mu \in \mathcal{A}^+$ annehmen. Weiterhin definieren wir noch eine numerische Funktion $L := \tilde{f} \mathbb{1}_{N^c} + \infty \mathbb{1}_N \in \overline{\mathcal{A}}^+$ mit $\mu(N) = 0 = \nu_s(N^c)$, so ist $\{L = \infty\} = N$ und die Lebesgue-Zerlegung schreibt sich in der Form $\nu = L\mu + \mathbb{1}_{\{L=\infty\}}\nu$, d.h. für alle $A \in \mathcal{A}$ gilt $\nu(A) = \int_A L d\mu + \nu(A \cap \{L = \infty\})$. \square

B04.07 **Definition.** Seien $\nu, \mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endliche Maße auf (Ω, \mathcal{A}) , wobei nicht notwendig $\nu \ll \mu$ gilt. Dann heißt jede positive numerische Funktion $L \in \overline{\mathcal{A}}^+$ mit

$$\mu(L = \infty) = 0 \text{ und } \nu = L\mu + \mathbb{1}_{\{L=\infty\}}\nu \quad (04.01)$$

ein *Dichtequotient (DQ)* von ν bezüglich μ . \square

B04.08 **Eigenschaft.** Seien $\nu, \mu \in \mathcal{M}_\sigma(\mathcal{A})$ σ -endliche Maße. Dann ist der Dichtequotient $L \in \overline{\mathcal{A}}^+$ von ν bezüglich μ eindeutig bis auf Gleichheit $\nu + \mu$ -fast überall. \square

B05 Endliche Produktmaße

B05.01 **Erinnerung.** Seien $(\mathcal{S}_i, \mathcal{S}_i), i \in \mathcal{I}$, messbare Räume mit beliebiger, nicht-leerer Indexmenge \mathcal{I} . Die Menge $\mathcal{S}_{\mathcal{I}} := \times_{i \in \mathcal{I}} \mathcal{S}_i$ aller Abbildungen $(s_i)_{i \in \mathcal{I}} : \mathcal{I} \rightarrow \cup_{i \in \mathcal{I}} \mathcal{S}_i$ sodass $s_i \in \mathcal{S}_i$ für alle $i \in \mathcal{I}$ gilt, heißt **Produktraum** oder kartesisches Produkt. Sind alle \mathcal{S}_i gleich, etwa $\mathcal{S}_i = \mathcal{S}$, dann schreiben wir $\mathcal{S}^{\mathcal{I}} := \mathcal{S}_{\mathcal{I}}$, im Fall $n := |\mathcal{I}| < \infty$, auch nur kurz $\mathcal{S}^n := \mathcal{S}^{\mathcal{I}}$. Für jedes $\mathcal{J} \subseteq \mathcal{I}$ bezeichne $\Pi_{\mathcal{S}_j} : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_j$ mit $(s_i)_{i \in \mathcal{I}} \mapsto (s_j)_{j \in \mathcal{J}}$ die **kanonische Projektion** und speziell für $j \in \mathcal{I}$ die **Koordinatenabbildung** mit $\Pi_{\mathcal{S}_j} : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_j$ mit $(s_i)_{i \in \mathcal{I}} \mapsto s_j$, sodass $\times_{i \in \mathcal{I}} E_i = \bigcap_{i \in \mathcal{I}} \Pi_{\mathcal{S}_i}^{-1}(E_i)$ für $E_i \subseteq \mathcal{S}_i, i \in \mathcal{I}$. \square

B05.02 **Definition.**

- (a) Für eine Familie $(\mathcal{A}_i)_{i \in \mathcal{I}}$ von Teil- σ -Algebren von \mathcal{A} mit beliebiger nicht-leerer Indexmenge \mathcal{I} bezeichnet $\bigwedge_{i \in \mathcal{I}} \mathcal{A}_i := \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ und $\bigvee_{i \in \mathcal{I}} \mathcal{A}_i := \sigma(\cup_{i \in \mathcal{I}} \mathcal{A}_i)$ die **größte σ -Algebra**, die in allen $\mathcal{A}_i, i \in \mathcal{I}$, enthalten ist, bzw. die **kleinste σ -Algebra**, die alle $\mathcal{A}_i, i \in \mathcal{I}$, enthält.
- (b) Die Produkt- σ -Algebra $\mathcal{S}_{\mathcal{I}} := \bigotimes_{i \in \mathcal{I}} \mathcal{S}_i$ auf dem Produktraum $\mathcal{S}_{\mathcal{I}}$ ist die kleinste σ -Algebra, sodass für jedes $i \in \mathcal{I}$ die Koordinatenabbildung $\Pi_{\mathcal{S}_i} : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_i$ $\mathcal{S}_{\mathcal{I}}\text{-}\mathcal{S}_i$ -messbar ist, d.h.

$$\mathcal{S}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathcal{S}_i := \bigvee_{i \in \mathcal{I}} \sigma(\Pi_{\mathcal{S}_i}) = \bigvee_{i \in \mathcal{I}} \Pi_{\mathcal{S}_i}^{-1}(\mathcal{S}_i).$$

Sind alle $(\mathcal{S}_i, \mathcal{S}_i)$ gleich, etwa $(\mathcal{S}_i, \mathcal{S}_i) = (\mathcal{S}, \mathcal{S})$, dann schreiben wir $\mathcal{S}^{\mathcal{I}} := \mathcal{S}_{\mathcal{I}}$, im Fall $n := |\mathcal{I}| < \infty$, auch nur $\mathcal{S}^n := \mathcal{S}^{\mathcal{I}}$. \square

B05.03 **Eigenschaft.** Für jedes $i \in \llbracket n \rrbracket$ sei \mathcal{E}_i ein Erzeuger der σ -Algebra \mathcal{S}_i auf \mathcal{S}_i , welcher eine Folge $(E_{ik})_{k \in \mathbb{N}}$ von Mengen mit $E_{ik} \uparrow \mathcal{S}_i$ enthält. Dann wird die Produkt- σ -Algebra $\mathcal{S}_{\llbracket n \rrbracket} = \bigotimes_{i \in \llbracket n \rrbracket} \mathcal{S}_i$ vom System aller Mengen $\{ \times_{i \in \llbracket n \rrbracket} E_i : E_i \in \mathcal{E}_i, i \in \llbracket n \rrbracket \}$ erzeugt. \square

B05.04 **Definition.** Sei $\mu_i \in \mathcal{M}(\mathcal{S}_i)$ ein Maß auf $(\mathcal{S}_i, \mathcal{S}_i)$ für jedes $i \in \llbracket n \rrbracket$. Ein Maß $\mu_{\llbracket n \rrbracket} \in \mathcal{M}(\mathcal{S}_{\llbracket n \rrbracket})$ auf $(\mathcal{S}_{\llbracket n \rrbracket}, \mathcal{S}_{\llbracket n \rrbracket})$ heißt **Produktmaß**, wenn für alle $E_i \in \mathcal{S}_i, i \in \llbracket n \rrbracket$ gilt $\mu_{\llbracket n \rrbracket}(\times_{i \in \llbracket n \rrbracket} E_i) = \mu_{\llbracket n \rrbracket}(\bigcap_{i \in \llbracket n \rrbracket} \Pi_{\mathcal{S}_i}^{-1}(E_i)) = \prod_{i \in \llbracket n \rrbracket} \mu_i(E_i)$. In dem Fall schreiben wir $\bigotimes_{i \in \llbracket n \rrbracket} \mu_i := \mu_{\llbracket n \rrbracket}$. Sind alle $\mu_i = \mu$ gleich, so schreiben wir $\mu^{\otimes n} := \mu_{\llbracket n \rrbracket}$. \square

B05.05 **Eigenschaft.**

- (i) (**Eindeutigkeit eines endlichen Produktmaßes**) Jeder Erzeuger \mathcal{E}_i von $\mathcal{S}_i, i \in \llbracket n \rrbracket$, sei \cap -stabil und enthalte eine Folge $(E_{ik})_{k \in \mathbb{N}}$ von Mengen mit $E_{ik} \uparrow \mathcal{S}_i$ und $\mu_i(E_{ik}) < \infty$ für jedes $k \in \mathbb{N}$. Dann gibt es höchstens ein Maß $\mu_{\llbracket n \rrbracket} \in \mathcal{M}(\mathcal{S}_{\llbracket n \rrbracket})$ auf $(\mathcal{S}_{\llbracket n \rrbracket}, \mathcal{S}_{\llbracket n \rrbracket})$ mit $\mu_{\llbracket n \rrbracket}(\times_{i \in \llbracket n \rrbracket} E_i) = \prod_{i \in \llbracket n \rrbracket} \mu_i(E_i)$ für alle $E_i \in \mathcal{E}_i, i \in \llbracket n \rrbracket$.
- (ii) (**Existenz eines Produktmaßes**) Sei jedes $\mu_i \in \mathcal{M}_\sigma(\mathcal{S}_i)$ ein σ -endliches Maß auf $(\mathcal{S}_i, \mathcal{S}_i), i \in \llbracket n \rrbracket$. Dann existiert genau ein Produktmaß $\mu_{\llbracket n \rrbracket}$ auf $(\mathcal{S}_{\llbracket n \rrbracket}, \mathcal{S}_{\llbracket n \rrbracket})$. Dabei ist $\mu_{\llbracket n \rrbracket} \in \mathcal{M}_\sigma(\mathcal{S}_{\llbracket n \rrbracket})$ auch σ -endlich. \square

B05.06 **Schreibweise.** Bei gegebener Abbildung $h : \mathcal{S}_1 \times \mathcal{S}_2 \rightarrow \mathcal{S}_3$ bezeichnen wir für jedes $s_1 \in \mathcal{S}_1$ und $s_2 \in \mathcal{S}_2$ mit $h_{s_1} : \mathcal{S}_2 \rightarrow \mathcal{S}_3$ bzw. $h^{s_2} : \mathcal{S}_1 \rightarrow \mathcal{S}_3$ die Abbildung $s_2 \mapsto h_{s_1}(s_2) := h(s_1, s_2)$ bzw. $s_1 \mapsto h^{s_2}(s_1) := h(s_1, s_2)$. \square

B05.07 **Definition.** Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\mathcal{S}, \mathcal{S})$ ein Messraum und $N \in \mathcal{A}$ eine μ -Nullmenge. Eine Abbildung $h : N^c := \Omega \setminus N \rightarrow \mathcal{S}$ heißt **μ -fast überall definiert** sowie **\mathcal{A} - \mathcal{S} -messbar**, falls $h^{-1}(\mathcal{S}) \subseteq \mathcal{A}$ gilt. \square

B05.08 **Bemerkung.** Sind $h, g \in \overline{\mathcal{A}}$ μ -fast-überall endlich, so ist die Funktion $g - h$ μ -fast überall definiert und \mathcal{A} - $\overline{\mathcal{B}}$ -messbar. Insbesondere gilt dies, wenn g und h μ -integrierbar sind. Ist f nun $\overline{\mathbb{R}}$ -wertig, μ -fast-überall definiert mit μ -Nullmenge N und \mathcal{A} - $\overline{\mathcal{B}}$ -messbar, so können wir $\tilde{f}(\omega) := 0$ für $\omega \in N$ und andernfalls $\tilde{f}(\omega) := f(\omega)$ definieren. Dann ist $\tilde{f} \in \overline{\mathcal{A}}$ numerisch. Ist \tilde{f} weiterhin μ -integrierbar, so definieren wir für f das Integral $\mu(f) = \int f d\mu := \int \tilde{f} d\mu$. \square

B05.09 **Eigenschaft.**

(i) (**Satz von Tonelli**) Seien $(\mathcal{S}_i, \mathcal{S}_i, \mu_i), i \in \llbracket 2 \rrbracket$, σ -endliche Maßräume und sei $h \in \overline{\mathcal{S}_1 \otimes \mathcal{S}_2}^+$ positiv numerisch. Dann ist $\mu_1(h^\bullet) : \mathcal{S}_2 \rightarrow \overline{\mathbb{R}}^+$ mit $s_2 \mapsto \mu_1(h^{s_2})$ aus $\overline{\mathcal{S}_2}^+$ und $\mu_2(h_\bullet) : \mathcal{S}_1 \rightarrow \overline{\mathbb{R}}^+$ mit $s_1 \mapsto \mu_2(h_{s_1})$ aus $\overline{\mathcal{S}_1}^+$. Es gilt

$$\begin{aligned} \mu_1 \otimes \mu_2(h) &= \mu_2(\mu_1(h^\bullet)) = \int \mu_1(h^{s_2}) \mu_2(ds_2) = \int \int h(s_1, s_2) \mu_1(ds_1) \mu_2(ds_2) \\ &= \int \mu_2(h_{s_1}) \mu_1(ds_1) = \mu_1(\mu_2(h_\bullet)) \end{aligned}$$

(ii) (**Satz von Fubini**) Seien $(\mathcal{S}_i, \mathcal{S}_i, \mu_i), i \in \llbracket 2 \rrbracket$, σ -endliche Maßräume und $h \in \mathcal{L}_1(\mu_1 \otimes \mu_2)$. Dann ist $\mu_2(h_\bullet) : s_1 \mapsto \mu_2(h_{s_1})$ μ_1 -fast überall definiert und \mathcal{S}_1 - $\overline{\mathcal{B}}$ -messbar und $\mu_1(h^\bullet) : s_2 \mapsto \mu_1(h^{s_2})$ μ_2 -fast überall definiert und \mathcal{S}_2 - $\overline{\mathcal{B}}$ -messbar. Es gilt

$$\mu_2(\mu_1(h^\bullet)) = \int \mu_1(h^{s_2}) \mu_2(ds_2) = \mu_1 \otimes \mu_2(h) = \int \mu_2(h_{s_1}) \mu_1(ds_1) = \mu_1(\mu_2(h_\bullet)).$$

(iii) Für jedes $i \in \llbracket n \rrbracket$ sei $(\mathcal{S}_i, \mathcal{S}_i, \mu_i)$ ein σ -endlicher Maßraum, $\mathbb{f}_i \in \mathcal{S}_i^+$ und $\nu_i := \mathbb{f}_i \mu_i$. Dann ist das Produktmaß $\nu_{\llbracket n \rrbracket} = \bigotimes_{i \in \llbracket n \rrbracket} \nu_i$ definiert, und absolut stetig bezüglich des Produktmaßes $\mu_{\llbracket n \rrbracket} = \bigotimes_{i \in \llbracket n \rrbracket} \mu_i$ mit Produktdichte $\prod_{i \in \llbracket n \rrbracket} \mathbb{f}_i$, d.h. es gilt $\nu_{\llbracket n \rrbracket} = (\prod_{i \in \llbracket n \rrbracket} \mathbb{f}_i) \mu_{\llbracket n \rrbracket}$. \square

B05.10 **Erinnerung.** Seien nun $\nu = \mathbb{P}_0$ und $\mu = \mathbb{P}$ Wahrscheinlichkeitsmaße auf $(\mathcal{S}, \mathcal{S})$, wobei nicht notwendig $\mathbb{P}_0 \ll \mathbb{P}$ gilt. Dann heißt jede positive, numerische Funktion $L \in \mathcal{S}^+$ mit $\mathbb{P}_0 = L \mathbb{P} + \mathbb{1}_{\{L=\infty\}} \mathbb{P}_0$ und $\mathbb{P}(L \in \mathbb{R}^+) = 1$ Dichtequotient von \mathbb{P}_0 bezüglich \mathbb{P} (vgl. Definition B04.07). Bezeichnet $\mu \in \mathcal{M}_\sigma(\mathcal{S})$ ein σ -endliches Maß mit $\mathbb{P}_i \ll \mu, i \in \llbracket 2 \rrbracket$, (z.Bsp. das endliche Maß $\mu = \mathbb{P}_0 + \mathbb{P}$) und bezeichnen $\mathbb{f}_i \in \mathcal{S}^+$ μ -Dichten von $\mathbb{P}_i, i \in \llbracket 2 \rrbracket$, so ist $L^* := \frac{\mathbb{f}_0}{\mathbb{f}_1} \mathbb{1}_{\{\mathbb{f}_1 \in \mathbb{R}_0^+\}} + \infty \mathbb{1}_{\{\mathbb{f}_1=0, \mathbb{f}_0 \in \mathbb{R}_0^+\}} \in \overline{\mathcal{S}}^+$ eine spezielle Festlegung des Dichtequotienten. In dem speziellen Fall $\mathbb{P}_0 \ll \mathbb{P}$ stimmt der Dichtequotient von \mathbb{P}_0 bezüglich \mathbb{P} mit der \mathbb{P} -Dichte von \mathbb{P}_0 überein und ist \mathbb{P} -bestimmt. \square

B05.11 **Eigenschaft.** Für jedes $i \in \llbracket n \rrbracket$ seien $\mathbb{P}_{0,i}, \mathbb{P}_{1,i} \in \mathcal{W}(\mathcal{S}_i)$ Wahrscheinlichkeitsmaße auf $(\mathcal{S}_i, \mathcal{S}_i)$ mit Dichtequotient L_i von $\mathbb{P}_{0,i}$ bezüglich $\mathbb{P}_{1,i}$. Dann ist das Produkt $L(x) := \prod_{i \in \llbracket n \rrbracket} L_i(x_i)$ für $x = (x_i)_{i \in \llbracket n \rrbracket}$ eine Festlegung des Dichtequotienten von $\mathbb{P}_0 := \bigotimes_{i \in \llbracket n \rrbracket} \mathbb{P}_{0,i}$ bezüglich $\mathbb{P}_1 := \bigotimes_{i \in \llbracket n \rrbracket} \mathbb{P}_{1,i}$. \square

B06 Übergangskerne

B06.01 **Definition.** Seien $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ messbare Räume. Eine Abbildung $\kappa : \Omega_1 \times \mathcal{A}_2 \rightarrow \overline{\mathbb{R}}^+$ heißt (σ) -endlicher **Übergangskern** von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, falls sie die folgenden zwei Bedingungen erfüllt:

- (Ük1) für alle $\omega_1 \in \Omega_1$ ist $\kappa_{\omega_1} : \mathcal{A}_2 \rightarrow \overline{\mathbb{R}}^+$ mit $A_2 \mapsto \kappa_{\omega_1}(A_2) := \kappa(\omega_1, A_2)$ ein (σ) -endliches Maß auf $(\Omega_2, \mathcal{A}_2)$, kurz $\kappa_{\omega_1} \in \mathcal{M}_\sigma(\mathcal{A}_2)$ bzw. $\kappa_{\omega_1} \in \mathcal{M}_\sigma(\mathcal{A}_2)$;
- (Ük2) für alle $A_2 \in \mathcal{A}_2$ ist $\kappa^{A_2} : \Omega_1 \rightarrow \overline{\mathbb{R}}^+$ mit $\omega_1 \mapsto \kappa^{A_2}(\omega_1) := \kappa(\omega_1, A_2)$ eine positive, numerische \mathcal{A}_1 -messbare Abbildung, kurz $\kappa^{A_2} \in \overline{\mathcal{A}_1}^+$.

Ist für jedes $\omega_1 \in \Omega_1$ das Maß in (Ük1) ein Wahrscheinlichkeitsmaß, $\kappa_{\omega_1} \in \mathcal{W}(\mathcal{A}_2)$, so heißt κ *Markovkern*. □

B06.02 **Bemerkung.** Es genügt die Bedingung (Ük2) nur für Mengen aus einem \cap -stabilen Erzeuger \mathcal{E} von \mathcal{A}_2 , der Ω_2 oder eine Folge $(E_n)_{n \in \mathbb{N}}$ von Mengen mit $E_n \uparrow \Omega_2$ enthält, zu fordern. □

B06.03 **Eigenschaft.** Sei κ ein endlicher Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, und sei $h \in \overline{\mathcal{A}_1 \otimes \mathcal{A}_2}^+$ positiv numerisch. Dann ist die Funktion $\kappa_*(h_\bullet) : \Omega_1 \rightarrow \overline{\mathbb{R}}^+$ mit $\omega_1 \mapsto \kappa_{\omega_1}(h_{\omega_1}) = \int h_{\omega_1} d\kappa_{\omega_1}$ wohldefiniert und aus $\overline{\mathcal{A}_1}^+$. □

B06.04 **Schreibweise.** Für $\mathbb{1}_A \in (\mathcal{A}_1 \otimes \mathcal{A}_2)^+$, also $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, ist nach **Eigenschaft** B06.03 die Funktion $\kappa_*(A_\bullet) = \kappa_*((\mathbb{1}_A)_\bullet) : \Omega_1 \rightarrow \overline{\mathbb{R}}^+$ mit $\omega_1 \mapsto \kappa_{\omega_1}(A_{\omega_1}) = \kappa_{\omega_1}((\mathbb{1}_A)_{\omega_1})$ wohldefiniert und aus $\overline{\mathcal{A}_1}^+$. □

B06.05 **Eigenschaft.** Sei $(\Omega_1, \mathcal{A}_1, \mu_1)$ ein endlicher Maßraum, $(\Omega_2, \mathcal{A}_2)$ ein Messraum und κ ein endlicher Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$. Dann existiert ein eindeutig bestimmtes, σ -endliches Maß $\mu \odot \kappa \in \mathcal{M}_\sigma(\mathcal{A}_1 \otimes \mathcal{A}_2)$ auf dem Produktraum $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ mit $\mu \odot \kappa(A) = \mu(\kappa_*(A_\bullet))$ für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, wobei für alle $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$ gilt

$$\mu \odot \kappa(A_1 \times A_2) = \mu(\mathbb{1}_{A_1} \kappa^{A_2}) = \int_{A_1} \kappa^{A_2} d\mu = \int_{A_1} \kappa(\omega_1, A_2) \mu(d\omega_1).$$

Ist κ ein Markovkern und μ ein Wahrscheinlichkeitsmaß, so ist $\mu \odot \kappa$ ein Wahrscheinlichkeitsmaß. □

B06.06 **Satz von Tonelli/Fubini für Übergangskerne.** Sei $(\Omega_1, \mathcal{A}_1, \mu)$ ein endlicher Maßraum, $(\Omega_2, \mathcal{A}_2)$ ein Messraum und κ ein endlicher Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$. Ist $h \in \overline{\mathcal{A}_1 \otimes \mathcal{A}_2}^+$ oder $h \in \mathcal{L}_1(\mu \odot \kappa)$ dann gilt

$$\begin{aligned} \mu \odot \kappa(h) &= \mu(\kappa_*(h_\bullet)) = \int \kappa_{\omega_1}(h_{\omega_1}) \mu(d\omega_1) = \int \left(\int h_{\omega_1} d\kappa_{\omega_1} \right) \mu(d\omega_1) \\ &= \int \int h(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2) \mu(d\omega_1). \end{aligned}$$

□

B06.07 **Schreibweise.** Betrachte einen Wahrscheinlichkeitsraum $(\Omega_1, \mathcal{A}_1, \mathbb{P})$, einen Messraum $(\Omega_2, \mathcal{A}_2)$, einen Markovkern κ von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$ und mit **Eigenschaft** B06.05 das eindeutig bestimmte Wahrscheinlichkeitsmaß $\mathbb{P} \odot \kappa$ auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. Dann bezeichnen wir mit

$$\kappa \mathbb{P}(A_2) := \mathbb{P}(\kappa^{A_2}) = \int \kappa^{A_2} d\mathbb{P} = \int \kappa(\omega_1, A_2) \mathbb{P}(d\omega_1), \quad \text{für alle } A_2 \in \mathcal{A}_2$$

die durch $\mathbb{P} \odot \kappa$ auf $(\Omega_2, \mathcal{A}_2)$ induzierte Randverteilung $\kappa \mathbb{P} \in \mathcal{W}(\mathcal{A}_2)$. □

C Bedingte Erwartung

In diesem Kapitel werden Begriffe, Notationen und Aussagen der Vorlesung *Wahrscheinlichkeitstheorie 1 (WT1)* wiederholt. Eine detaillierte Darstellung findet sich zum Beispiel in Klenke [2020] und Witting [1985].

C01 Diskret- oder stetig-verteilte Zufallsvariablen

C01.01 **Erinnerung.** Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Für $B \in \mathcal{A}$ mit $\mathbb{P}(B) \in \mathbb{R}_0^+$ heißt $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ die *bedingte Wahrscheinlichkeit* von A bei gegebenem B . \square

C01.02 **Eigenschaft.**

- (i) Sei (S, X) eine diskret-verteilte Zufallsvariable mit Werten in $(\mathcal{S} \times \mathcal{X}, 2^{\mathcal{S} \times \mathcal{X}})$ und gemeinsamer Zähl-dichte $\mathbb{p}^{(S,X)} : \mathcal{S} \times \mathcal{X} \rightarrow [0, 1]$ (bezüglich des Produktzählmaßes $\zeta_{\mathcal{S}} \otimes \zeta_{\mathcal{X}}$). Für jedes $s \in \mathcal{S}$ mit strikt-positiver Randzähl-dichte von S in s , also $\mathbb{p}^S(s) \in \mathbb{R}_0^+$, ist die Abbildung

$$\mathbb{p}^{X|S=s} : \mathcal{X} \rightarrow [0, 1] \text{ mit } x \mapsto \mathbb{p}^{X|S=s}(x) := \frac{\mathbb{p}^{(S,X)}(x, s)}{\mathbb{p}^S(s)} \quad (01.01)$$

eine Zähl-dichte (bezüglich des Zählmaßes $\zeta_{\mathcal{X}}$).

- (ii) Sei (S, X) ein stetig-verteilter Zufallsvektor mit gemeinsamer Dichte $f^{(S,X)} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^+$ (bezüglich des Lebesgue-Maßes λ^{m+n}). Für jedes $s \in \mathbb{R}^m$ mit strikt-positiver Randdichte von S in s , also $f^S(s) \in \mathbb{R}_0^+$, ist

$$f^{X|S=s} : \mathbb{R}^n \rightarrow \mathbb{R}^+ \text{ mit } x \mapsto f^{X|S=s}(x) := \frac{f^{(S,X)}(s, x)}{f^S(s)} \quad (01.02)$$

eine Dichte (bezüglich des Lebesgue-Maßes λ^n). \square

C01.03 **Bemerkung.** Da die Mengen $\{s \in \mathcal{S} : \mathbb{p}^S(s) = 0\}$ bzw. $\{s \in \mathbb{R}^m : f^S(s) = 0\}$ Nullmengen bzgl. der Verteilung \mathbb{P}^S von S sind, sind die Abbildungen $\mathbb{p}^{X|S=s}$ und $f^{X|S=s}$ tatsächlich \mathbb{P}^S -fast überall, also nur außerhalb einer \mathbb{P}^S -Nullmenge, definiert. Wir können auf diesen Nullmengen beliebige Dichten, zum Beispiel die Randdichten $\mathbb{p}^{X|S=s} = \mathbb{p}^X$ und $f^{X|S=s} = f^X$, wählen. \square

C01.04 **Definition.**

- (a) Sei (S, X) eine diskret-verteilte Zufallsvariable mit Werten in $(\mathcal{S} \times \mathcal{X}, 2^{\mathcal{S} \times \mathcal{X}})$. Für jedes $s \in \mathcal{S}$ mit strikt-positiver Randzähl-dichte von S in s heißt die Verteilung $\mathbb{P}^{X|S=s}$ mit Zähl-dichte $\mathbb{p}^{X|S=s}$ in (01.01) *bedingte Verteilung* (mit *bedingter Zähl-dichte*) von X gegeben $S = s$. Für $h \in \overline{2^{\mathcal{X}}}$ bzw. $h \in \mathcal{L}_1(\mathbb{P}^{X|S=s})$ wird $\mathbb{E}^{X|S=s}(h) := \mathbb{P}^{X|S=s}(h) = \sum_{x \in \mathcal{X}} h(x) \mathbb{p}^{X|S=s}(x) = \int h \mathbb{p}^{X|S=s} d\zeta_{\mathcal{X}}$ *bedingter Erwartungswert* von $h(X)$ bei gegebenem $S = s$ genannt.
- (b) Sei (S, X) ein stetig-verteilter Zufallsvektor mit Werten in \mathbb{R}^{n+m} . Für jedes $s \in \mathbb{R}^m$ mit strikt-positiver Randdichte von S in s heißt die Verteilung $\mathbb{P}^{X|S=s}$ mit Dichte $f^{X|S=s}$ in (01.02) *bedingte Verteilung* bzw. *bedingte Dichte* von X gegeben $S = s$. Für $h \in \overline{\mathcal{B}^n}$

bzw. $h \in \mathcal{L}_1(\mathbb{P}^{X|S=s})$ wird $\mathbb{E}^{X|S=s}(h) := \mathbb{P}^{X|S=s}(h) = \int h(x) \mathbb{f}^{X|S=s}(x) \lambda^n(dx) = \int h \mathbb{f}^{X|S=s} d\lambda^n$ *bedingter Erwartungswert* von $h(X)$ gegeben $S = s$ genannt. □

C01.05 **Bemerkung.** Betrachten wir den stetigen Fall (für den diskreten Fall folgen die Aussagen analog). Da für $h \in \mathcal{L}_1(\mathbb{P}^X)$ auch außerhalb einer \mathbb{P}^S -Nullmenge $\int |h| \mathbb{f}_s^{(S,X)} d\lambda^n < \infty$ gilt, ist für jedes s außerhalb einer \mathbb{P}^S -Nullmenge der bedingte Erwartungswert $\mathbb{E}^{X|S=s}(h) = \mathbb{P}^{X|S=s}(h)$ von $h(X)$ gegeben $S = s$, wie auch zuvor die bedingte Dichte, definiert, und auf der Nullmenge können wir diesen beliebig fortsetzen. Verschiedene Versionen (Festlegungen) unterscheiden sich damit nur auf einer Nullmenge. Wir sehen später, dass wir eine Festlegung derart wählen können, dass die Abbildung $s \mapsto \varphi(s) := \mathbb{E}^{X|S=s}(h) = \mathbb{P}^{X|S=s}(h)$ messbar ist. In diesem Fall gilt für alle $B \in \mathcal{B}^m$:

$$\begin{aligned} \mathbb{E}^S(\mathbb{1}_B \varphi) &= \mathbb{P}^S(\mathbb{1}_B \varphi) = \int \mathbb{1}_B \varphi \mathbb{f}^S d\lambda^m = \int_{\mathbb{R}^m} \mathbb{1}_B(s) \int_{\mathbb{R}^n} h(x) \mathbb{f}^{(S,X)}(s, x) \lambda^n(dx) \lambda^m(ds) \\ &= \int \mathbb{1}_B h \mathbb{f}^{(S,X)} d\lambda^{m+n} = \mathbb{P}^{(S,X)}(\mathbb{1}_B h) = \mathbb{E}^{(S,X)}(\mathbb{1}_B h) \end{aligned}$$

Die Messbarkeit und die letzte Gleichheit sind der Ausgangspunkt für den folgenden allgemeinen Zugang, der uns erlaubt, auch den diskreten sowie den stetigen Fall abzudecken. Die folgenden Resultate und Eigenschaften lassen sich dann auf die bisher vorgestellten bedingten Dichten, Verteilungen und Erwartungswerte übertragen. □

C01.06 **Beispiel.**

(a) Sei $X = |W_1 - W_2|$ und $S = W_1 + W_2$ der Absolutbetrag der Differenz bzw. die Summe der Augenzahlen von zwei unabhängigen fairen Würfeln (W_1, W_2). Dann ist (S, X) diskretverteilt mit Werten in $\llbracket 0, 5 \rrbracket \times \llbracket 2, 12 \rrbracket$ und bedingten Zähldichten von X gegeben $S = s$:

		s												
		2	3	4	5	6	7	8	9	10	11	12	$\mathbb{P}^X(x)$	
x	0	1	0	$\frac{1}{3}$	0	$\frac{1}{5}$	0	$\frac{1}{5}$	0	$\frac{1}{3}$	0	1	$\frac{6}{36}$	
	1	0	1	0	$\frac{1}{2}$	0	$\frac{1}{3}$	0	$\frac{1}{2}$	0	1	0	$\frac{10}{36}$	
	2	0	0	$\frac{2}{3}$	0	$\frac{2}{5}$	0	$\frac{2}{5}$	0	$\frac{2}{3}$	0	0	$\frac{8}{36}$	
	3	0	0	0	$\frac{1}{2}$	0	$\frac{1}{3}$	0	$\frac{1}{2}$	0	0	0	$\frac{6}{36}$	
	4	0	0	0	0	$\frac{2}{5}$	0	$\frac{2}{5}$	0	0	0	0	$\frac{4}{36}$	
	5	0	0	0	0	0	$\frac{1}{3}$	0	0	0	0	0	$\frac{2}{36}$	
$\mathbb{P}^S(s)$		$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1	

(b) Seien X und S Zufallsvektoren mit

$$Y := \begin{pmatrix} X \\ S \end{pmatrix} \sim N_{(\mu, \Sigma)} \text{ mit } \mu = \begin{pmatrix} \mu_X \\ \mu_S \end{pmatrix} \in \mathbb{R}^{n+m} \text{ und } \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XS} \\ \Sigma_{SX} & \Sigma_S \end{pmatrix} > 0,$$

also $X \sim N_{(\mu_X, \Sigma_X)}$, $S \sim N_{(\mu_S, \Sigma_S)}$ und $\text{Cov}(X, S) = \Sigma_{XS} = \Sigma_{SX}^t$. Dann ist die bedingte Verteilung von X gegeben $S = s$ eine $N_{(\mu_{X|S=s}, \Sigma_{X|S=s})}$ -Verteilung mit

$$\mu_{X|S=s} := \mu_X + \Sigma_{XS} \Sigma_S^{-1} (s - \mu_S) \in \mathbb{R}^m \text{ und } \Sigma_{X|S=s} := \Sigma_X - \Sigma_{XS} \Sigma_S^{-1} \Sigma_{SX} > 0.$$

Definitionsgemäß ist $\mu_{X|S=s}$ gerade der bedingte Erwartungswert von X gegeben $S = s$. Zum Beweis der Aussage benutzen wir, dass für $\Sigma > 0$ gilt $\Sigma_X > 0$, $\Sigma_S > 0$ und $E :=$

$\Sigma_X - \Sigma_{XS}\Sigma_S^{-1}\Sigma_{SX} > 0$ sowie mit $F := \Sigma_{XS}\Sigma_S^{-1}$ auch

$$\Sigma^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}F \\ -FE^{-1} & \Sigma_S^{-1} + F^tE^{-1}F \end{pmatrix} > 0.$$

Für die gemeinsame Dichte $f_{N(\mu, \Sigma)}$ von $Y = (X^t, S^t)^t$ und die Randdichte $f_{N(\mu_S, \Sigma_S)}$ von S gilt mit $y = (x^t, s^t)$ und Normierungskonstante c

$$\begin{aligned} \frac{f_{N(\mu, \Sigma)}(y)}{f_{N(\mu_S, \Sigma_S)}(s)} &= c \exp \left\{ -\frac{1}{2}(y - \mu)^t \Sigma^{-1}(y - \mu) + \frac{1}{2}(s - \mu_S)^t \Sigma^{-1}(s - \mu_S) \right\} \\ &= c \exp \left\{ -\frac{1}{2}(x - (\mu_X + F(s - \mu_S)))^t E^{-1}(x - (\mu_X + F(s - \mu_S))) \right\}. \end{aligned}$$

Da $\mu_{X|S=s} = \mu_X + F(s - \mu_S)$ und $\Sigma_{X|S=s} = E$ folgt die Behauptung. \square

C02 Positive numerische Zufallsvariablen

Seien im Folgenden stets $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, \mathbb{E} die Erwartung bzgl. \mathbb{P} und $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra von \mathcal{A} .

C02.01 **Erinnerung.** Wir schreiben kurz $X \in \overline{\mathcal{A}}^+$, wenn X eine positive numerische Zufallsvariable auf (Ω, \mathcal{A}) , also $X : \Omega \rightarrow \overline{\mathbb{R}}^+$ eine \mathcal{A} - $\overline{\mathcal{B}}^+$ -messbare Abbildung, ist. Insbesondere, ist $\overline{\mathcal{F}}^+ \subseteq \overline{\mathcal{A}}^+$ und für $Y \in \overline{\mathcal{F}}^+$ somit $\mathbb{E}(Y)$ definiert. \square

C02.02 **Eigenschaft.** Zu jedem $X \in \overline{\mathcal{A}}^+$ existiert eine Lösung $Y \in \overline{\mathcal{F}}^+$ der Radon-Nikodym-Gleichung $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ für jedes $F \in \mathcal{F}$, wobei Y bis auf Gleichheit \mathbb{P} -f.ü. eindeutig ist. \square

C02.03 **Definition.** Für $X \in \overline{\mathcal{A}}^+$ heißt jede Abbildung $Y : \Omega \rightarrow \overline{\mathbb{R}}^+$ *Festlegung (Version) des bedingten Erwartungswertes* von X gegeben \mathcal{F} , symbolisch $\mathbb{E}(X|\mathcal{F}) := Y$, wenn sie die folgenden zwei Bedingungen erfüllt:

(bE1) Y ist \mathcal{F} - $\overline{\mathcal{B}}^+$ -messbar, also $Y \in \overline{\mathcal{F}}^+$ und

(bE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ für jedes $F \in \mathcal{F}$. (Radon-Nikodym-Gleichung)

Jede Abbildung $\mathbb{E}(\bullet|\mathcal{F}) : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathcal{F}}^+$ mit $X \mapsto \mathbb{E}(X|\mathcal{F})$ heißt *Festlegung der bedingten Erwartung* bzgl. \mathbb{P} gegeben \mathcal{F} . Die von $\mathbb{E}(\bullet|\mathcal{F})$ implizierte Abbildung $\mathbb{P}(\bullet|\mathcal{F}) : \mathcal{A} \rightarrow \overline{\mathcal{F}}^+$ mit $A \mapsto \mathbb{P}(A|\mathcal{F}) := \mathbb{E}(\mathbb{1}_A|\mathcal{F})$ wird *Festlegung der bedingten Verteilung* von \mathbb{P} gegeben \mathcal{F} genannt. Mit (bE2) ist jede Festlegung $\mathbb{P}(A|\mathcal{F})$ Lösung der Radon-Nikodym-Gleichung, d.h. $\mathbb{E}(\mathbb{1}_F \mathbb{P}(A|\mathcal{F})) = \int_F \mathbb{P}(A|\mathcal{F}) d\mathbb{P} = \mathbb{P}(F \cap A)$ für alle $F \in \mathcal{F}$. \square

C02.04 **Bemerkung.** Nach **Eigenschaft** C02.02 unterscheiden sich für jedes $X \in \overline{\mathcal{A}}^+$ Festlegungen des bedingten Erwartungswertes von X gegeben \mathcal{F} nur auf einer \mathbb{P} -Nullmenge. Diese Eigenschaft überträgt sich im Allgemeinen nicht auf Festlegungen der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} , da wir zu jedem $X \in \overline{\mathcal{A}}^+$ eine Ausnahmemenge erhalten, und deren Vereinigung im Allgemeinen keine Nullmenge mehr ist. Betrachten wir eine Festlegung $\mathbb{P}(\bullet|\mathcal{F})$ der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} , so ist eine wünschenswerte Eigenschaft, dass für jedes $\omega \in \Omega$ die entsprechende Abbildung $\mathbb{P}(\bullet|\mathcal{F})(\omega) : \mathcal{A} \rightarrow \overline{\mathbb{R}}^+$ mit $A \mapsto \mathbb{P}(A|\mathcal{F})(\omega)$ ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) ist. \square

C02.05 **Erinnerung** (vgl. **Definition** B06.01). Seien $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ messbare Räume. Eine Abbildung $\kappa : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ heißt *Markovkern* von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, falls sie die folgenden zwei Bedingungen erfüllt:

(Mk1) für alle $\omega_1 \in \Omega_1$ ist $\kappa_{\omega_1} : \mathcal{A}_2 \rightarrow [0, 1]$ mit $A_2 \mapsto \kappa_{\omega_1}(A_2) := \kappa(\omega_1, A_2)$ ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{A}_2)$, kurz $\kappa_{\omega_1} \in \mathcal{W}(\mathcal{A}_2)$;

(Mk2) für alle $A_2 \in \mathcal{A}_2$ ist $\kappa^{A_2} : \Omega_1 \rightarrow [0, 1]$ mit $\omega_1 \mapsto \kappa^{A_2}(\omega_1) := \kappa(\omega_1, A_2)$ eine \mathcal{A}_1 - $\mathcal{B}_{[0,1]}$ -messbare Abbildung, kurz $\kappa^{A_2} \in \mathcal{A}_1^+$. \square

C02.06 **Schreibweise** (vgl. B06.07). Betrachte einen Wahrscheinlichkeitsraum $(\Omega_1, \mathcal{A}_1, \mathbb{P})$, einen messbaren Raum $(\Omega_2, \mathcal{A}_2)$ und einen Markovkern κ von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$. Dann existiert nach **Eigenschaft** B06.05 ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\mathbb{P} \odot \kappa$ auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ mit

$$\mathbb{P} \odot \kappa(A_1 \times A_2) = \mathbb{P}(\mathbb{1}_{A_1} \kappa^{A_2}) = \int_{A_1} \kappa(\omega_1, A_2) \mathbb{P}(d\omega_1), \quad \text{für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Ist $h \in \overline{\mathcal{A}_1 \otimes \mathcal{A}_2}^+$ oder $h \in \mathcal{L}(\mathbb{P} \odot \kappa)$ dann gilt

$$\mathbb{P} \odot \kappa(h) = \mathbb{P}(\kappa_\bullet(h_\bullet)) = \int_{\Omega_1} \int_{\Omega_2} h(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2) \mathbb{P}(d\omega_1)$$

(vgl. **Satz** B06.06). Weiterhin bezeichnen wir mit $\kappa \mathbb{P}$ die durch $\kappa \odot \mathbb{P}$ auf $(\Omega_2, \mathcal{A}_2)$ induzierte Randverteilung, also $\kappa \mathbb{P}(A_2) = \mathbb{P}(\kappa^{A_2}) = \int_{\Omega_1} \kappa(\omega_1, A_2) \mathbb{P}(d\omega_1)$ für alle $A_2 \in \mathcal{A}_2$. \square

C02.07 **Definition.**

- (a) $\mathbb{P}(\bullet | \mathcal{F})$ heißt *reguläre Festlegung* der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} , wenn die Abbildung $(\omega, A) \mapsto \mathbb{P}(A | \mathcal{F})(\omega)$ die Eigenschaften (Mk1) und (Mk2) erfüllt, also $\mathbb{P}(\bullet | \mathcal{F})$ ein Markovkern von (Ω, \mathcal{F}) nach (Ω, \mathcal{A}) ist.
- (b) $\mathbb{E}(\bullet | \mathcal{F})$ heißt *reguläre Festlegung* der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} , wenn die implizierte Festlegung $\mathbb{P}(\bullet | \mathcal{F})$ der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} regulär ist, und für jedes $\omega \in \Omega$ die Abbildung $X \mapsto \mathbb{E}(X | \mathcal{F})(\omega)$ die Erwartung bzgl. $\mathbb{P}(\bullet | \mathcal{F})(\omega)$ ist. \square

C02.08 **Schreibweise.** Ist $\mathbb{E}(\bullet | \mathcal{F})$ eine reguläre Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} , und $\mathbb{P}(\bullet | \mathcal{F})$ die implizierte Festlegung der bedingten Verteilung bzgl. \mathbb{P} gegeben \mathcal{F} , so verwenden wir für $X \in \overline{\mathcal{A}}^+$ wie bisher auch beide Schreibweisen $\mathbb{E}(X | \mathcal{F}) = \mathbb{P}(X | \mathcal{F})$. \square

C02.09 **Eigenschaft.**

- (i) Zu jeder regulären Festlegung der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} existiert eine sie implizierende reguläre Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} .
- (ii) Zu jedem Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}^k, \mathcal{B}^k)$ und Teil- σ -Algebra $\mathcal{F} \subseteq \mathcal{B}^k$ von \mathcal{B}^k existiert eine reguläre Festlegung der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} . \square

C02.10 **Bemerkung.** Bevor wir Eigenschaften der bedingten Erwartung festhalten, werden wir in einfachen Situationen konstruktiv Versionen der bedingten Erwartung bzw. Verteilung bestimmen. Dabei zeigt sich, dass auch eine reguläre Festlegung der bedingten Verteilung nicht eindeutig ist, aber in vielen für uns interessanten Situation bis auf \mathbb{P} -f.ü. Gleichheit eindeutig bestimmt ist. \square

C02|01 Einfache Bedingungen

Wir betrachten zunächst eine Teil- σ -Algebra \mathcal{F} von \mathcal{A} , die nur von einer einzelnen messbaren Menge erzeugt ist, das heißt, $\mathcal{F} = \{\emptyset, B, B^c, \Omega\} = \sigma(\{B\})$ mit $B \in \mathcal{A}$.

C02.11 **Erinnerung.** Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Zu \mathbb{P} existiert nach Definition A04.01 eine eindeutig bestimmte Erwartung $\mathbb{E} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$, die (E1) linear und (E2) monoton konvergent ist mit (E3) $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$ für alle $A \in \mathcal{A}$. \square

C02.12 **Bedingte Erwartung.** Sei $B \in \mathcal{A}$ mit $\mathbb{P}(B) \in \mathbb{R}_0^+$. Dann ist

$$\mathbb{P}(\cdot | B) : \mathcal{A} \rightarrow [0, 1] \text{ mit } A \mapsto \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) . Entsprechend existiert zu $\mathbb{P}(\cdot | B)$ eine eindeutig bestimmte Erwartung, die wir mit $\mathbb{E}(\cdot | B)$ bezeichnen. Mit (E3) erfüllt diese

$$\forall A \in \mathcal{A} : \mathbb{E}(\mathbb{1}_A | B) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}(\mathbb{1}_A \mathbb{1}_B)}{\mathbb{P}(B)} = \mathbb{P}(\mathbb{1}_A | B).$$

Da für jedes $X \in \overline{\mathcal{A}}^+$ weiterhin gilt $X\mathbb{1}_B \in \overline{\mathcal{A}}^+$ erfüllt die Erwartung $\mathbb{E}(\cdot | B)$ auch

$$\forall X \in \overline{\mathcal{A}}^+ : \mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)} = \mathbb{P}(X|B).$$

Falls für $B^c = \Omega \setminus B \in \mathcal{A}$ auch $\mathbb{P}(B^c) \in \mathbb{R}_0^+$ gilt, so können wir analog das Wahrscheinlichkeitsmaß $\mathbb{P}(\cdot | B^c)$ auf (Ω, \mathcal{A}) und die entsprechende Erwartung $\mathbb{E}(\cdot | B^c)$ bzgl. $\mathbb{P}(\cdot | B^c)$ betrachten. Für jedes $X \in \overline{\mathcal{A}}^+$ ist dann

$$\mathbb{E}(X|\mathcal{F}) : \Omega \rightarrow \overline{\mathbb{R}}^+ \text{ mit } \omega \mapsto \mathbb{E}(X|\mathcal{F})(\omega) := \mathbb{E}(X|B)\mathbb{1}_B(\omega) + \mathbb{E}(X|B^c)\mathbb{1}_{B^c}(\omega)$$

messbar bzgl. der Teil- σ -Algebra $\mathcal{F} := \sigma(\{B\}) = \{\emptyset, B, B^c, \Omega\} \subseteq \mathcal{A}$, also eine einfache numerische Zufallsvariable auf (Ω, \mathcal{F}) , kurz $\mathbb{E}(X|\mathcal{F}) \in \overline{\mathcal{F}}^+$. Für jedes $\omega \in \Omega$ ist entweder $\omega \in B$ und $\mathbb{E}(X|\mathcal{F})(\omega) = \mathbb{E}(X|B)$ oder $\omega \in B^c$ sowie $\mathbb{E}(X|\mathcal{F})(\omega) = \mathbb{E}(X|B^c)$, so dass

$$\mathbb{E}(\cdot | \mathcal{F}) : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathcal{F}}^+ \text{ mit } X \mapsto \mathbb{E}(X|\mathcal{F}) := \mathbb{E}(X|B)\mathbb{1}_B + \mathbb{E}(X|B^c)\mathbb{1}_{B^c}. \quad (02.01)$$

Wir halten fest, dass gegeben $\omega \in B$, gilt $\mathbb{E}(\cdot | B) = \mathbb{E}(\cdot | \mathcal{F})(\omega)$. Für jedes $X \in \overline{\mathcal{A}}^+$ ist $\mathbb{E}(X|\mathcal{F})$ insbesondere Lösung der Radon-Nikodym-Gleichung (einfaches nachrechnen!)

$$\forall F \in \mathcal{F} : \mathbb{E}(\mathbb{1}_F \mathbb{E}(X|\mathcal{F})) = \mathbb{E}(\mathbb{1}_F X). \quad (02.02)$$

Damit erfüllt die Abbildung $\mathbb{E}(\cdot | \mathcal{F})$ in (02.01) die Bedingungen (bE1) sowie (bE2), und sie ist somit eine Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} . \square

C02.13 **Reguläre Festlegung.** Wir halten weiterhin fest, dass die von der Festlegung $\mathbb{E}(\cdot | \mathcal{F})$ in (02.01) implizierte Festlegung der bedingten Verteilung

$$\mathbb{P}(\cdot | \mathcal{F}) : \mathcal{A} \rightarrow \mathcal{F}^+ \text{ mit } A \mapsto \mathbb{P}(A|\mathcal{F}) := \mathbb{E}(\mathbb{1}_A|\mathcal{F})$$

ein **Markovkern** von (Ω, \mathcal{F}) nach (Ω, \mathcal{A}) ist. In der Tat erfüllt die Abbildung $\Omega \times \mathcal{A} \rightarrow \mathbb{R}^+$ mit

$$(\omega, A) \mapsto \mathbb{P}(A|\mathcal{F})(\omega) = \mathbb{P}(A|B)\mathbb{1}_B(\omega) + \mathbb{P}(A|B^c)\mathbb{1}_{B^c}(\omega),$$

die Bedingungen (Mk1) und (Mk2). Betrachte (Mk1). Für $\omega \in \Omega$ ist entweder $\omega \in B$, und somit $\mathbb{P}(\cdot | \mathcal{F})(\omega) = \mathbb{P}(\cdot | B)$, oder $\omega \in B^c$ und $\mathbb{P}(\cdot | \mathcal{F})(\omega) = \mathbb{P}(\cdot | B^c)$. In beiden Fällen ist $\mathbb{P}(\cdot | \mathcal{F})(\omega) \in \mathcal{W}(\mathcal{A})$ ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) . Wir halten fest, dass gegeben $\omega \in B$, gilt $\mathbb{P}(\cdot | B) = \mathbb{P}(\cdot | \mathcal{F})(\omega)$. Andererseits, gilt (Mk2). Da für jedes $A \in \mathcal{A}$ nach Konstruktion die Zufallsvariable $\mathbb{P}(A|\mathcal{F}) = \mathbb{E}(\mathbb{1}_A|\mathcal{F})$ die Bedingung (bE1) erfüllt, die Abbildung $\mathbb{P}(A|\mathcal{F}) : \Omega \rightarrow [0, 1]$ mit $\omega \mapsto \mathbb{P}(A|\mathcal{F})(\omega)$ also \mathcal{F} -messbar ist. Folglich ist die **Festlegung** $\mathbb{P}(\cdot | \mathcal{F})$ der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} **regulär**. Andererseits ist per Konstruktion für jedes $\omega \in \Omega$ die Abbildung $\mathbb{E}(\cdot | \mathcal{F})(\omega)$ die Erwartung bzgl. $\mathbb{P}(\cdot | \mathcal{F})(\omega)$, so dass auch die **Festlegung** $\mathbb{E}(\cdot | \mathcal{F})$ in (02.01) **regulär** ist. \square

C02.14 **Bedingte Verteilung geben eine Nullmenge B .** Für $B \in \mathcal{A}$ mit $\mathbb{P}(B) = 0$ ist $\mathbb{P}(\cdot|B)$ und damit auch $\mathbb{E}(\cdot|\mathcal{F})$ sowie $\mathbb{P}(\cdot|\mathcal{F})$ bisher nicht definiert. Wir vereinbaren im Folgenden $\mathbb{P}(\cdot|B) := \mathbb{P}$ und somit $\mathbb{E}(\cdot|B) = \mathbb{E}$. Die Abbildung (02.01) wird dann zu

$$\mathbb{E}(\cdot|\mathcal{F}) : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathcal{F}}^+ \text{ mit } X \mapsto \mathbb{E}(X|\mathcal{F}) := \mathbb{E}(X)\mathbb{1}_B + \mathbb{E}(X|B^c)\mathbb{1}_{B^c}. \tag{02.03}$$

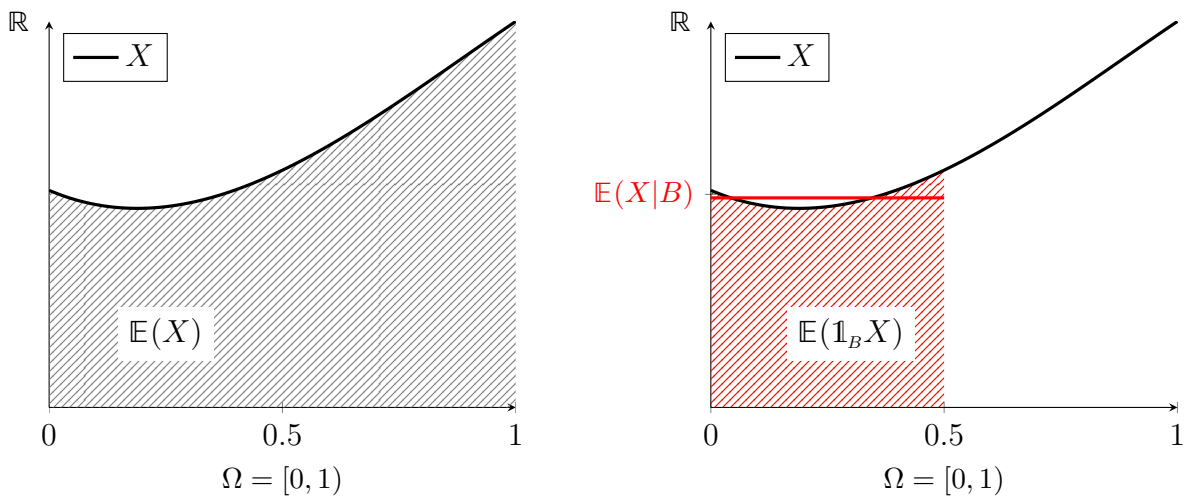
$\mathbb{E}(\cdot|\mathcal{F})$ erfüllt für alle $X \in \overline{\mathcal{A}}^+$ weiterhin die Bedingungen (bE1)-(bE2), d.h. $\mathbb{E}(X|\mathcal{F}) \in \overline{\mathcal{F}}^+$ sowie $\mathbb{E}(\mathbb{1}_F \mathbb{E}(X|\mathcal{F})) = \mathbb{E}(\mathbb{1}_F X)$ für alle $F \in \mathcal{F}$. Damit ist $\mathbb{E}(\cdot|\mathcal{F})$ in (02.03) eine Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} . Außerdem sind (Mk1) und (Mk2) immer noch von

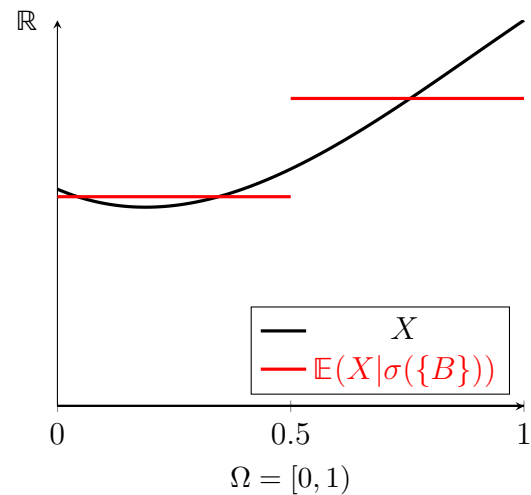
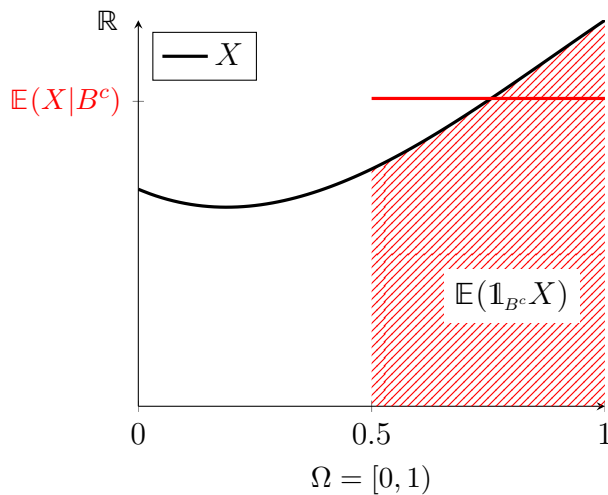
$$\mathbb{P}(\cdot|\mathcal{F}) : \mathcal{A} \rightarrow \overline{\mathcal{F}}^+ \text{ mit } A \mapsto \mathbb{P}(\cdot|\mathcal{F}) := \mathbb{P}(A)\mathbb{1}_B + \mathbb{P}(A|B^c)\mathbb{1}_{B^c} \tag{02.04}$$

erfüllt. Damit ist $\mathbb{P}(\cdot|\mathcal{F})$ weiterhin eine **reguläre Festlegung** der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} . Per Konstruktion ist für jedes $\omega \in \Omega$ die Abbildung $\mathbb{E}(\cdot|\mathcal{F})(\omega)$ die Erwartung bzgl. $\mathbb{P}(\cdot|\mathcal{F})(\omega)$, so dass auch die **Festlegung** $\mathbb{E}(\cdot|\mathcal{F})$ in (02.03) **regulär** ist. \square

C02.15 **Anmerkung.** Alle bisher getroffenen Aussagen für $\mathbb{E}(\cdot|\mathcal{F})$ und $\mathbb{P}(\cdot|\mathcal{F})$ gelten damit für beliebige Teil- σ -Algebren von \mathcal{A} der Form $\mathcal{F} = \sigma(\{B\})$ mit $B \in \mathcal{A}$. Offensichtlich ist in der Situation $\mathbb{P}(B) = 0$ die Wahl $\mathbb{P}(\cdot|B) = \mathbb{P}$ willkürlich. Die Aussagen gelten genauso, wenn wir $\mathbb{P}(\cdot|B) = \tilde{\mathbb{P}}$ für ein beliebiges $\tilde{\mathbb{P}} \in \mathcal{W}(\mathcal{A})$ und mit der Erwartung $\tilde{\mathbb{E}}$ bzgl. $\tilde{\mathbb{P}}$ auch $\mathbb{E}(\cdot|B) = \tilde{\mathbb{E}}$ setzen. Es ist wichtig zu bemerken, dass in der Situation $\mathbb{P}(B) = 0$ eine andere Wahl $\mathbb{E}(\cdot|\mathcal{F})$ und $\mathbb{P}(\cdot|\mathcal{F})$ nur auf der Nullmenge B bzgl. \mathbb{P} ändert, also $\mathbb{E}(\cdot|\mathcal{F})$ und $\mathbb{P}(\cdot|\mathcal{F})$ eindeutig sind bis auf Gleichheit \mathbb{P} -f.ü.. \square

C02.16 **Skizze.** Betrachte den Wahrscheinlichkeitsraum $([0, 1), \mathcal{B}_{[0,1)}, U_{[0,1)})$. Die nächsten Graphiken stellen eine positive numerische Zufallsvariable $X : [0, 1) \rightarrow \mathbb{R}$ sowie für $\mathcal{F} = \sigma(\{B\})$ mit $B = [0, 0.5)$ und $B^c = [0.5, 1)$ die vorgestellte reguläre Festlegung des bedingten Erwartungswertes von X gegeben \mathcal{F} dar.





□

C02|02 Abzählbare Bedingungen

Wir betrachten nun eine Teil- σ -Algebra \mathcal{F} von \mathcal{A} , die von einer abzählbaren und messbaren Partition $\{B_i, i \in \mathcal{I}\}$ von Ω erzeugt ist, das heißt $\mathcal{F} = \sigma(\{B_i, i \in \mathcal{I}\})$.

C02.17 Bedingte Erwartung. Für eine abzählbare und messbare Partition $\{B_i, i \in \mathcal{I}\}$ von Ω , wobei wir wie bisher $\mathbb{P}(\cdot | B_i) = \mathbb{P}$ für $\mathbb{P}(B_i) = 0$ setzen, sei für jedes $i \in \mathcal{I}$ weiterhin $\mathbb{E}(\cdot | B_i)$ die Erwartung bzgl. $\mathbb{P}(\cdot | B_i)$. Für $\mathcal{F} = \sigma(\{B_i, i \in \mathcal{I}\})$ erfüllt dann die Abbildung

$$\mathbb{E}(\cdot | \mathcal{F}) : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathcal{F}}^+ \text{ mit } X \mapsto \mathbb{E}(X | \mathcal{F}) := \sum_{i \in \mathcal{I}} \mathbb{E}(X | B_i) \mathbf{1}_{B_i} \quad (02.05)$$

die Bedingungen (bE1) und (bE2). Betrachte (bE1). Für $X \in \overline{\mathcal{A}}^+$ gilt $\mathbb{E}(X | \mathcal{F}) \in \overline{\mathcal{F}}^+$. In der Tat sind die Abbildungen $f : \Omega \rightarrow \mathcal{I}$ mit $\omega \mapsto f(\omega) = i \Leftrightarrow \omega \in B_i$ und $g : \mathcal{I} \rightarrow \overline{\mathbb{R}}^+$ mit $i \mapsto \mathbb{E}(X | B_i)$ \mathcal{F} - $2^{\mathcal{I}}$ - bzw. $2^{\mathcal{I}}$ - $\overline{\mathcal{B}}^+$ -messbar, so dass $\mathbb{E}(X | \mathcal{F}) = g \circ f \in \overline{\mathcal{F}}^+$ eine \mathcal{F} - $\overline{\mathcal{B}}^+$ -messbare Abbildung ist. Andererseits, ist $F \in \mathcal{F}$, so existiert $\mathcal{J} \subseteq \mathcal{I}$ mit $F = \bigcup_{j \in \mathcal{J}} B_j$ und (bE2) gilt, da

$$\begin{aligned} \mathbb{E}(\mathbf{1}_F \mathbb{E}(X | \mathcal{F})) &= \sum_{j \in \mathcal{J}} \mathbb{E}(\mathbf{1}_{B_j} \mathbb{E}(X | \mathcal{F})) = \sum_{j \in \mathcal{J}} \mathbb{E}(\mathbf{1}_{B_j} \mathbb{E}(X | B_j)) \\ &= \sum_{j \in \mathcal{J}} \mathbb{P}(B_j) \mathbb{E}(X | B_j) = \sum_{j \in \mathcal{J}} \mathbb{E}(\mathbf{1}_{B_j} X) = \mathbb{E}(\mathbf{1}_F X). \end{aligned}$$

Damit erfüllt die Abbildung $\mathbb{E}(\cdot | \mathcal{F})$ in (02.05) die Bedingungen (bE1) sowie (bE2), und sie ist somit eine Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} . Wir halten weiterhin fest, dass die von der Festlegung $\mathbb{E}(\cdot | \mathcal{F})$ in (02.05) implizierte Festlegung der bedingten Verteilung

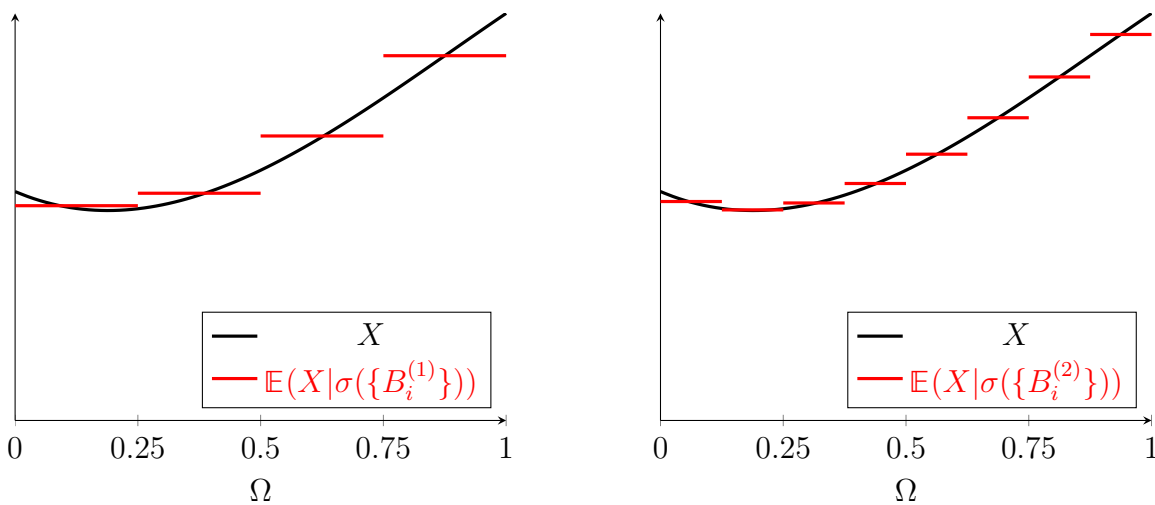
$$\mathbb{P}(\cdot | \mathcal{F}) : \mathcal{A} \rightarrow \overline{\mathcal{F}}^+ \text{ mit } A \mapsto \mathbb{P}(A | \mathcal{F}) := \mathbb{E}(\mathbf{1}_A | \mathcal{F}) = \sum_{i \in \mathcal{I}} \mathbb{P}(A | B_i) \mathbf{1}_{B_i} \quad (02.06)$$

ein Markovkern von (Ω, \mathcal{F}) nach (Ω, \mathcal{A}) ist. In der Tat die Abbildung $\Omega \times \mathcal{A} \rightarrow [0, 1]$ mit

$$(\omega, A) \mapsto \mathbb{P}(A | \mathcal{F})(\omega) = \sum_{i \in \mathcal{I}} \mathbb{P}(A | B_i) \mathbf{1}_{B_i}(\omega),$$

erfüllt die Bedingungen (Mk1) und (Mk2). Betrachte (Mk1). Da für jedes $\omega \in \Omega$ genau ein $i \in \mathcal{I}$ mit $\omega \in B_i$ existiert, ist nach Konstruktion $\mathbb{P}(\bullet | \mathcal{F})(\omega) = \mathbb{P}(\bullet | B_i)$ ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) . Andererseits erfüllt für jedes $A \in \mathcal{A}$ nach Konstruktion die Zufallsvariable $\mathbb{P}(A | \mathcal{F}) = \mathbb{E}(\mathbb{1}_A | \mathcal{F})$ die Bedingung (bE1), die Abbildung $\mathbb{P}(A | \mathcal{F}) : \Omega \rightarrow [0, 1]$ mit $\omega \mapsto \mathbb{P}(A | \mathcal{F})(\omega)$ ist also \mathcal{F} -messbar. Folglich ist die Festlegung $\mathbb{P}(\bullet | \mathcal{F})$ in (02.06) der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} regulär. Andererseits ist per Konstruktion für jedes $\omega \in \Omega$ die Abbildung $\mathbb{E}(\bullet | \mathcal{F})(\omega)$ die Erwartung bzgl. $\mathbb{P}(\bullet | \mathcal{F})(\omega)$, so dass auch die Festlegung $\mathbb{E}(\bullet | \mathcal{F})$ in (02.05) regulär ist. \square

C02.18 **Skizze.** Betrachte wie in Skizze C02.16 den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}_{[0,1]}, \mathbb{U}_{[0,1]})$. Die nächsten Graphiken stellen die vorgestellte reguläre Festlegung des bedingten Erwartungswertes von X gegeben $\mathcal{F} = \sigma(\{B_i^{(j)}\})$, $j \in \llbracket 2 \rrbracket$, für die zwei Partitionen $B_i^{(1)} := [\frac{i-1}{4}, \frac{i}{4})$, $i \in \llbracket 4 \rrbracket$ und $B_i^{(2)} := [\frac{i-1}{8}, \frac{i}{8})$, $i \in \llbracket 8 \rrbracket$ von $[0, 1)$ dar.



\square

C02|03 Induzierte bedingte Verteilung

C02.19 **Erinnerung.** Ist X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem messbaren Raum $(\mathcal{X}, \mathcal{X})$, so heißt $\mathbb{P}^X = \mathbb{P} \circ X^{-1} = \mathbb{P}(X^{-1}(\bullet))$ die von \mathbb{P} auf $(\mathcal{X}, \mathcal{X})$ induzierte Verteilung. \square

C02.20 **Definition.** Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem messbaren Raum $(\mathcal{X}, \mathcal{X})$ und $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra.

(a) Zu jeder Festlegung $\mathbb{P}(\bullet | \mathcal{F})$ der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} , heißt die von X induzierte Abbildung $\mathbb{P}^{X|\mathcal{F}} : \mathcal{X} \rightarrow \overline{\mathcal{F}}^+$ mit $B \mapsto \mathbb{P}^{X|\mathcal{F}}(B) := \mathbb{P}(X^{-1}(B) | \mathcal{F})$ Festlegung der bedingten Verteilung von X gegeben \mathcal{F} , wobei für alle $F \in \mathcal{F}$ und $B \in \mathcal{X}$ somit $\mathbb{E}(\mathbb{1}_{X^{-1}(B)} \mathbb{1}_F) = \mathbb{E}(\mathbb{P}^{X|\mathcal{F}}(B) \mathbb{1}_F)$ gilt. Ist weiterhin

$$(\omega, B) \mapsto \mathbb{P}^{X|\mathcal{F}}(B)(\omega)$$

ein Markovkern von (Ω, \mathcal{F}) nach $(\mathcal{X}, \mathcal{X})$, so wird diese Festlegung regulär genannt.

(b) Zu jeder Festlegung $\mathbb{E}(\bullet | \mathcal{F})$ der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} , heißt $\mathbb{E}^{X|\mathcal{F}}(h) := \mathbb{E}(h(X) | \mathcal{F})$ für $h \in \overline{\mathcal{X}}^+$ Festlegung des bedingten Erwartungswertes von $h(X)$ gegeben \mathcal{F} und die von X induzierte Abbildung $\mathbb{E}^{X|\mathcal{F}} : \overline{\mathcal{X}}^+ \rightarrow \overline{\mathcal{F}}^+$ mit $h \mapsto \mathbb{E}^{X|\mathcal{F}}(h)$ Festlegung der bedingten Erwartung von X gegeben \mathcal{F} , wobei für alle $F \in \mathcal{F}$ und $h \in \overline{\mathcal{X}}^+$ die

Radon-Nikodym-Gleichung (bE2) gilt $\mathbb{E}(h(X)\mathbb{1}_F) = \mathbb{E}(\mathbb{E}(h(X)|\mathcal{F})\mathbb{1}_F) = \mathbb{E}(\mathbb{E}^{X|\mathcal{F}}(h)\mathbb{1}_F)$. Die Festlegung $\mathbb{E}^{X|\mathcal{F}}$ heißt *regulär*, falls die induzierte Festlegung $\mathbb{P}^{X|\mathcal{F}}$ regulär ist, und für jedes $\omega \in \Omega$, $\mathbb{E}^{X|\mathcal{F}}(\cdot)(\omega)$ die Erwartung bzgl. $\mathbb{P}^{X|\mathcal{F}}(\cdot)(\omega)$ ist. \square

C02.21 **Bemerkung.** $\mathbb{P}^{X|\mathcal{F}}$ erfüllt definitionsgemäß die Bedingung (Mk2). Sei zusätzlich die Festlegung $\mathbb{P}(\cdot|\mathcal{F})$ regulär, so dass (Mk1) erfüllt ist, also für alle $\omega \in \Omega$ die Abbildung

$$A \mapsto \mathbb{P}(A|\mathcal{F})(\omega)$$

ein Wahrscheinlichkeitsmaß ist, kurz $\mathbb{P}(\cdot|\mathcal{F})(\omega) \in \mathcal{W}(\mathcal{A})$. Damit ist auch $\mathbb{P}^{X|\mathcal{F}}(\cdot)(\omega) = \mathbb{P}(X^{-1}(\cdot)|\mathcal{F})(\omega) \in \mathcal{W}(\mathcal{X})$ für jedes $\omega \in \Omega$. Die Festlegung $\mathbb{P}^{X|\mathcal{F}}$ erfüllt also auch (Mk1). Sie ist somit ein Markovkern, also regulär. Für eine reguläre Festlegung $\mathbb{E}(\cdot|\mathcal{F})$ ist definitionsgemäß die induzierte Festlegung $\mathbb{P}(\cdot|\mathcal{F})$ und somit auch $\mathbb{P}^{X|\mathcal{F}}$ regulär. Für jedes $\omega \in \Omega$ ist nach dem Transformationssatz (Eigenschaft A04.05 (viii)) die induzierte Festlegung $\mathbb{E}^{X|\mathcal{F}}(\cdot)(\omega)$ gerade die Erwartung bzgl. $\mathbb{P}^{X|\mathcal{F}}(\cdot)(\omega)$. \square

C02|04 Bedingte Erwartung gegeben eine Zufallsvariable

C02.22 **Erinnerung.** Sei $\mathcal{F} = \sigma(S)$ für eine Zufallsvariable S auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem messbaren Raum $(\mathcal{S}, \mathcal{S})$. Wir bezeichnen wie bisher mit \mathbb{P}^S das Bildmaß und mit \mathbb{E}^S die entsprechende Erwartung. Ist Y eine $\sigma(S)$ -messbare, positive numerische Zufallsvariable, also $Y \in \overline{\sigma(S)}^+$, so existiert nach Eigenschaft A02.06 (iv) ein $\varphi \in \overline{\mathcal{S}}^+$ mit $Y = \varphi(S)$, also $Y(\omega) = \varphi(S(\omega))$, $\omega \in \Omega$. Die Abbildung φ ist durch Y nur auf $S(\Omega)$ eindeutig festgelegt, für $s \notin S(\Omega)$ kann sie beliebig (messbar) fortgesetzt werden. \square

C02.23 **Definition.** Für $S : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ sei $\mathbb{E}(X|\sigma(S)) \in \overline{\sigma(S)}^+$ eine Festlegung des bedingten Erwartungswertes von $X \in \overline{\mathcal{A}}^+$ gegeben $\sigma(S)$ und $\varphi \in \overline{\mathcal{S}}^+$ wie in C02.22, also $\mathbb{E}(X|\sigma(S))(\omega) = \varphi(S(\omega))$, $\omega \in \Omega$. $\mathbb{E}(X|S) := \varphi$ und $\mathbb{E}(X|S = s) := \varphi(s)$ für $s \in \mathcal{S}$, heißen *Festlegung des bedingten Erwartungswertes von X gegeben S bzw. gegeben $S = s$* . Jede Abbildung $\mathbb{E}(\cdot|S) : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathcal{S}}^+$ mit $X \mapsto \mathbb{E}(X|S)$ heißt *Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben S* . Die von $\mathbb{E}(\cdot|S)$ implizierte Abbildung $\mathbb{P}(\cdot|S) : \mathcal{A} \rightarrow \overline{\mathcal{S}}^+$ mit $A \mapsto \mathbb{P}(A|S) := \mathbb{E}(\mathbb{1}_A|S)$ sowie $\mathbb{P}(\cdot|S = s) : \mathcal{A} \rightarrow \mathbb{R}^+$ mit $A \mapsto \mathbb{P}(A|S = s) := \mathbb{E}(\mathbb{1}_A|S = s)$ wird *Festlegung der bedingten Verteilung von \mathbb{P} gegeben S bzw. gegeben $S = s$* genannt. Die entsprechende Radon-Nikodym-Gleichheit (bE2) lautet $\mathbb{E}^S(\mathbb{1}_B \mathbb{P}(A|S)) = \int_B \mathbb{P}(A|S) d\mathbb{P}^S = \mathbb{P}(S^{-1}(B) \cap A)$ für alle $B \in \mathcal{S}$ und $A \in \mathcal{A}$. Die Festlegung $\mathbb{P}(\cdot|S)$ heißt *regulär*, falls $(s, A) \mapsto \mathbb{P}(A|S = s)$ ein Markovkern von $(\mathcal{S}, \mathcal{S})$ nach (Ω, \mathcal{A}) ist. Analog wird die Festlegung $\mathbb{E}(\cdot|S)$ *regulär* genannt, wenn die implizierte Festlegung $\mathbb{P}(\cdot|S)$ regulär ist, und für jedes $s \in \mathcal{S}$, $\mathbb{E}(\cdot|S = s)$ die Erwartung bzgl. $\mathbb{P}(\cdot|S = s)$ ist. \square

C02.24 **Eigenschaft.** Für $X \in \overline{\mathcal{A}}^+$ seien $\varphi(S)$ und $\tilde{\varphi}(S)$ mit $\varphi, \tilde{\varphi} \in \overline{\mathcal{S}}^+$ zwei Festlegungen des bedingten Erwartungswertes $\mathbb{E}(X|\sigma(S))$ von X gegeben $\sigma(S)$. Nach Eigenschaft C02.02 gilt $\varphi(S) = \tilde{\varphi}(S)$ \mathbb{P} -f.ü. und damit auch $\varphi = \tilde{\varphi}$ \mathbb{P}^S -f.ü.. Im Gegensatz zu $\mathbb{E}(X|\sigma(S))$ hängt φ nicht nur von $\sigma(S)$, sondern auch von der speziellen Wahl der Zufallsvariable S ab. \square

C02.25 **Bemerkung.** $\mathbb{P}(\cdot|S)$ erfüllt definitionsgemäß die Bedingung (Mk2). Sei zusätzlich die Festlegung $\mathbb{P}(\cdot|\sigma(S))$ regulär, so dass auch (Mk1) erfüllt ist, also für alle $\omega \in \Omega$ die Abbildung

$$A \mapsto \mathbb{P}(A|\sigma(S))(\omega) = \mathbb{P}(A|S = S(\omega))$$

ein Wahrscheinlichkeitsmaß ist, kurz $\mathbb{P}(\cdot | S = S(\omega)) \in \mathcal{W}(\mathcal{A})$. Damit gilt $\mathbb{P}(\cdot | S = s) \in \mathcal{W}(\mathcal{A})$ für jedes $s \in S(\Omega)$. Für jedes surjektive S erfüllt die Festlegung $\mathbb{P}(\cdot | S)$ also auch (Mk1). Sie ist somit ein Markovkern, also regulär. Ist S nicht surjektiv, aber $S(\Omega) \in \mathcal{S}$, so ist zum Beispiel die Festlegung $\mathbb{P}(\cdot | S) \mathbb{1}_{S(\Omega)} + \mathbb{P} \mathbb{1}_{S \setminus S(\Omega)}$ ein Markovkern. Analog, ist die Festlegung $\mathbb{E}(\cdot | \sigma(S))$ regulär, also $\mathbb{P}(\cdot | \sigma(S))$ ist regulär und für jedes $\omega \in \Omega$ ist $\mathbb{E}(\cdot | \sigma(S))(\omega) = \mathbb{E}(\cdot | S = S(\omega))$ die Erwartung bzgl. $\mathbb{P}(A | \sigma(S))(\omega) = \mathbb{P}(A | S = S(\omega))$. Ist S surjektiv, so ist auch $\mathbb{E}(\cdot | S)$ regulär, falls $S(\Omega) = \mathcal{S}$ gilt. Ist S nicht surjektiv, aber $S(\Omega) \in \mathcal{S}$, so ist die Festlegung $\mathbb{E}(\cdot | S) \mathbb{1}_{S(\Omega)} + \mathbb{E}(\cdot) \mathbb{1}_{S \setminus S(\Omega)}$ regulär. Andererseits, ist die Festlegung $\mathbb{P}(\cdot | S)$ regulär, also ein Markovkern von $(\mathcal{S}, \mathcal{S})$ nach (Ω, \mathcal{A}) , dann ist auch $(\omega, A) \mapsto \mathbb{P}(A | S = S(\omega)) =: \mathbb{P}(A | \sigma(S))(\omega)$ eine reguläre Festlegung der bedingten Verteilung von \mathbb{P} gegeben $\sigma(S)$, also ein Markovkern $(\Omega, \sigma(S))$ nach (Ω, \mathcal{A}) . Offensichtlich gilt (Mk2), da die Hintereinanderausführung messbarer Abbildungen messbar ist. Weiterhin gilt für alle $\omega \in \Omega$ auch $\mathbb{P}(A | S = S(\omega)) \in \mathcal{W}(\mathcal{A})$, also (Mk1). □

C02.26 **Beispiel.** Sei nun $S : (\Omega, \mathcal{A}) \rightarrow (\{0, 1\}, 2^{\{0,1\}})$ eine einfache Zufallsvariable, also $\{S = 1\} = S^{-1}(\{1\}) \in \mathcal{A}$ und $\sigma(S) = \sigma(\{S^{-1}(\{1\})\}) =: \mathcal{F}$. Eine Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben $\sigma(S)$ ist somit

$$\begin{aligned} \mathbb{E}(\cdot | \sigma(S)) : \overline{\mathcal{A}}^+ &\rightarrow \overline{\sigma(S)}^+ \text{ mit} \\ X &\mapsto \mathbb{E}(X | \sigma(S)) = \mathbb{E}(X | S^{-1}(\{1\})) \mathbb{1}_{S^{-1}(\{1\})} + \mathbb{E}(X | S^{-1}(\{0\})) \mathbb{1}_{S^{-1}(\{0\})} = \varphi(S) \\ \text{und } \varphi &\in \overline{2^{\{0,1\}}}^+ \text{ mit} \\ s &\mapsto \varphi(s) = \mathbb{E}(X | S^{-1}(\{1\})) \mathbb{1}_{\{1\}}(s) + \mathbb{E}(X | S^{-1}(\{0\})) \mathbb{1}_{\{0\}}(s). \end{aligned}$$

Damit ist $\mathbb{E}(\cdot | S) = \mathbb{E}(\cdot | S^{-1}(\{1\})) \mathbb{1}_{\{1\}} + \mathbb{E}(\cdot | S^{-1}(\{0\})) \mathbb{1}_{\{0\}}$ und $\mathbb{E}(\cdot | S = s) = \mathbb{E}(\cdot | S^{-1}(\{1\})) \mathbb{1}_{\{1\}}(s) + \mathbb{E}(\cdot | S^{-1}(\{0\})) \mathbb{1}_{\{0\}}(s)$, $s \in \{0, 1\}$ eine *Festlegung des bedingten Erwartungswertes von X gegeben S bzw. gegeben $S = s$* . Wir halten fest, dass insbesondere $\mathbb{E}(\cdot | S = 1) = \mathbb{E}(\cdot | S^{-1}(\{1\}))$ und $\mathbb{E}(\cdot | S = 0) = \mathbb{E}(\cdot | S^{-1}(\{0\}))$ gilt. □

C02|05 Induzierte bedingte Verteilung gegeben eine Zufallsvariable

C02.27 **Erinnerung.** Sei (S, X) eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in dem Produktraum $\mathcal{S} \times \mathcal{X}$ versehen mit der Produkt- σ -Algebra $\mathcal{S} \otimes \mathcal{X}$. Wir bezeichnen mit $\mathbb{P}^{(S, X)}$ die von (S, X) auf $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$ induzierte gemeinsame Verteilung. Seien wie bisher $\Pi_x : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{X}$ und $\Pi_s : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{S}$ mit $(s, x) \mapsto \Pi_x(s, x) := x$ bzw. $(s, x) \mapsto \Pi_s(s, x) := s$ die entsprechenden Koordinatenabbildungen. Für die Randverteilungen von X und S gilt dann $\mathbb{P}^X = \mathbb{P} \circ X^{-1} = \mathbb{P} \circ (\Pi_x(S, X))^{-1} = \mathbb{P}^{(S, X)} \circ \Pi_x^{-1}$ bzw. $\mathbb{P}^S = \mathbb{P}^{(S, X)} \circ \Pi_s^{-1}$. Die Statistiken Π_x und Π_s sind offensichtlich surjektiv. Unser Ziel ist eine bedingte Verteilung von X gegeben S einzuführen unter Verwendung der Definitionen C02.20 und C02.23. Geben wir uns eine Festlegung $\mathbb{P}(\cdot | \sigma(S))$ vor, so impliziert diese mit Definition C02.20 eine Festlegung $\mathbb{P}^{X | \sigma(S)}$ mit Werten in $\overline{\sigma(S)}^+$, welche weiterhin mit Definition C02.23 eine Festlegung $\mathbb{P}^{X | S}$ mit Werten in $\overline{\mathcal{S}}^+$ impliziert. Ist die Festlegung $\mathbb{P}(\cdot | \sigma(S))$ regulär, so auch $\mathbb{P}^{X | \sigma(S)}$, aber wie wir in Bemerkung C02.25 gesehen haben, ist im Allgemeinen $\mathbb{P}^{X | S}$ nicht regulär. Andererseits, können wir uns eine Festlegung $\mathbb{P}^{(S, X)}(\cdot | \sigma(\Pi_s))$ vorgeben, so impliziert diese mit Definition C02.20 eine Festlegung $\mathbb{P}^{X | \sigma(\Pi_s)}$ mit Werten in $\overline{\sigma(\Pi_s)}^+$, welche weiterhin mit Definition C02.23 eine Festlegung $\mathbb{P}^{X | S}$ mit Werten in $\overline{\mathcal{S}}^+$ impliziert. Da Π_s surjektiv ist, folgt die Regularität der Festlegung $\mathbb{P}^{X | S}$, falls wir mit einer regulären Festlegung $\mathbb{P}^{(S, X)}(\cdot | \sigma(\Pi_s))$ angefangen haben. □

C02.28 **Definition.** Sei $\mathbb{P}^{(S, X)}$ die gemeinsame Verteilung von (S, X) auf $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$.

- (a) Zu jeder Festlegung $\mathbb{P}^{(s,X)}(\cdot | \sigma(\Pi_s))$ auf $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$ der bedingten Verteilung von $\mathbb{P}^{(s,X)}$ gegeben $\sigma(\Pi_s)$, heißt die Abbildung

$$\mathbb{P}^{X|S} : \mathcal{X} \rightarrow \overline{\mathcal{S}}^+ \text{ mit } A \mapsto \mathbb{P}^{X|S}(A) := \varphi \text{ und}$$

$$\mathbb{P}^{X|\sigma(\Pi_s)}(A) = \mathbb{P}^{(X,S)}(\Pi_s^{-1}(A) | \sigma(\Pi_s)) = \varphi(\Pi_s)$$

Festlegung der bedingten Verteilung von X gegeben S und $\mathbb{P}^{X|S=s}(A) = \varphi(s)$ Festlegung der bedingten Wahrscheinlichkeit von $X \in A$ bei gegebenem $S = s$. Die entsprechende Radon-Nikodym-Gleichheit (bE2) lautet $\mathbb{P}^S(\mathbb{P}^{X|S}(A)\mathbb{1}_B) = \int_B \mathbb{P}^{X|S}(A) d\mathbb{P}^S = \mathbb{P}(X^{-1}(A) \cap S^{-1}(B)) = \mathbb{P}^{(s,X)}(B \times A)$ für alle $A \in \mathcal{X}$ und $B \in \mathcal{S}$. Ist weiterhin $(s, A) \mapsto \mathbb{P}^{X|S=s}(A)$ ein Markovkern von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{X}, \mathcal{X})$, so wird diese Festlegung *regulär* genannt, wobei dann auf Grund der Definition C02.03 (bE2) $\mathbb{P}^S \circ \mathbb{P}^{X|S} = \mathbb{P}^{(s,X)}$ gilt (vgl. Schreibweise C02.06). Besitzt in diesem Fall für ein $s \in \mathcal{S}$ die Verteilung $\mathbb{P}^{X|S=s}$ ein endliches erstes absolutes Moment, so schreiben wir auch $\mathbb{E}^{X|S=s}(X) := \mathbb{E}^{X|S=s}(\text{id}_X) = \int_{\mathcal{X}} x \mathbb{P}^{X|S=s}(dx)$.

- (b) Zu jeder Festlegung $\mathbb{E}^{(s,X)}(\cdot | \sigma(\Pi_s))$ der bedingten Erwartung bzgl. $\mathbb{P}^{(s,X)}$ gegeben $\sigma(\Pi_s)$, heißt die Abbildung

$$\mathbb{E}^{X|S} : \overline{\mathcal{X}}^+ \rightarrow \overline{\mathcal{S}}^+ \text{ mit } h \mapsto \mathbb{E}^{X|S}(h) := \varphi \text{ und}$$

$$\mathbb{E}^{X|\sigma(\Pi_s)}(h) = \mathbb{E}_{(s,X)}(h(\Pi_X) | \sigma(\Pi_s)) = \varphi(\Pi_s)$$

und $\mathbb{E}^{X|S=s}(h) = \varphi(s)$ *Festlegung der bedingten Erwartung von X gegeben S bzw. Festlegung des bedingten Erwartungswertes von $h(X)$ gegeben $S = s$.* Die Festlegung $\mathbb{E}^{X|S}$ heißt *regulär*, falls die induzierte Festlegung $\mathbb{P}^{X|S}$ regulär ist, und für jedes $s \in \mathcal{S}$, $\mathbb{E}^{X|S=s}$ die Erwartung bzgl. $\mathbb{P}^{X|S=s}$ ist. \square

C03 Integrierbare Zufallsvariablen

C03.01 **Erinnerung.** Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra und $X \in \overline{\mathcal{A}}$ eine numerische Zufallsvariable. Mit Hilfe der Zerlegung $X = X^+ - X^-$ mit $X^+, X^- \in \overline{\mathcal{A}}^+$ haben wir in Definition A04.03 (b) für X mit $\mathbb{E}(|X|) \in \mathbb{R}^+$ also $\mathbb{E}(X^+) \in \mathbb{R}^+$ und $\mathbb{E}(X^-) \in \mathbb{R}^+$ den Erwartungswert $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$ definiert. Im Folgenden sei $\mathcal{L}_1(\mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}} : \mathbb{E}(|X|) \in \mathbb{R}^+\}$ und zu \mathbb{P} bezeichne $\mathbb{E} : \mathcal{L}_1(\mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ die eindeutig bestimmte Erwartung. Da auch $\overline{\mathcal{F}} \subseteq \overline{\mathcal{A}}$ ist $\mathcal{L}_1(\mathcal{F}, \mathbb{P}) \subseteq \mathcal{L}_1(\mathcal{A}, \mathbb{P})$. Sei $X \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$, dann gilt $\mathbb{E}(X^+) \in \mathbb{R}^+$ und für jede Festlegung $\mathbb{E}(X^+ | \mathcal{F})$ gilt (bE1), $\mathbb{E}(X^+ | \mathcal{F}) \in \overline{\mathcal{F}}^+$ und (bE2), $\mathbb{E}(\mathbb{1}_F \mathbb{E}(X^+ | \mathcal{F})) = \mathbb{E}(\mathbb{1}_F X^+)$ für alle $F \in \mathcal{F}$, also insbesondere mit $F = \Omega$ auch $\mathbb{E}(\mathbb{E}(X^+ | \mathcal{F})) = \mathbb{E}(X^+) \in \mathbb{R}^+$. Somit kann $\mathbb{E}(X^+ | \mathcal{F}) \in \overline{\mathcal{F}}^+$ und analog $\mathbb{E}(X^- | \mathcal{F}) \in \overline{\mathcal{F}}^+$ gewählt werden. Damit erfüllt $Y := \mathbb{E}(X^+ | \mathcal{F}) - \mathbb{E}(X^- | \mathcal{F}) \in \overline{\mathcal{F}}$, als auch die Radon-Nikodym-Gleichung (bE2), wobei insbesondere $Y \in \mathcal{L}_1(\mathcal{F}, \mathbb{P})$. \square

C03.02 **Definition.** Für $X \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ heißt jede Abbildung $Y : \Omega \rightarrow \overline{\mathbb{R}}$ *Festlegung (Version) des bedingten Erwartungswertes* von X gegeben \mathcal{F} , symbolisch $\mathbb{E}(X | \mathcal{F}) := Y$, wenn sie die folgenden zwei Bedingungen erfüllt:

(bE1) Y ist \mathcal{F} - $\overline{\mathcal{B}}$ -messbar, also $Y \in \overline{\mathcal{F}}$ und

(bE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ für jedes $F \in \mathcal{F}$. (Radon-Nikodym-Gleichung)

Jede Abbildung $\mathbb{E}(\cdot | \mathcal{F}) : \mathcal{L}_1(\mathcal{A}, \mathbb{P}) \rightarrow \overline{\mathcal{F}}$ mit $X \mapsto \mathbb{E}(X | \mathcal{F})$ heißt *Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F}* . $\mathbb{E}(\cdot | \mathcal{F})$ heißt *reguläre Festlegung* der bedingten Erwartung

bzgl. \mathbb{P} gegeben \mathcal{F} , wenn die implizierte Festlegung $\mathbb{P}(\cdot | \mathcal{F})$ der bedingten Verteilung von \mathbb{P} gegeben \mathcal{F} regulär ist, und für jedes $\omega \in \Omega$ die Abbildung $\mathbb{E}(\cdot | \mathcal{F})(\omega) : \mathcal{L}_1(\mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ die Erwartung bzgl. $\mathbb{P}(\cdot | \mathcal{F})(\omega)$ ist. □

C03.03 **Bemerkung.**

- (i) Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem messbaren Raum $(\mathcal{X}, \mathcal{X})$ und $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra. Dann definieren wir analog zu Definition C02.20 (b) nun für $h \in \mathcal{L}_1(\mathcal{X}, \mathbb{P}^x)$ eine Festlegung $\mathbb{E}^{X|\mathcal{F}}(h) \in \overline{\mathcal{F}}$ des bedingten Erwartungswertes von $h(X)$ gegeben \mathcal{F} sowie eine (reguläre) Festlegung $\mathbb{E}^{X|\mathcal{F}} : \mathcal{L}_1(\mathcal{X}, \mathbb{P}^x) \rightarrow \overline{\mathcal{F}}$ mit $h \mapsto \mathbb{E}^{X|\mathcal{F}}(h)$ der bedingten Erwartung von X gegeben \mathcal{F} .
- (ii) Für $S : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ definieren wir nun für $X \in \mathcal{L}_1(\mathcal{X}, \mathbb{P})$ wie in Definition C02.23 eine Festlegung $\mathbb{E}(X|S) \in \overline{\mathcal{S}}$ und $\mathbb{E}(X|S=s) \in \overline{\mathbb{R}}$ des bedingten Erwartungswertes von X gegeben S bzw. gegeben $S=s$ sowie eine (reguläre) Festlegung $\mathbb{E}(\cdot | S) : \mathcal{L}_1(\mathcal{A}, \mathbb{P}) \rightarrow \overline{\mathcal{S}}$ mit $X \mapsto \mathbb{E}(X|S)$ der bedingten Erwartung bzgl. \mathbb{P} gegeben S .
- (iii) Sei $(S, X) : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$ mit gemeinsamer Verteilung $\mathbb{P}^{(S,X)}$. Analog zu Definition C02.28 (b) definiere für $h \in \mathcal{L}_1(\mathcal{X}, \mathbb{P}^x)$ eine Festlegung $\mathbb{E}^{X|S}(h) \in \overline{\mathcal{S}}$ und $\mathbb{E}^{X|S=s}(h) \in \overline{\mathbb{R}}$ des bedingten Erwartungswertes von $h(X)$ gegeben S bzw. gegeben $S=s$ sowie eine (reguläre) Festlegung $\mathbb{E}^{X|S} : \mathcal{L}_1(\mathcal{X}, \mathbb{P}^x) \rightarrow \overline{\mathcal{S}}$ mit $h \mapsto \mathbb{E}^{X|S}(h)$ der bedingten Erwartung von X gegeben S . Ist analog zu Definition C02.28 (a) $\mathbb{P}^{X|S}$ eine reguläre Festlegung der bedingten Verteilung von X gegeben S , und besitzt für ein $s \in \mathcal{S}$ die Verteilung $\mathbb{P}^{X|S=s}$ zum Beispiel ein endliches erstes absolutes Moment, so schreiben wir auch $\mathbb{E}^{X|S=s}(X) := \mathbb{E}^{X|S=s}(\text{id}_1) = \int_{\mathcal{X}} x \mathbb{P}^{X|S=s}(dx)$. □

C03.04 **Eigenschaft.** Seien $X, Y \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ und $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra. Für jede Festlegung der bedingten Erwartung bzgl. \mathbb{P} gegeben \mathcal{F} gelten die folgenden Aussagen \mathbb{P} -f.s.,

- (i) Für alle $a, b \in \mathbb{R}$ gilt $\mathbb{E}(aX + bY | \mathcal{F}) = a\mathbb{E}(X | \mathcal{F}) + b\mathbb{E}(Y | \mathcal{F})$; (linear)
- (ii) Für $X \leq Y$ gilt $\mathbb{E}(X | \mathcal{F}) \leq \mathbb{E}(Y | \mathcal{F})$; (monoton)
- (iii) $|\mathbb{E}(X | \mathcal{F})| \leq \mathbb{E}(|X| | \mathcal{F})$; (Dreiecksungleichung)
- (iv) Für $S \in \overline{\mathcal{A}}$ mit $\mathbb{E}(|S| | \mathcal{F}) \in \mathcal{B}^+$ gilt $\mathbb{P}(|S| = \infty) = 0$. (endlich)
- (v) Für $\phi : \mathbb{R} \rightarrow \mathbb{R}$ konvex mit $\phi(X) \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ gilt $\phi(\mathbb{E}(X | \mathcal{F})) \leq \mathbb{E}(\phi(X) | \mathcal{F})$. (Ungleichung von Jensen)
- (vi) Für $X_n \uparrow X$ \mathbb{P} -f.ü. gilt $\sup_{n \in \mathbb{N}} \mathbb{E}(X_n | \mathcal{F}) = \mathbb{E}(X | \mathcal{F})$. (monotone Konvergenz)
- (vii) Für $X_n \rightarrow X$ \mathbb{P} -f.ü. mit $|X_n| \leq Y$, $n \in \mathbb{N}$, gilt $\lim_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{F}) = \mathbb{E}(X | \mathcal{F})$ \mathbb{P} -f.s. und in $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$. (dominierte Konvergenz)

Ist die Festlegung regulär, also $\mathbb{E}(\cdot | \mathcal{F})(\omega)$ eine Erwartung für alle $\omega \in \Omega$, so gelten die Aussagen (i)-(vii) für alle $\omega \in \Omega$. □

C03.05 **Eigenschaft.** Seien $X, Y \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ und $\mathcal{G} \subseteq \mathcal{F} \subseteq \mathcal{A}$ Teil- σ -Algebren. Für jede Festlegung der bedingten Erwartung gelten die folgenden Aussagen \mathbb{P} -f.s.,

- (i) Für $\mathbb{E}(|XY|) < \infty$ und $Y \in \mathcal{F}$ gelten

$$\mathbb{E}(XY | \mathcal{F}) = Y\mathbb{E}(X | \mathcal{F}) \text{ und } \mathbb{E}(Y | \mathcal{F}) = \mathbb{E}(Y | \sigma(Y)) = Y$$
- (ii) $\mathbb{E}(\mathbb{E}(X | \mathcal{F}) | \mathcal{G}) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{F}) = \mathbb{E}(X | \mathcal{G})$; (Turmeigenschaft)
- (iii) Sind $\sigma(X)$ und \mathcal{F} unabhängig, so gilt $\mathbb{E}(X | \mathcal{F}) = \mathbb{E}(X)$; (unabhängig)

- (iv) Für $\overline{\mathcal{F}} := \{A \in \mathcal{A} \mid \mathbb{P}(A) \in \{0, 1\}\}$ gilt $\mathbb{E}(X|\overline{\mathcal{F}}) = \mathbb{E}(X)$.
- (v) $\mathbb{E}(\mathbb{E}(X|\mathcal{F})) = \mathbb{E}(X)$. (Satz von der totalen Wahrscheinlichkeit)
- Seien weiterhin S und T Zufallsvariablen von (Ω, \mathcal{A}) nach $(\mathcal{S}, \mathcal{S})$ sowie $(\mathcal{T}, \mathcal{T})$.
- (vi) Für $h \in \mathcal{S}$ mit $\mathbb{E}(|h(S)X|) < \infty$ gilt $\mathbb{E}(h(S)X|S) = h(S)\mathbb{E}(X|S)$ \mathbb{P}^S -f.s.
- (vii) Sind (X, S) und (Y, T) unabhängig, so gilt $\mathbb{E}(XY|(S, T)) = \mathbb{E}(X|S)\mathbb{E}(Y|T)$ $\mathbb{P}^{(S, T)}$ -f.s.. □

C03.06 **Eigenschaft.** Es seien $\mathbb{P}_0, \mathbb{P} \in \mathcal{W}(\mathcal{A})$ Wahrscheinlichkeitsmaße mit $\mathbb{P}_0 \ll \mathbb{P}$. Für jede Zufallsvariable $S : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ gilt:

- (i) $\mathbb{P}_0^S \ll \mathbb{P}^S$.
- (ii) $\frac{d\mathbb{P}_0^S}{d\mathbb{P}^S} = \mathbb{P} \left(\frac{d\mathbb{P}_0}{d\mathbb{P}} \middle| S \right)$ \mathbb{P}^S -f.s.
- (iii) Ist $h \in \overline{\mathcal{A}}$, für die $\mathbb{P}_0(h)$ existiert, so gilt $\mathbb{P} \left(h \frac{d\mathbb{P}_0}{d\mathbb{P}} \middle| S \right) = \mathbb{P}_0(h|S) \frac{d\mathbb{P}_0^S}{d\mathbb{P}^S}$ \mathbb{P}^S -f.s.. Insbesondere gilt $\mathbb{P}_0(h|S) = \frac{\mathbb{P}(hf|S)}{\mathbb{P}(f|S)}$ \mathbb{P}^S -f.s. mit $f := d\mathbb{P}_0/d\mathbb{P}$.

Bezeichnet $\tilde{\mathbb{P}} \in \mathcal{W}(\mathcal{A})$ ein weiteres Wahrscheinlichkeitsmaß mit $\tilde{\mathbb{P}} \ll \mathbb{P}$, L einen Dichtequotienten (DQ) von $\tilde{\mathbb{P}}$ bezüglich \mathbb{P} , L^S einen DQ von $\tilde{\mathbb{P}}^S$ bezüglich \mathbb{P}^S , sowie \mathcal{N} bzw. \mathcal{N}^S die singulären Bereiche von $\tilde{\mathbb{P}}$ bezüglich \mathbb{P} bzw. von $\tilde{\mathbb{P}}^S$ bezüglich \mathbb{P}^S , so gilt mit $\varphi := \tilde{\mathbb{P}}(\mathcal{N}^c|S)$

- (iv) $\tilde{\mathbb{P}}(\mathcal{N}) = \tilde{\mathbb{P}}^S(\mathcal{N}^S) + \mathbb{P}(L^S(S)(1 - \varphi(S)))$,
- (v) $\tilde{\mathbb{P}}(L|S) = L^S\varphi$ \mathbb{P}^S -f.s.. □

C03|01 Reguläre Festlegungen

C03.07 **Erinnerung.** Sei (S, X) eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in dem Produktraum $\mathcal{S} \times \mathcal{X}$ versehen mit der Produkt- σ -Algebra $\mathcal{S} \otimes \mathcal{X}$. Eine reguläre Festlegung der bedingten Verteilung $\mathbb{P}^{X|S}$ von X gegeben S , wenn sie existiert, ist aufgefasst als Abbildung $(s, A) \mapsto \mathbb{P}^{X|S=s}(A)$ ein Markovkern von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{X}, \mathcal{X})$, der Lösung der Radon-Nikodym-Gleichung $\mathbb{P}^S \odot \mathbb{P}^{X|S} = \mathbb{P}^{(S, X)}$ ist, d.h. für jedes $A \in \mathcal{X}$ gilt

$$\mathbb{P}^{(S, X)}(B \times A) = \mathbb{P}(S^{-1}(B) \cap X^{-1}(A)) = \int_B \mathbb{P}^{X|S}(A) d\mathbb{P}^S = \mathbb{P}^S(\mathbb{1}_B \mathbb{P}^{X|S}(A)) \quad \forall B \in \mathcal{S}$$

Häufig ist es nützlich, neben $\mathbb{P}^{X|S}$ auch eine reguläre Festlegung der bedingten Verteilung $\mathbb{P}^{(S, X)|S}$ von (S, X) bei gegebenen S zu betrachten, also einen Markovkern $(s, D) \mapsto \mathbb{P}^{(S, X)|S=s}(D)$ von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$, der für jedes $D \in \mathcal{S} \otimes \mathcal{X}$ Lösung der entsprechenden Radon-Nikodym-Gleichung

$$\begin{aligned} \mathbb{P}^{(S, X)}(D \cap (B \times \mathcal{X})) &= \mathbb{P}((S, X)^{-1}(D) \cap S^{-1}(B)) = \int_B \mathbb{P}^{(S, X)|S}(D) d\mathbb{P}^S \\ &= \mathbb{P}^S(\mathbb{P}^{(S, X)|S}(D)\mathbb{1}_B) \quad \forall B \in \mathcal{S} \end{aligned}$$

ist. Dann ergibt sich die bedingte Verteilung $\mathbb{P}^{X|S=s}$ aus der bedingten Verteilung $\mathbb{P}^{(S, X)|S=s}$ wieder als Randverteilung gemäß $\mathbb{P}^{X|S=s}(A) = \mathbb{P}^{(S, X)|S=s}(\mathcal{S} \times A)$, $A \in \mathcal{X}$. Wir bezeichnen weiterhin mit $D_s := \{x \in \mathcal{X} : (s, x) \in D\}$ bzw. $f_s(x) := f(s, x)$ den *s-Schnitt* von $D \subseteq \mathcal{S} \times \mathcal{X}$ bzw. $f : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}$, jeweils bei gegebenem $s \in \mathcal{S}$. □

C03.08 **Eigenschaft.** Sei (S, X) eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$. Es existiere eine reguläre Festlegung $\mathbb{P}^{X|S}$ der bedingten Verteilung von X bei gegebenen S .

(i) Dann ist $(s, D) \mapsto \mathbb{P}^{(S,X)|S=s}(D) := \mathbb{P}^{X|S=s}(D_s)$ eine reguläre Festlegung der bedingten Verteilung von (S, X) bei gegebenen S .

(ii) Für jede Funktion $f \in \mathcal{S} \otimes \mathcal{X}$, für die $\mathbb{P}^{(S,X)}(f) = \mathbb{E}(f(S, X))$ existiert, gilt

$$\mathbb{P}^{(S,X)|S=s}(f) = \int f_s d\mathbb{P}^{X|S=s} = \mathbb{P}^{X|S=s}(f_s) \quad \text{für } \mathbb{P}^S\text{-f.a. } s \in \mathcal{S},$$

sowie für alle $A \in \mathcal{X}$ und $B \in \mathcal{S}$ gilt

$$\mathbb{P}^{(S,X)}(f \mathbf{1}_{B \times A}) = \mathbb{P}^S(\mathbf{1}_B \mathbb{P}^{(S,X)|S}(\mathbf{1}_A f)) = \int_B \left(\int_A f_s d\mathbb{P}^{X|S=s} \right) \mathbb{P}^S(ds). \quad \square$$

C03.09 **Bemerkung.** Intuitiv liegt es nahe, $\mathbb{P}^{X|S}(f_s)$ für $\mathbb{P}^{(X,S)|S}(f)$ zu schreiben. Das ist jedoch nicht gerechtfertigt, da sich $\mathbb{P}^{X|S}(f_s)$ nicht als \mathcal{S} -messbare Lösung einer Radon-Nikodym-Gleichung gewinnen lässt, und somit nicht in der üblichen Weise als bedingter Erwartungswert formuliert werden kann (siehe Witting [1985, S.125] für eine weiterführende Diskussion). Andererseits kann zur vereinfachenden Berechnung im Integranden $f(S, X)$ durch $f(s, X)$ ersetzt werden, also die symbolische Schreibweise $S = s$ als ein Einsetzen verstanden werden. Die Rechtfertigung eines derartigen Einsetzens folgt daraus, dass $(s, D) \mapsto \mathbb{P}^{X|S=s}(D_s)$ unter den Voraussetzungen von **Eigenschaft** C03.20 eine Version der regulären Festlegung $\mathbb{P}^{(S,X)|S}$ ist. Für $D = \{s\} \times \mathcal{X}$ gilt damit $\mathbb{P}^{(S,X)|S=s}(D) = \mathbb{P}^{X|S=s}(\mathcal{X}) = 1$ für \mathbb{P}^S -f.a. $s \in \mathcal{S}$. Die bedingte Verteilung $\mathbb{P}^{(S,X)|S}$ konzentriert sich also für \mathbb{P}^S -f.a. $s \in \mathcal{S}$ auf der Menge $\{S = s\}$. □

C03.10 **Beispiel.**

(a) Seien $X, S \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ unabhängig. Dann gilt \mathbb{P} -f.s.

$$\mathbb{E}(X + S | \sigma(S)) = \mathbb{E}(X | \sigma(S)) + \mathbb{E}(S | \sigma(S)) = \mathbb{E}(X) + S.$$

Allgemeiner, sei $f \in \mathcal{S} \otimes \mathcal{X}$, für die $\mathbb{P}^{(S,X)}(f) = \mathbb{E}(f(S, X))$ existiert. Da $X \perp\!\!\!\perp S$ (siehe auch **Eigenschaft** C03.11) ist der Markovkern $(s, A) \mapsto \mathbb{P}^X(A)$ von $(\mathcal{S}, \mathcal{S})$ nach $(\mathcal{X}, \mathcal{X})$ eine reguläre Festlegung der bedingten Verteilung von X bei gegebenen S . Unter Verwendung von **Eigenschaft** C03.08 (ii) gilt damit $\mathbb{E}(f(S, X) | S = s) = \mathbb{P}^{(S,X)|S=s}(f) = \mathbb{P}^X(f_s) = \mathbb{E}(f(s, X))$ für \mathbb{P}^S -f.a. $s \in \mathcal{S}$.

(b) Sei $(X_i)_{i \in \llbracket n \rrbracket}$ eine unabhängige Familie aus $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$ mit $\mathbb{E}(X_i) = 0, i \in \llbracket n \rrbracket$. Für $m \in \llbracket n \rrbracket$ setze $\mathcal{F}_m := \sigma((X_i)_{i \in \llbracket m \rrbracket})$ und $S_m := \sum_{i \in \llbracket m \rrbracket} X_i$, dann gilt \mathbb{P} -f.s.

$$\mathbb{E}(S_n | \mathcal{F}_m) = \sum_{i \in \llbracket n \rrbracket} \mathbb{E}(X_i | \mathcal{F}_m) = \sum_{i \in \llbracket m \rrbracket} X_i + \sum_{i \in \llbracket m+1, n \rrbracket} \mathbb{E}(X_i) = S_m.$$

Da $\sigma(S_m) \subseteq \mathcal{F}_m$ folgt mit **Eigenschaft** C03.05 (ii) auch \mathbb{P} -f.s.

$$\mathbb{E}(S_n | \sigma(S_m)) = \mathbb{E}(\mathbb{E}(S_n | \mathcal{F}_m) | \sigma(S_m)) = \mathbb{E}(S_m | \sigma(S_m)) = S_m. \quad \square$$

C03.11 **Eigenschaft.** Sei (S, X) eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$. Existiert eine reguläre Festlegung $\mathbb{P}^{X|S}$ der bedingten Verteilung von X bei gegebenen S , so sind die folgenden drei Aussagen äquivalent:

(i) X und S sind unabhängig, kurz $X \perp\!\!\!\perp S$, unter \mathbb{P} .

(ii) Es gibt eine Festlegung $\mathbb{P}^{X|S}$ die unabhängig von S ist.

(iii) $\mathbb{P}^{X|S} = \mathbb{P}^X$ \mathbb{P}^S -f.ü..

In dem Fall $X \perp\!\!\!\perp S$ unter \mathbb{P} gibt es für jedes $g : (\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X}) \rightarrow (\mathcal{T}, \mathcal{T})$ eine \mathbb{P}^S -Nullmenge N derart, dass $\mathbb{P}^{g(S,X)|S=s}(C) = \mathbb{P}^{g(S,X)}(C)$ für alle $C \in \mathcal{T}$ und $s \in N^c$ gilt. □

C03|02 Bedingte Dichten

C03.12 **Vorbemerkung.** Eine explizite Darstellung einer regulären Festlegung der bedingten Verteilung $\mathbb{P}^{X|S}$ von X bei gegebenem S , und somit eine Möglichkeit des Nachweises der Existenz, können wir angeben, wenn die gemeinsame Verteilung eine Dichte bezüglich eines Produktmaßes σ -endlicher Maße besitzt. Wir bezeichnen weiterhin mit $D_s := \{x \in \mathcal{X} : (s, x) \in D\}$ bzw. $f_s(x) := f(s, x)$ den s -Schnitt von $D \subseteq \mathcal{S} \times \mathcal{X}$ bzw. $f : \mathcal{S} \times \mathcal{X} \rightarrow \overline{\mathbb{R}}$, jeweils bei gegebenem $s \in \mathcal{S}$. \square

C03.13 **Definition.** Seien (S, X) eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$ und $\mathbb{P}^{(S, X)} \ll \nu \otimes \mu$ mit $\nu \otimes \mu$ -Dichte $f^{(S, X)} \in (\mathcal{S} \otimes \mathcal{X})^+$ für σ -endliche Maße $\mu \in \mathcal{M}_\sigma(\mathcal{X})$ und $\nu \in \mathcal{M}_\sigma(\mathcal{S})$. Weiter bezeichne $f^X \in \mathcal{X}^+$ und $f^S \in \mathcal{S}^+$ Festlegungen der Randdichten (bzgl. μ bzw. ν), d.h. $f^S(s) = \int (f^{(S, X)})_s d\mu = \mu((f^{(S, X)})_s)$ ν -f.ü. und $f^X(x) = \nu((f^{(S, X)})^x)$ μ -f.ü.. Die $\mathcal{S} \otimes \mathcal{X}$ - \mathbb{R}^+ - bzw. \mathcal{X} - \mathbb{R}^+ -messbaren Abbildungen

$$\begin{aligned} \mathbb{f}^{X|S} &:= (f^{(S, X)} / f^S) \mathbb{1}_{\{f^S \in \mathbb{R}_0^+\}} + f^X \mathbb{1}_{\{f^S = 0\}} : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}^+ \text{ mit} \\ (s, x) &\mapsto \mathbb{f}^{X|S=s}(x) := \frac{f^{(S, X)}(s, x)}{f^S(s)} \mathbb{1}_{\{f^S(s) \in \mathbb{R}_0^+\}} + f^X(x) \mathbb{1}_{\{f^S(s) = 0\}} \end{aligned} \quad (03.01)$$

und $\mathbb{f}^{X|S=s} : \mathcal{X} \rightarrow \mathbb{R}^+$ heißen *reguläre Festlegungen der bedingten μ -Dichte* von X bei gegebenem S bzw. $S = s$. \square

C03.14 **Eigenschaft.** Unter den Annahmen von *Definition C03.13* gilt

- (i) Es gibt reguläre Festlegungen $\mathbb{P}^{X|S}$ und $\mathbb{P}^{(S, X)|S}$ von X bzw. (S, X) bei gegebenem S . Diese lassen sich explizit angeben und zwar bei festem $s \in \mathcal{S}$ mit $\mathbb{P}^{X|S=s} = \mathbb{f}^{X|S=s} \mu \in \mathcal{W}(\mathcal{X})$ und $\mathbb{P}^{(S, X)|S=s} \in \mathcal{W}(\mathcal{S} \otimes \mathcal{X})$ mit $D \mapsto \mathbb{P}^{(S, X)|S=s}(D) = \mu(\mathbb{1}_{D_s} \mathbb{f}^{X|S=s}) = \int_{D_s} \mathbb{f}^{X|S=s} d\mu$.
- (ii) Für jede Funktion $h \in \overline{(\mathcal{S} \otimes \mathcal{X})}$, für die $\mathbb{P}^{(S, X)}(h) = \mathbb{E}(h(S, X))$ existiert, gilt

$$\mathbb{P}^{(S, X)|S=s}(h) = \mu(h_s \mathbb{f}^{X|S=s}) = \mathbb{f}^{X|S=s} \mu(h_s) \quad \text{für } \mathbb{P}^S\text{-f.a. } s \in \mathcal{S},$$

sowie für all $A \in \mathcal{X}$ und $B \in \mathcal{S}$ gilt

$$\mathbb{P}^{(S, X)}(h \mathbb{1}_{B \times A}) = \nu(\mathbb{1}_B \mathbb{f}^S \mathbb{P}^{(S, X)|S}(\mathbb{1}_A h)) = \int_B \mu(\mathbb{1}_A h_s \mathbb{f}^{X|S=s}) f^S(s) \nu(ds). \quad \square$$

C03.15 **Bemerkung.** Die in der *Definition C01.04 (a)* und *(b)* eingeführten Begriffe können wir nun in den allgemeineren Ansatz einbetten. Dazu sei (S, X) eine Zufallsvariable mit gemeinsamer Verteilung $\mathbb{P}^{(S, X)}$ auf $(\mathcal{S} \times \mathcal{X}, \mathcal{S} \otimes \mathcal{X})$.

- (i) Seien $\mathcal{S} \times \mathcal{X}$ abzählbar, also (S, X) *diskret-verteilt*, wie in *Definition C01.04 (a)*. Für jedes $s \in \mathcal{S}$ bezeichnen wir die Zähl-dichte $\mathbb{p}^{X|S=s} : 2^{\mathcal{X}} \rightarrow [0, 1]$ (bzgl. des Zählmaßes ζ_x) mit $x \mapsto \mathbb{p}^{X|S=s}(x) = \frac{\mathbb{p}^{(S, X)}(s, x)}{\mathbb{p}^S(s)} \mathbb{1}_{\{\mathbb{p}^S(s) \in \mathbb{R}_0^+\}} + \mathbb{p}^X(x) \mathbb{1}_{\{\mathbb{p}^S(s) = 0\}}$ als *reguläre Festlegung der bedingten Zähl-dichte von X bei gegebenem $S = s$* . Dann ist $(s, A) \mapsto \mathbb{P}^{X|S=s}(A)$ ein Markovkern von $(\mathcal{S}, 2^{\mathcal{S}})$ nach $(\mathcal{X}, 2^{\mathcal{X}})$, der *(bE1)*, also $\mathbb{P}^{X|S}(A) : 2^{\mathcal{S}} \rightarrow [0, 1]$, als auch die Bedingung *(bE2)* erfüllt. In der Tat, für jedes $A \in 2^{\mathcal{X}}$ und jedes $B \in 2^{\mathcal{S}}$ gilt

$$\begin{aligned} \mathbb{P}^{(S, X)}(\mathbb{1}_A(X) \mathbb{1}_B(S)) &= \sum_{s \in \mathcal{S}} \mathbb{1}_B(s) \sum_{x \in \mathcal{X}} \mathbb{1}_A(x) \mathbb{p}^{(S, X)}(x, s) \\ &= \sum_{s \in \mathcal{S}} \mathbb{1}_B(s) \mathbb{p}^S(s) \mathbb{P}^{X|S=s}(A) = \mathbb{P}^S(\mathbb{P}^{X|S}(A) \mathbb{1}_B). \end{aligned}$$

(ii) Seien (S, X) *stetig-verteilt* mit Werten in $(\mathbb{R}^{m+n}, \mathcal{B}^{m+n})$, wie in **Definition** C01.04 (b). Für jedes $s \in \mathbb{R}^m$ bezeichnen wir die Dichte $f^{X|S=s} \in (\mathcal{B}^n)^+$ (bzgl. des Lebesgue-Maßes λ^n) mit $x \mapsto f^{X|S=s}(x) = \frac{f^{(S,X)}(s,x)}{f^S(s)} \mathbb{1}_{\{f^S(s) \in \mathbb{R}_0^+\}} + f^X(x) \mathbb{1}_{\{f^S(s)=0\}}$ als *reguläre Festlegung der bedingten Lebesgue-Dichte von X bei gegebenem $S = s$* . Dann ist $(s, A) \mapsto \mathbb{P}^{X|S=s}(A)$ ein Markovkern von $(\mathbb{R}^m, \mathcal{B}^m)$ nach $(\mathbb{R}^n, \mathcal{B}^n)$, der (bE1), also $\mathbb{P}^{X|S}(A) \in (\mathcal{B}^m)^+$, als auch die Bedingung (bE2) erfüllt. In der Tat, für jedes $A \in \mathcal{B}^n$ und jedes $B \in \mathcal{B}^m$ gilt

$$\begin{aligned} \mathbb{P}^{(S,X)}(\mathbb{1}_A(X)\mathbb{1}_B(S)) &= \int_{\mathbb{R}^m} \mathbb{1}_B(s) \int_{\mathbb{R}^n} \mathbb{1}_A(x) f^{(S,X)}(s,x) dx ds \\ &= \int_{\mathbb{R}^m} \mathbb{1}_B(s) f^S(s) \mathbb{P}^{X|S=s}(A) ds = \mathbb{E}^S(\mathbb{P}^{X|S}(A)\mathbb{1}_B). \end{aligned}$$

□

C03.16 **Eigenschaft.** Unter den Annahmen von **Definition** C03.13 sind X und S genau dann unabhängig, kurz $X \perp\!\!\!\perp S$, unter \mathbb{P} , wenn $f^{X|S} = f^X$ \mathbb{P}^S -f.ü. gilt. □

C03.17 **Beispiel.**

(a) Unter Verwendung von **Eigenschaft** C03.16 sind im **Beispiel** C01.06 (a) X und S somit nicht unabhängig, da zum Bsp. $\mathbb{P}^{X|S=2} \neq \mathbb{P}^X$ mit $\mathbb{P}^S(\{2\}) = \mathbb{P}^S(2) = 1/36 > 0$ gilt. Andererseits im **Beispiel** C01.06 (b) sind X und S unabhängig, d.h. $\mathbb{P}^{X|S} = \mathbb{P}^X$ gilt \mathbb{P}^S -f.s., genau dann, wenn $\Sigma_{XS} = 0$ gilt, also wenn X und S unkorreliert sind.

(b) Für $a \in \mathbb{R}_0^+$ seien $X \sim \text{Exp}_a$ und $Y \sim \text{Exp}_a$ unabhängig. Dann sind X und $S = X + Y$ abhängig, und es gilt $\mathbb{P}^{X|S=s} = U_{[0,s]}$ für λ -f.a. $s \in \mathbb{R}^+$. Dagegen sind $W := X/(X + Y)$ und S unabhängig, und es gilt $\mathbb{P}^{W|S} = U_{[0,1]}$ λ -f.ü.. (Übung.) □

C03|03 Beste Vorhersage

C03.18 **Vorbemerkung.** Sei $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ ein Hilbertraum versehen mit der induzierten Norm $\|\cdot\|_{\mathbb{H}}$ und \mathbb{U} ein abgeschlossener linearer Unterraum von \mathbb{H} . Für jedes Element $h \in \mathbb{H}$ existiert dann ein eindeutig bestimmtes Element $u_h \in \mathbb{U}$ mit $\|h - u_h\|_{\mathbb{H}} = \inf_{u \in \mathbb{U}} \|h - u\|_{\mathbb{H}}$. Die Abbildung $\Pi_{\mathbb{U}} : \mathbb{H} \rightarrow \mathbb{U}$ mit $h \mapsto \Pi_{\mathbb{U}}(h) := u_h$ heißt *orthogonale Projektion*. Bezeichnen wir mit \mathbb{U}^\perp das orthogonale Komplement von \mathbb{U} in \mathbb{H} , so gilt für alle $h, g \in \mathbb{H}$:

- (i) $\Pi_{\mathbb{U}} \circ \Pi_{\mathbb{U}} = \Pi_{\mathbb{U}}$, (Projektion)
- (ii) $h - \Pi_{\mathbb{U}}(h) \in \mathbb{U}^\perp$, (orthogonal)
- (iii) $h = \Pi_{\mathbb{U}}(h) + (h - \Pi_{\mathbb{U}}(h))$ die eindeutig bestimmte Zerlegung von h in eine orthogonale direkte Summe von Elementen aus \mathbb{U} und \mathbb{U}^\perp ,
- (iv) $\Pi_{\mathbb{U}}$ ist linear und selbst-adjungiert, also $\langle \Pi_{\mathbb{U}}(h), g \rangle_{\mathbb{H}} = \langle h, \Pi_{\mathbb{U}}(g) \rangle_{\mathbb{H}}$. □

C03.19 **Eigenschaft.** Für jede Teil- σ -Algebra $\mathcal{F} \subseteq \mathcal{A}$ ist $\mathcal{L}_2(\mathcal{F}, \mathbb{P})$ ein abgeschlossener linearer Unterraum von $\mathcal{L}_2(\mathcal{A}, \mathbb{P})$. □

C03.20 **Eigenschaft.** Sei $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra und $\mathbb{E}(\cdot | \mathcal{F})$ eine Festlegung der bedingten Erwartung. Dann ist $\mathbb{E}(\cdot | \mathcal{F}) : \mathcal{L}_2(\mathcal{A}, \mathbb{P}) \rightarrow \mathcal{L}_2(\mathcal{F}, \mathbb{P})$ eine orthogonale Projektion, also für alle $X \in \mathcal{L}_2(\mathcal{A}, \mathbb{P})$ und $Y \in \mathcal{L}_2(\mathcal{F}, \mathbb{P})$ gilt

$$\|X - Y\|_{\mathcal{L}_2}^2 = \mathbb{E}(|X - Y|^2) \geq \mathbb{E}(|X - \mathbb{E}(X | \mathcal{F})|^2) = \|X - \mathbb{E}(X | \mathcal{F})\|_{\mathcal{L}_2}^2,$$

wobei Gleichheit genau dann gilt, wenn $Y = \mathbb{E}(X | \mathcal{F})$ \mathbb{P} -f.s.. □

C03.21 **Anmerkung.** Für jede Statistik $S : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ und $\mathcal{F} = \sigma(S)$ ist $\mathbb{E}(X|\sigma(S)) = \varphi(S)$ nach **Eigenschaft** C03.20 in $\mathcal{L}_2(\sigma(S), \mathbb{P})$ die beste Vorhersage von X , da für alle $h \in \mathcal{L}_2(\mathcal{S}, \mathbb{P}^S)$ gilt $\mathbb{E}|X - \mathbb{E}(X|\sigma(S))|^2 \leq \mathbb{E}|X - h(S)|^2$. Nach **Eigenschaft** A04.12 ist $Z^* = \mathbb{E}(X) + \text{Cov}(X, S)(\text{Cov}(S))^{-1}(S - \mathbb{E}(S))$ für Zufallsvektoren X und S in \mathcal{L}_2 die beste lineare Vorhersage von X durch S ist, also für alle A und b gilt $\mathbb{E}|X - Z^*|^2 \leq \mathbb{E}|X - (AS + b)|^2$ (**Definition** A04.10). Offensichtlich ist eine beste Vorhersage, die auch linear ist, eine beste lineare Vorhersage, aber die Umkehrung gilt im Allgemeinen nicht. \square

C03.22 **Beispiel.**

(a) Seien X und S wie in **Beispiel** C01.06 (b) gemeinsam normal-verteilt. Dann stimmt die beste Vorhersage $\mathbb{E}(X|\sigma(S)) = \mu_X + \Sigma_{XS}\Sigma_S^{-1}(S - \mu_S)$ mit der besten linearen Vorhersage in **Eigenschaft** A04.12 überein.

(b) Seien U und S unabhängig und identisch $N_{(0,1)}$ -verteilt, und $X := S^2 + S + U$. Dann ist für die beste Vorhersage $\mathbb{E}(X|\sigma(S)) = S^2 + S$ (benutze **Beispiel** C03.10 (a)) der mittlere quadratische Vorhersagefehler $\mathbb{E}|X - \mathbb{E}(X|\sigma(S))|^2 = \mathbb{E}(U)^2 = 1$. Die beste lineare Vorhersage von X durch S ist in diesem Fall $Z^* = S + 1$, da $\text{Cov}(X, S) = 1$, $\mathbb{E}(X) = 1$, $\text{Var}(S) = 1$ und $\mathbb{E}(S) = 0$ ist. Damit gilt $\mathbb{E}|X - Z^*|^2 = \mathbb{E}(S^4) + \mathbb{E}(U^2) - 2\mathbb{E}(S^2) + 1 = 3 > 1 = \mathbb{E}|X - \mathbb{E}(X|\sigma(S))|^2$. \square

C03.23 **Eigenschaft.** Sei $p \in [1, \infty]$, $\mathcal{F} \subseteq \mathcal{A}$ eine Teil- σ -Algebra und $\mathbb{E}(\cdot | \mathcal{F})$ eine Festlegung der bedingten Erwartung. Dann ist die lineare Abbildung $\mathbb{E}(\cdot | \mathcal{F}) : \mathcal{L}_p(\mathcal{A}, \mathbb{P}) \rightarrow \mathcal{L}_p(\mathcal{F}, \mathbb{P})$ eine Kontraktion, also $\|\mathbb{E}(X | \mathcal{F})\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_p}$, und damit beschränkt und stetig. Folglich für jede in $\mathcal{L}_p(\mathcal{A}, \mathbb{P})$ konvergente Folge $(X_n)_{n \in \mathbb{N}}$ ist $(\mathbb{E}(X_n | \mathcal{F}))_{n \in \mathbb{N}}$ ebenso in $\mathcal{L}_p(\mathcal{F}, \mathbb{P})$ konvergent. \square

C03.24 **Erinnerung.** Eine Familie $(X_i)_{i \in \mathcal{I}}$ von Zufallsvariablen in $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$ mit beliebiger nicht-leerer Indexmenge \mathcal{I} , für die gilt

$$\inf \left\{ \sup_{i \in \mathcal{I}} \mathbb{P}(|X_i| \mathbb{1}_{\{|X_i| \geq a\}}) : a \in \mathbb{R}^+ \right\} = 0.$$

heißt *gleichgradig integrierbar* (**Definition** B03.03). \square

C03.25 **Eigenschaft.** Für beliebige nicht-leere Indexmengen \mathcal{I} und \mathcal{J} seien $(X_i)_{i \in \mathcal{I}}$ eine gleichgradig integrierbare Familie von Zufallsvariablen in $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$ und $(\mathcal{F}_j)_{j \in \mathcal{J}}$ eine Familie von Teil- σ -Algebren von \mathcal{A} . Dann ist die Familie $(\mathbb{E}(X_i | \mathcal{F}_j))_{i \in \mathcal{I}, j \in \mathcal{J}}$ gleichgradig integrierbar in $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$. Insbesondere für $X \in \mathcal{L}_1(\mathcal{A}, \mathbb{P})$ und $\mathcal{J} = \{\mathcal{F} : \mathcal{F} \text{ ist Teil-}\sigma\text{-Algebra von } \mathcal{A}\}$ ist die Familie $(\mathbb{E}(X | \mathcal{F}))_{\mathcal{F} \in \mathcal{J}}$ in $\mathcal{L}_1(\mathcal{A}, \mathbb{P})$ gleichgradig integrierbar. \square

Literaturverzeichnis

- H. Bauer. *Maß- und Integrationstheorie*. Berlin etc.: Walter de Gruyter, 2., überarbeitete Auflage, 1992.
- F. Comte. *Estimation non-paramétrique*. Spartacus-idh, Paris, 2015.
- J. Elstrodt. *Maß- und Integrationstheorie*. Berlin: Springer, 7., überarbeitete und ergänzte Auflage, 2011.
- H.-O. Georgii. *Stochastik*. De Gruyter, 5. Auflage, 2015.
- A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Spektrum, 3., überarbeitete und ergänzte Auflage, 2012.
- A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Spektrum, 4., überarbeitete und ergänzte Auflage, 2020.
- U. Krenzel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig: Vieweg, 8., erweiterte Auflage, 2005.
- E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York., 1998.
- 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.
- E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <http://www.R-project.org/>.
- M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.
- J. Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer New York, 2003.
- A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. New York, NY: Springer. xii, 214 p., 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.

Index

Kleinste-Quadrate-Schätzer

- gewöhnlicher (gKQS), 5
- verallgemeinerter (vKQS), 5

Lineares Modell

Lokations-Skalen-Modell

- normal $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, 63

Lineares Modell

- gewöhnliches, 7, 16

$$IM_{X \mathbb{R}^p \times \mathbb{R}_{\leq}^{(n,n)}}, 3$$

Lokations-Skalen-Modell

- normal $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, 65

gewöhnliches

$$IM_{X \mathbb{R}^p \times \mathbb{R}_0^+}, 3$$

- normales $N_{X \mathbb{R}^k \times \mathbb{R}_0^+}$, 35, 36

Lokations-Modell

- normal $N_{\mathbb{R} \times \{\sigma_0^2\}}^n$, 40, 42, 53, 57–59

Lokations-Skalen-Modell

- normal $N_{\mathbb{R} \times \mathbb{R}_0^+}^n$, 37, 46, 64

Lokations-Skalen-Modell

$$IM_{\mathbb{1}_n \mathbb{R} \times \mathbb{R}_0^+}, 3$$

Lokations-Modell

$$IM_{\mathbb{1}_n \mathbb{R} \times \{\sigma_0^2\}}, 3$$

Skalen-Modell

$$IM_{\{\mu_0 \mathbb{1}_n\} \times \mathbb{R}_0^+}, 3$$

Nonparametric

- regression, 74
- density estimation, 73, 74
- regression, 78

Verteilung

- Bernoulli B_p , 5
- Binomial $\text{Bin}_{(n,\pi)}$, 57
- Exponential Exp_λ , 43
- Exponential Exp_λ , 38, 44
- Multinomial $\text{MN}_{(n,p)}$, 67
- Poisson Poi_λ , 44
- Poisson Poi_λ , 37, 45, 61
- Uniform $U_{[0,\theta]}$, 35, 36
- Uniform $U_{[\theta - \frac{1}{2b}, \theta + \frac{1}{2b}]}$, 42