Ruprecht-Karls-Universität Heidelberg

Institute of Applied Mathematics

Prof. Dr. Jan Johannes, Sandra Schluttenhofer

Seminar: Nonparametric Statistics

Preliminary Discussion: 15.10.2019 in Seminar Room 4.414 (Mathematikon), 14:15.

Date: The seminar will be held as a *Blockseminar* on two days at the end of November and the beginning of December.

Field: Applied Mathematics, Stochastics

Language: The seminar will be in English, if there is at least one non-German speaking participant. Otherwise the presentations may be in German.

Requirements: The seminar is for advanced Bachelor students and Master students who want to specialize in statistics and are already familiar with the topics typically covered in the lecture *Introduction to Probability and Statistics*. Knowledge from the lectures *Probability Theory I* and *Statistics I* is useful, but not a prerequisite.

Description of the Seminar:
This seminar introduces the asymptotic theory of nonparametric statistics. A typical problem of nonparametric statistics is the estimation of a function that is assumed to lie in a function class. Examples are Hölder classes and Sobolev balls. The topic of the seminar is the asymptotic theory of optimal estimation in such settings. We study the performance of estimators and prove lower bounds for minimax risks that are available in the model. A main part of the seminar will be different approaches to obtain such lower bounds. The seminar follows the book:

Alexandre B. Tsybakov: Introduction to Nonparametric Statistics. Springer 2009.

which is availiable at `https://ins.sjtu.edu.cn/files/common/20121209191850_7.pdf`

Each participant is expected to give a 60 minutes talk using the blackboard. A handout containing the most important definitions and results as well as short sketches of the proofs should be prepared for the other participants.

 Contact: schluttenhofer@math.uni-heidelberg.de

Ruprecht-Karls-Universität Heidelberg

Institute of Applied Mathematics

Prof. Dr. Jan Johannes, Sandra Schluttenhofer

Topics

Part I: Upper Bounds

Density Estimation - Upper Bound

- **Kernel Density Estimators**
  - What is a Kernel Density Estimator? (Sec. 1.2)
  - Mean Squared Risk of a Kernel Density Estimator (Sec. 1.2.1)
  - Integrated Squared Risk of a Kernel Density Estimator (Sec. 1.2.3)
- **Fourier Analysis of KDEs - Which Kernel should we choose?**
  - Tools From Fourier Analysis (Sec. 1.3)
  - Admissable/Inadmissable Kernels (Sec. 1.3)
- **Bandwidth Selection for KDEs - Which bandwidth should we choose?**
  - Cross-Validation (Sec. 1.4)
  - Goldenshluger-Lepski (Sec. 4.2. in *Nonparametric Estimation* (F. Comte))

Regression Estimation - Upper Bound

- **Local Polynomial Estimators**
  - The Nadaraya-Watson Estimator (Sec. 1.5)
  - Local Polynomial Estimators (Sec. 1.6)
  - Mean Squared and Integrated Squared Risk of Local Polynomial Estimators (Sec. 1.6.1)
- **Projection Estimators**
  - Projection Estimators (Sec. 1.7)
  - Sobolev Classes (Sec. 1.7.1)
  - Integrated Squared Risk of Projection Estimators (Sec. 1.7.2)

Contact: schluttenhofer@math.uni-heidelberg.de

Prof. Dr. Jan Johannes, Sandra Schluttenhofer

Topics

Part II: Lower Bounds

- **Lower Bounds based on two hypotheses I - Pointwise Risk**
  - General Reduction Scheme (Sec. 2.2)
  - Lower Bounds based on two hypotheses (Sec. 2.3)
- **Lower Bounds based on two hypotheses II - Pointwise Risk**
  - Tools: Distances between probability measures (Sec. 2.4, only Kullback-divergence)
  - **Either:** application of the method to regression estimation (Sec. 2.5)
  - **Or:** application of the method to density estimation (Ex. 2.8)
- **Lower Bounds based on many hypotheses I - Global Risk**
  - Why two hypotheses are not enough (Sec. 2.6)
  - Lower Bounds based on many hypotheses (Sec. 2.6)
- **Lower Bounds based on many hypotheses II - Global Risk**
  - Tools: Hoeffding's Inequality, Varshamov-Gilbert Bound
  - **Either:** application of the method to regression estimation (Sec. 2.6.1)
  - **Or:** application of the method to density estimation (Ex. 2.10)
- **Lower Bounds based on Fano's Method**
  - Fano's Lemma (Sec. 2.7.1)
  - Application of the method to regression estimation (Thm. 2.11)
- **Lower Bounds based on Assouad's Cube Method**
  - Assouad's Lemma (Sec. 2.7.2)
  - Application of the method to regression estimation (Example 2.2)

Contact: schluttenhofer@math.uni-heidelberg.de

Seminar: Nonparametric Statistics
Sandra Schluttenhofer
Wintersemester 2019/20

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Overview of the Seminar

	Density Estimation	**Regression Estimation**
observations	$X_1, \dots, X_n \overset{\text{iid}}{\sim} p = \frac{dP}{d\lambda}$	$(X_1, Y_1), \dots, (X_n, Y_n)$ with $Y_i = f(X_i) + \xi_i$ ξ_i iid with $\mathbb{E}(\xi_i) = 0$
goal	Estimate p.	Estimate f.
(known) typical (parametric) assumption	e.g. $p(x) = \lambda \exp(-\lambda x) \mathbb{I}_{(0,\infty)}(x)$	e.g. $f(x) = ax + b$
	we only need to estimate the parameters λ resp. (a, b)!	
(new) typical (nonparametric) assumption	$p \in \mathcal{P}$ e.g. Lipschitz continuos densities	$f \in \mathcal{F}$ e.g. continuous functions
	1st Goal of the Seminar: Learn about Nonparametric Estimation Methods. Construct estimators $\hat{p}$ and $\hat{f}$.	
estimation methods	kernel density estimation	local polynomial estimation projection estimation
	2nd Goal of the Seminar: Learn how to derive Upper Bounds for the Risk of an Estimator. How good are the estimators?	
mean squared risk	$\sup_{p \in \mathcal{P}} \mathbb{E} \left(\hat{p}(x) - p(x)\right)^2 \leq$ **?**	$\sup_{f \in \mathcal{F}} \mathbb{E} \left(\hat{f}(x) - f(x)\right)^2 \leq$ **?**
integrated mean squared risk	$\sup_{p \in \mathcal{P}} \mathbb{E} \int \left(\hat{p}(x) - p(x)\right)^2 dx \leq$ **?**	$\sup_{f \in \mathcal{F}} \mathbb{E} \int \left(\hat{f}(x) - f(x)\right)^2 dx \leq$ **?**
	3rd Goal of the Seminar: Learn how to derive Lower Bounds for the Minimax Risk. We want to find "the best" estimator!	
mean squared risk	$\inf_{\hat{p}} \sup_{p \in \mathcal{P}} \mathbb{E} \left(\hat{p}(x) - p(x)\right)^2 \geq$ **?**	$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left(\hat{f}(x) - f(x)\right)^2 \geq$ **?**
integrated mean squared risk	$\inf_{\hat{p}} \sup_{p \in \mathcal{P}} \mathbb{E} \int \left(\hat{p}(x) - p(x)\right)^2 dx \geq$ **?**	$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \int \left(\hat{f}(x) - f(x)\right)^2 dx \geq$ **?**