Seminar on Empirical Processes

Summer Semester 2024

Dr. Housen Li and Prof. Dr. Axel Munk

Key information

Time:	12/04/2024 – 12/07/2024, on Fridays, 10:15–12:00
Location:	Seminar Room 5.101 (IMS)
Possible Modules:	B.Mat.3444: Seminar on mathematical statistics
	M.Mat.4844: Seminar on mathematical statistics
	B.Mat.3447: Seminar on statistical foundations of data science
	M.Mat.4847: Seminar on statistical foundations of data science
Instructors:	Dr. Housen Li
Intended Audience:	**Advanced Bachelor and beginning Master students**
Language:	English

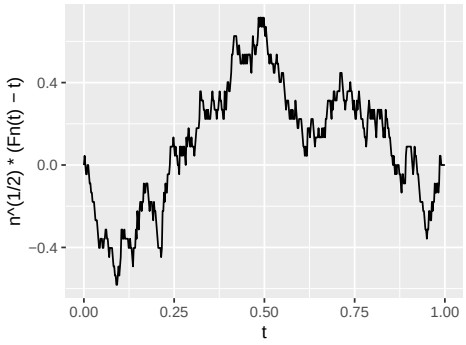
Prerequisites

Participants must have successfully attended

- Measure & probability theory (B.Mat.1400).

It is further considered helpful to have attended

- Stochastics (B.Mat.2410), and
- Statistical data science (B.Mat.2420), and/or
- Statistical foundations of data science (B.Mat.3147).



Description

Modern data do not just occur as vectors, but rather as functions or even more generally, as random objects taking values in metric spaces. Many important statistical quantities can be written as functionals of such data. A simple but fundamental example is the sample mean of real valued data, which is a simple functional (an integral) of the empirical cumulative distribution function (e.c.d.f.). After centring and rescaling, this leads to the empirical process (see the plot), a stochastic process. One generic approach to studying the properties of such a statistic is to separately investigate the properties of the functional (mainly using functional analytic tools) and of the underlying empirical process. Classical Glivenko–Cantelli and Donsker theorems establish almost sure uniform and weak convergence of the empirical process, respectively. As a consequence, properties of the sample mean and also of more complex functionals of the empirical process can be derived.

The modern empirical process approach is to view the empirical process resulting from the e.c.d.f. as a process indexed in indicator functions on intervals $(-\infty, t]$ rather than scalars t. This allows remarkable generalisations to empirical processes indexed in function spaces. The developed theory provides powerful techniques that can be employed to understand the properties of modern statistical methods (e.g. bootstrap) in a broad range of scenarios. These techniques are fundamental to statistical and machine learning theory and have manifold applications nowadays, e.g. in nonparametric regression and the theory of statistical optimal transport.

In this seminar we cover the basic relevant mathematical concepts and main results of empirical process theory. Our focus will be on understanding the fundamental principles and techniques rather than obtaining results in most generality. Topics include: concentration of measure, maximal and isoperimetric inequalities, metric entropy and Vapnik–Chervonenkis dimension, Talagrand's (generic) chaining, central limit theorems for processes in (non-)separable metric spaces, and central limit theorems in high dimensions. As application of the theory, we consider M-estimators, empirical risk minimisation, and bootstrap.

Application

To provide participants with the material to be presented at an early stage, we ask you to preregister for this seminar. To this end, please email Housen Li (housen.li@mathematik.uni-goettingen.de) and indicate your interest to give a seminar talk. Please include information about relevant courses you have taken in your email. Deadline for preregistration is 8$^{\text{th}}$ March 2024.

A preparatory meeting, during which topics will be assigned to participating students, is scheduled on 15$^{\text{th}}$ March 2024, 10:15–12:00. This meeting will be held virtually via Zoom (dialling-in information will be provided to registered students via email in due time). The seminar has a limited number of participants. In case of outnumbering, participants will be chosen based on the information provided in their preregistration email.

Recommended literature

Main Reference

Topics for presentations will be assigned along the lines of

- Sen, B. (2022): Lecture Notes "A Gentle Introduction to Empirical Process Theory and Applications", `http://www.stat.columbia.edu/~bodhi/Talks/Emp-Proc-Lecture-Notes.pdf`.

References for further reading

- Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2023). High-dimensional data bootstrap. Annual Review of Statistics and Its Application, 10, 427-449.
- Giné, E., and Nickl, R. (2021). Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge University Press.
- Shorack, G. R. and Wellner, J. A. (2009). Empirical Processes with Applications to Statistics. Society for Industrial and Applied Mathematics.
- van der Vaart, A. W. and Wellner, J. A. (2023). Weak Convergence and Empirical Processes: With Applications to Statistics. Second Edition. Springer.
- Vershynin, R. (2018). High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press.