technische universität
dortmund

Seminar "Empirische Prozesstheorie"

Seminar “Empirical processes”

Marc Ditzhaus

Lehrstuhl für mathematische Statistik und industrielle Anwendung

Dortmund, 07.07.2021

Empirical distribution function

- Starting point in 1930-1940: *empirical distribution function*.
- Let $X_1, \ldots, X_n$ be i.i.d. with distribution function F. Then Theorem I (Glivenko-Cantelli, 1933):

$$||\widehat{F}_n - F||_\infty = \sup_{-\infty<x<\infty} |\widehat{F}_n(x) - F(x)| \to 0 \text{ almost surely.}$$

Theorem II (Donsker, 1952):

$$\sqrt{n}(\widehat{F}_n - F) \xrightarrow{d} B_0 \circ F \text{ in } D(\mathbb{R}),$$

where B_0 is a Brownian bridge on $[0, 1]$ and $D(\mathbb{R})$ denotes the Skorokhod space.

Applications

- Various estimators for interesting quantities and test statistics can be written in terms of a the empirical distribution function:
  - Sample median = $\hat{F}_n^{-1}(0.5)$ (in the same way: general quantiles)
  - Sample mean = $\int x \, \mathrm{d}\hat{F}_n(x)$
  - Mann-Whitney-Test or the estimator for the concordance measure:

$$\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{1}\{X_i > Y_j\} = \int \left(1 - \hat{F}_n(y)\right) \mathrm{d}\hat{G}_n(y),$$

    where $\hat{F}_n$ and $\hat{G}_m$ are the empirical distribution functions of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, respectively.
  - and a lot more...

Application II

- In all previous situations we have

$$\widehat{\theta}_n = \Phi(\hat{F}_n) \quad \text{(one-sample)}$$
$$T_n = \Phi(\hat{F}_n, \hat{G}_n) \quad \text{(two-sample)}$$

- To obtain asymptotic confidence intervals or asymtotically valid tests, we can combine the Donsker Theorem and a functional delta-Method, e.g.

$$\sqrt{n}\Big(\Phi(\hat{F}_n) - \Phi(F)\Big) \xrightarrow{d} \Phi'_F(B_0 \circ F).$$

Possible topics

Source: van der Vaart and Wellner (1996). Weak Convergence and Empirical Processes. Springer.

- Introduction (Stochastic convergence for general spaces)
- General Donsker and Glivenko-Cantelli results for

$$\mathcal{F} \ni f \mapsto \frac{1}{n}\sum_{i=1}^{n} f(X_i) \qquad \left(\text{Recall } \mathbb{R} \ni t \mapsto \hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbf{1}\{X_i \leq t\}}_{f_t(X_i)}\right)$$

- Functional Delta-Method
- Bootstrap for one-sample settings
- Bootstrap and Permutation for k-sample settings.
- Application of the theory in recent research

Modus/dates

- Required Qualification: Knowledge about probability theory (e.g. stochastic convergence)
- 2-3 blocks during the lecture time.
- Currently planned as a digital seminar.
- First meeting in the first lecture week (to find dates etc., maybe a first introduction about basic concepts by myself)
- talk (40-60 minutes) and a short seminar paper.

Registration

- Binding registration via Email **until 24th September**:

  marc.ditzhaus@tu-dortmund.de

  please add the following information
  - Prior knowledge about probability theory (which lecture, course)
  - Bachelor/Master
  - Are you interested (later) in a Master thesis using empirical processes, then you can contact me anytime=)
- Further Information on my homepage