Lecture Empirical Processes

Jun.-Prof. Dr. Dennis Dobler
dobler@statistik.tu-dortmund.de
TU Dortmund University

General Information

Suitable study subjects (modules):
M.Sc. Statistik (MS 6/7), M.Sc. Data Science (MD E1), M.Sc. Econometrics (ME 7)

ECTS: 4.5

Language: English

Lectures: one every week

Exercises: one every second week

Exam: preferred oral, depending on number of participants

Empirical Process

- i.i.d. random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with values in $\mathbb{R}^d$
- empirical measure $\mathbb{P}_n(A) = \frac{1}{n} \sum_{i=1}^n 1_A(\boldsymbol{X}_i)$

Empirical Process

- i.i.d. random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with values in $\mathbb{R}^d$
- empirical measure $\mathbb{P}_n(A) = \frac{1}{n} \sum_{i=1}^n 1_A(\boldsymbol{X}_i)$
- for functions $g : \mathbb{R}^d \to \mathbb{R}^k$:

  empirical process $\mathbb{P}_n g = \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{X}_i)$, $g \in \mathcal{F}$

Empirical Process

- i.i.d. random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with values in $\mathbb{R}^d$
- empirical measure $\mathbb{P}_n(A) = \frac{1}{n} \sum_{i=1}^n 1_A(\boldsymbol{X}_i)$
- for functions $g : \mathbb{R}^d \to \mathbb{R}^k$:

  empirical process $\mathbb{P}_n g = \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{X}_i)$, $g \in \mathcal{F}$

$\Rightarrow$ Question (1): why?

$\Rightarrow$ Question (2): law of large numbers and central limit theorem uniformly in $g \in \mathcal{F}$?

Application

Kolmogorov-Smirnov Goodness-of-Fit Test (on $\mathbb{R}^1$)

- null hypothesis $H_0 : F = F_0$ vs. alternative hypothesis $H_a : F \neq F_0$.
- statistic: $D_n = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F_0(x)| = \sup_{x \in \mathbb{R}} |\mathbb{P}_n 1_{\{\,\cdot\, \leq x\}} - F_0(x)|$

$\Rightarrow$ or supremum over $\mathcal{F} = \{g_x : y \mapsto 1_{\{y \leq x\}} : x \in \mathbb{R}\}$

Application

Kolmogorov-Smirnov Goodness-of-Fit Test (on $\mathbb{R}^1$)

- null hypothesis $H_0 : F = F_0$ vs. alternative hypothesis $H_a : F \neq F_0$.
- statistic: $D_n = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F_0(x)| = \sup_{x \in \mathbb{R}} |\mathbb{P}_n 1_{\{\,\cdot\, \leq x\}} - F_0(x)|$

$\Rightarrow$ or supremum over $\mathcal{F} = \{g_x : y \mapsto 1_{\{y \leq x\}} : x \in \mathbb{R}\}$

Maximum Likelihood Estimator (on $\mathbb{R}^1$)

- $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F_\eta$, density f_η, $\eta \in \mathbb{R}$ unknown
- log-likelihood $\ell(\eta; X_1, \ldots, X_n) = \sum_{i=1}^n \log f_\eta(X_i) = n \cdot \mathbb{P}_n \log f_\eta$

$\Rightarrow$ $\mathcal{F} = \{g_\eta = \log f_\eta : \eta \in \mathbb{R}\}$

Lecture Empirical Processes

Aims (selection):

(1) control $\mathbb{P}_n$ over $\mathcal{F}$ (uniform LLN & CLT)

(2) suitable delta-method

Lecture Empirical Processes

Aims (selection):

(1) control $\mathbb{P}_n$ over $\mathcal{F}$ (uniform LLN & CLT)
(2) suitable delta-method
(3) various important applications
(4) resampling

Lecture Empirical Processes

Aims (selection):

(1) control $\mathbb{P}_n$ over $\mathcal{F}$ (uniform LLN & CLT)
(2) suitable delta-method
(3) various important applications
(4) resampling

Challenges (selection):

(1) control size of $\mathcal{F}$
(2) use framework without measurability

Lecture Empirical Processes

Aims (selection):

(1) control $\mathbb{P}_n$ over $\mathcal{F}$ (uniform LLN & CLT)

(2) suitable delta-method

(3) various important applications

(4) resampling

Challenges (selection):

(1) control size of $\mathcal{F}$

(2) use framework without measurability

References (selection):

- AW van der Vaart & JA Wellner (2023). Weak Convergence and Empirical Processes: With Applications to Statistics, 2nd Edition, Springer.
- MR Kosorok (2008). Introduction to empirical processes and semiparametric inference, Springer.