Statistical Data Analysis

Dr. Jana de Wiljes

10. November 2021

Universität Potsdam

Kolmogorov-distance

Def: The Kolmogorov-distance between the empirical cdf $\widehat{F}_n(t)$ and the theoretical cdf F is defined as follows

$$D_n := \sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F(t)| \tag{1}$$

Theorem of Gliwenko-Cantelli

Theorem: For the Kolmogorov-distance D_n the following holds

$$D_n \to 0 \text{ for } n \to \infty \text{ almost everywhere} \tag{2}$$

i.e.,

$$\mathbb{P}\left[\lim_{n\to} D_n = 0\right] = 1 \tag{3}$$

Proof

Proof

Proof

Proof

A statistical model

Def: A statistical model is a triple $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ where

- $\mathcal{X}$ is the sample space
- $\mathcal{A} \subset 2^{\mathcal{X}}$ is a σ-algebra on $\mathcal{X}$
- Θ is the parameter space
- for every $\theta \in \Theta$ $\mathbb{P}_\theta$ is a probability measure on $(\mathcal{X}, \mathcal{A})$

Estimator

Def: Let $\Theta \subset \mathbb{R}^r$. A estimator is a measurable map

$$\widehat{\theta} : \mathcal{X} \to \Theta, \quad x \mapsto \widehat{\theta}(x) \tag{4}$$

Example

Example

Maximum-Likelihood estimator

Def: The Maximum-Likelihood estimator is defined via

$$\widehat{\theta}_{\mathsf{ML}} = \arg\max_{\theta \in \Theta} L(\theta) \tag{5}$$

Abbildung 1: Daniel Bernoulli, Joseph-Louis Lagrange, Carl-Friedrich Gau and Ronald Fisher

Example

Example

A-posteriori-distribution

Def: The a-posteriori-distribution of θ is the conditional distribution given the information $X_1 = x_1, \dots, X_n = x_n$, i.e.,

$$q(\theta_i | x_1, \dots, x_n) := \mathbb{P}[\theta = \theta_i | X_1 = x_1, \dots, X_n = x_n], \quad i = 1, 2, \dots$$

Bayes method

Def: The Bayes estimator is defined as the expectation of the a-posteriori-distribution

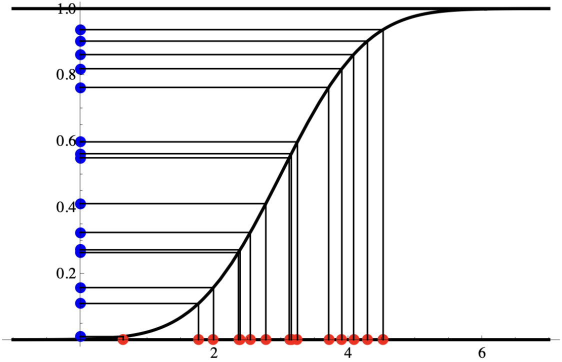
$$\widehat{\theta}_{\text{Bayes}} = \sum_i \theta_i q(\theta_i | x_1, \ldots, x_n)$$

Example

Example

Maximum-spacing method

Lemma: Let the cdf F_θ be continuous and strictly monoton increasing. Under $\mathbb{P}_\theta$ the random variables $F_\theta(X_1), \ldots, F_\theta(X_n)$ are independent and uniformly distributed on the $(0,1)$ interval.



Proof

Maximum-spacing method

Lemma: Let $z_1, \ldots, z_k \in [0, 1]$ be number that are subject to the condition $z_1 + \cdots + z_k = 1$. Then

$$z_1 \cdot \cdots \cdot z_k \leq \frac{1}{k^k}. \tag{6}$$

Equality is attained only if all the numbers all the numbers are equal to $\frac{1}{k}$.

Proof

Maximum-spacing method

Lemma: The maximum-spacing method is defined via

$$\widehat{\theta}_{MS} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n+1} \left(F_\theta(x_{(i)}) - F_\theta(x_{(i-1)})\right) \tag{7}$$

Example

Example