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

Note: We will show the theorem only for special case of cdf F being continuous

Proof: Let F be continuous and further $m \in \mathbb{N}$

Step 1: Since F is continuous and is monotonly increasing from 0 to 1 we can find real numbers z_k

$$z_0 \quad z_1 < z_2 < \dots z_{m-1} \quad \leftarrow \quad z_m$$

with $F(z_1) = 1/m, \dots, F(z_k) = k/m, \dots, F(z_{m-1}) = \frac{m-1}{m}$

$1/m$

Further we set $z_0 = -\infty$ and $z_m = +\infty$ and denote

$$F(z_0) = 0 \quad \text{and} \quad F(z_m) = 1$$

z_k	z	z_{k+1}
1	5	8
$F(z_k)$	$F(z)$	$F(z_{k+1})$
5	6	9
$\hat{F}_n(z_k)$	$\hat{F}_n(z)$	$\hat{F}(z_{k+1})$
7	8	9

$8 - 6 = 2$

$9 - 9 = 0$

$8 - 6 \geq 9 - 9$

$2 \geq 0$

Step 2: The plan is to now approximate the differences between $\hat{F}_n(z)$ and $F(z)$ at an arbitrary z by the differences at the values z_k. For every $z \in \mathbb{R}$ it is possible to find a k with $[z_k, z_{k+1}]$. Due to the fact that $\hat{F}_n$ and F are monoton the following holds

$$[\hat{F}_n(z) - F(z)] \leq \hat{F}_n(z_{k+1}) - F(z_k) = \hat{F}_n(z_{k+1}) - F(z_{k+1}) + \frac{1}{m}$$

Proof

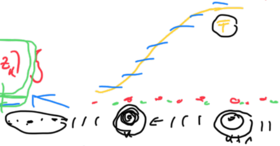
On the other hand we have $[z_k, z_{k+1}]$

$$\hat{F}_n(z) - F(z) \geq \hat{F}_n(z_k) - F(z_{k+1}) \geq \hat{F}_n(z_k) - F(z_k) - \boxed{\frac{1}{m}}$$

$\geq$ $\leq$

Step 3 Define $m \in \mathbb{N}$ and $k = 0, \dots, m$ the event

$$A_{m,k} := \{\omega \in \Omega : \lim_{n \to \infty} \hat{F}_n(z_k, \omega) = F(z_k)\}$$



Note that $\hat{F}_n(z_k, \omega)$ is a random variable and therefore dependent on the outcome $\omega \in \Omega$

$\mathbb{P}[A_{k,m}] = 1$ $\quad \forall m \in \mathbb{N},\ k = 0, \dots, m$ This we know from part 3 of the proposition from yesterday - as we showed pointwise convergence almost everywhere

Step 4: Define $A_m := \bigcap_{k=0}^{m} A_{m,k}$ $\Rightarrow$ (basic rules of probability theory) $\mathbb{P}[A_m] = 1$ $\quad \forall m \in \mathbb{N}$

$$A := \bigcap_{m=1}^{\infty} A_m \Rightarrow \mathbb{P}[A] = 1$$

Proof

Step 5: Consider an abitrary outcome $\omega \in A_m$. Then there ex. a $n(\omega, m) \in \mathbb{N}$ (due to the definition of $A_{m,k}$ with the property

$$\rightarrow |\vec{F}_n(z_k, \omega) - F(z_k)| < 1/m \quad \forall n > n(\omega, m) \text{ and } k = 0, \dots, m$$

From step 2 it follows that

$$\rightarrow D_n(\omega) = \sup_{z \in \mathbb{R}} \boxed{|\vec{F}_n(z, \omega) - F(z)|} \leq \frac{2}{m} \quad \text{(for all } \omega \in A_m \text{ and } n > n(\omega, m))$$

$\boxed{\text{For an abitray } \omega \in A \Rightarrow \omega \in A_m \ \forall m \in \mathbb{N}}$

Therefore we can rephrase the conclusion above to

For all $m \in \mathbb{N}$ $\exists\, n(\omega, m) \in \mathbb{N}$ so that $n > n(\omega, m)$ the inequality

$\boxed{0 \leq D_n(\omega) \leq 2/m}$ holds

That the means: $\lim\limits_{n \to \infty} D_n(\omega) = 0 \quad \forall \omega \in A$

Since the probability for A to take place is 1

we attain $\mathbb{P}[\{\omega \in \Omega : \boxed{\lim\limits_{n \to \infty} P_n(\omega) = 0}\}] \geq \boxed{\mathbb{P}[A] = 1} \Rightarrow D_n \xrightarrow[n \to \infty]{a.s.} 0$ □

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 $(\mathcal{X}, \mathcal{A})$

$X_1, \ldots, X_n$

$\mathcal{X} = \mathbb{R}^n$

$X_i(\omega) \in \mathbb{R}$

$(x_1, \ldots, x_n) \in \mathbb{R}^n$

Bern(p_1) Bern(p_2)

$X: \Omega \to \mathbb{R}^n = \mathcal{X}$

$\mathbb{P}(X_i = x_i)$

$\{\omega : X_i(\omega) = x_i\} \in \mathcal{A}$

$\mathbb{P}(X_1 = x_1, \ldots X_n = x_n)$ for X_i independent

$= \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \ldots \cdot \mathbb{P}(X_n = x_n)$

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

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

θ

Example: Assuming a normal distribution and we have samples $x_1, \dots, x_n$. $\mathcal{X} = \mathbb{R}^n$

$\boxed{X_i \sim N(\mu, \sigma^2)}$ i.i.d

$$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} = \boxed{\mathbb{R} \times (0, \infty)}$$

$\rightsquigarrow (\mu, \sigma^2) \in \Theta \rightsquigarrow \mathbb{P}_{\mu, \sigma^2}$ (Choice of μ, σ^2 leads to different probability measure)

$$\boxed{L(x_1, \dots, x_n, \theta) := \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}}$$

Example

Example

Maximum-Likelihood estimator

Def: The Maximum-Likelihood estimator is defined via

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

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

Example

$$L(x,\theta) = \boxed{\mathbb{P}_\theta \; [X = x]} \qquad X : \Omega \to \mathcal{X}$$

$$X_1, \dots, X_n \quad \text{i.i.d} \qquad \boxed{h_\theta(t)} = \boxed{\mathbb{P}_\theta \; [X_i = t]}$$

$$\boxed{L(x_1, \dots, x_n; \theta)} = \boxed{\mathbb{P}_\theta} [X_1 = x_1, \dots, X_n = x_n] = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n)$$

Ex: $X_i \sim \mathrm{Bern}(\theta)$ iid.

$$\boxed{L(x_1, \dots, x_n, \theta)} = \mathbb{P}_\theta [X_1 = x_1] \cdot \mathbb{P}[X_2 = x_2] \dots \mathbb{P}[X_n = x_n] = \theta^s (1-\theta)^{n-s}$$

s = number of success

$\boxed{s = 0, \quad s = n}$

$$\frac{dL(\dots, \theta)}{d\theta} = s\theta^{s-1}(1-\theta)^{n-s} - (n-s)\theta^s(1-\theta)^{n-s-1}$$

$$\frac{dL(\dots)}{d\theta} \overset{!}{=} 0 \quad \leadsto \quad \theta = \frac{s}{n} \quad \leadsto \quad \theta_{ML} = \boxed{\frac{s}{n}}$$

Example

A-posteriori-distribution

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

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

Bayes method

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

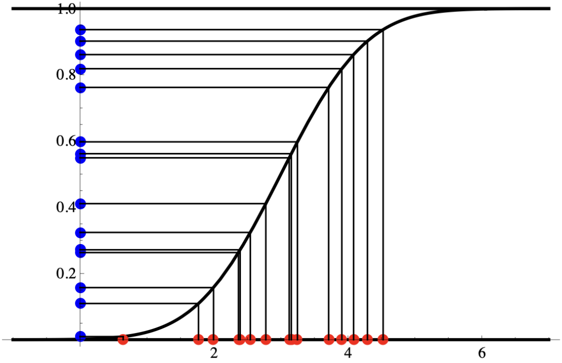
$$\widehat{\theta}_{\text{Bayes}} = \sum_i \theta_i q(\theta_i | x_1, \ldots, x_n)$$

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

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

Example

Example