NMAI059 Probability and statistics 1
Class 11

Robert Šámal

Overview

Statistics – modelling

Statistics – point estimation

Random sample

- without repetition
  $\Omega = \{$all n-tuples of citizens of Czechia$\}$
  For $\omega = (\omega_1, \dots, \omega_n)$ we put $X_i = I(\omega_i$ is left-handed$)$.
- with repetition
  $\Omega = \{$all n-tuples of citizens of Czechia, with repetition$\}$
  For $\omega = (\omega_1, \dots, \omega_n)$ we put $X_i = I(\omega_i$ is left-handed$)$.
- variants (stratified sample)
  We want to proportionally represent various subsets (age, education, home address, etc.).
  Not studied further in this course.

Statistika – model

- independent measurements – using i.i.d. $X_1, \ldots, X_n \sim F$ *random sample* from CDF F
- nonparametric models: large class of F
- parametric models: $F \in \{F_\vartheta : \vartheta \in \Theta\}$
- examples
  - $Pois(\lambda)$ (parameter $\vartheta = \lambda$, $\Theta = \mathbb{R}^+$)
  - $U(a, b)$ (parameter $\vartheta = (a, b)$, $\Theta = \mathbb{R}^2$)
  - $N(\mu, \sigma^2)$ (parameter $\vartheta = (\mu, \sigma)$, $\Theta = \mathbb{R} \times \mathbb{R}^+$)
- "All models are wrong, but some are useful." (George Box)

Confirmatory data analysis

- point estimates
- interval estimates
- hypothesis testing
- (linear) regression

- *statistics* – any function of a random sample, e.g., arithmetic mean, median, maximum, etc. That is $T = T(X_1, \ldots, X_n)$.

Overview

Statistics – modelling

Statistics – point estimation

Sample mean and variance

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$\widehat{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Estimator

Definition

Estimator is an arbitrary statistics.

Properties of estimators

Definition

Estimator $T_n = T_n(X_1, \ldots, X_n)$ *of* $g(\vartheta)$ *je*

- *unbiased – pokud* $g(\vartheta) = \mathbb{E}\big(T_n\big)$ *(for each* ϑ*)*
- *asymptotically unbiased*
  – if $g(\vartheta) = \lim\limits_{n\to\infty} \mathbb{E}\big(T_n\big)$
- *consistent – if* $T_n \xrightarrow{P} \vartheta$*.*
- *bias is defined as* $bias_\vartheta := \mathbb{E}\big(T_n\big) - g(\vartheta)$
- *mean squared error, MSE is* $MSE := \mathbb{E}\big((T - g(\vartheta))^2\big)$

Theorem

$MSE = bias_\vartheta^2 + var_\vartheta(T_n)$

Properties of estimators

Properties of estimators

Properties of sample mean and variance

Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value μ and variance σ^2.

1. *$\bar{X}_n$ is a consistent unbiased estimator of μ*
2. *$\bar{S}_n$ is a consistent asymptotically unbiased estimator of σ^2*
3. *$\widehat{S}_n$ is a consistent unbiased estimator of σ^2*

Method of moments

- $m_r(\vartheta) := \mathbb{E}\big(X^r\big)$ for $X \sim F_\vartheta$ ... r-th moment
- $\widehat{m_r(\vartheta)} := \frac{1}{n}\sum_{i=1}^n X_i^r$ for a random sample $X_1, \dots, X_n$ z F_ϑ ... r-th sample moment

Theorem

$\widehat{m_r(\vartheta)}$ *is unbiased consistent estimator of* $m_r(\vartheta)$

- Estimator using the method of moments is obtained by solving system of equations (k is the number of parameters)

$$m_r(\vartheta) = \widehat{m_r(\vartheta)} \qquad r = 1, \dots, k.$$

Method of moments – examples

Maximal likelihood, ML

- random sample $X = (X_1, \ldots, X_n)$ from a distribution with parameter ϑ
- possible realization $x = (x_1, \ldots, x_n)$
- … joint pmf$p_X(x; \vartheta)$
- … joint pdf $f_X(x; \vartheta)$
- *likelihood* $L(x; \vartheta)$ denotes p_X or f_X
- before: we have a fixed ϑ, study $L(x; \vartheta)$ as a function of x
- now: we have a fixed x and study $L(x; \vartheta)$ as a function of ϑ

Maximal Likelihood principle
choose ϑ that maximizes $L(x; \vartheta)$.

Maximal likelihood

- **Metoda MV (ML):**
  choose ϑ that maximizes $L(x;\vartheta)$.
- for convenience we put $\ell(x;\vartheta) = \log(L(x;\vartheta))$
- by independence of X_1, X_2, etc. we have

  $L(x;\vartheta) =$

  $\ell(x;\vartheta) =$

ML example – proportion of left-handed