NMAI059 Probability and statistics 1
Class 11

Robert Šámal

Overview

Statistics – modelling

Statistics – point estimation

Random sample

W_i = weight of the i-th person $\in \mathbb{R}$
H_i = height —''— $\in \mathbb{R}$
S_i = salary —''— $\in \mathbb{R}$

- without repetition
  $\Omega = \{$all n-tuples of citizens of Czechia$\}$
  For $\omega = (\omega_1, \dots, \omega_n)$ we put $X_i^{(\omega)} = I(\omega_i$ 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.

ω_i = state of a computer + input to an alg.

random var.

$\omega = (\omega_1 \dots \omega_n)$

$x_i = X_i(\omega)$ $\quad \langle 0, 1$ $\quad X_i(\omega)$ = running time from state ω_i

realization of X_i ---- numbers

Statisti~~ka~~cs – model

indep. ident. distr.

- independent measurements – using i.i.d. $X_1, \dots, X_n \sim F$ r.v.
  random sample from CDF F

- nonparametric models: large class of F
  ecdf — we sample F -----
  F(0) = 0
  F(∞) = 1
- parametric models: $F \in \{F_\vartheta : \vartheta \in \Theta\}$
- examples
  - $Pois(\lambda)$ (parameter $\vartheta = \lambda$, $\Theta = \mathbb{R}^+$) # kids in a family; # email/day
  - $U(a, b)$ (parameter $\vartheta = (a, b)$, $\Theta = \mathbb{R}^2$)
  - $N(\mu, \sigma^2)$ (parameter $\vartheta = (\mu, \sigma)$, $\Theta = \mathbb{R} \times \mathbb{R}^+$) --- height
  --- running time
  Exp(λ)

- "All models are wrong, but some are useful." (George Box)

Confirmatory data analysis

- point estimates ← try to guess ϑ ↙ don't know
  or poss. some fction of ϑ $(g(\vartheta))$
- interval estimates
- hypothesis testing
- (linear) regression

$X_1, \ldots, X_n \sim F_\vartheta$

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

when we get $x_1 = X_1(\omega)$, $x_2 = X_2(\omega)$, ... realizations

we use $T(x_1, \ldots x_n)$ as our best guess.

Overview

Statistics – modelling

Statistics – point estimation

Sample mean and variance

$$\mathbb{E}X_1 = \sum_x x \cdot P(X_1 = x)$$

$$= \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

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

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

looks more sensible

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

has better properties,

gives truer answers

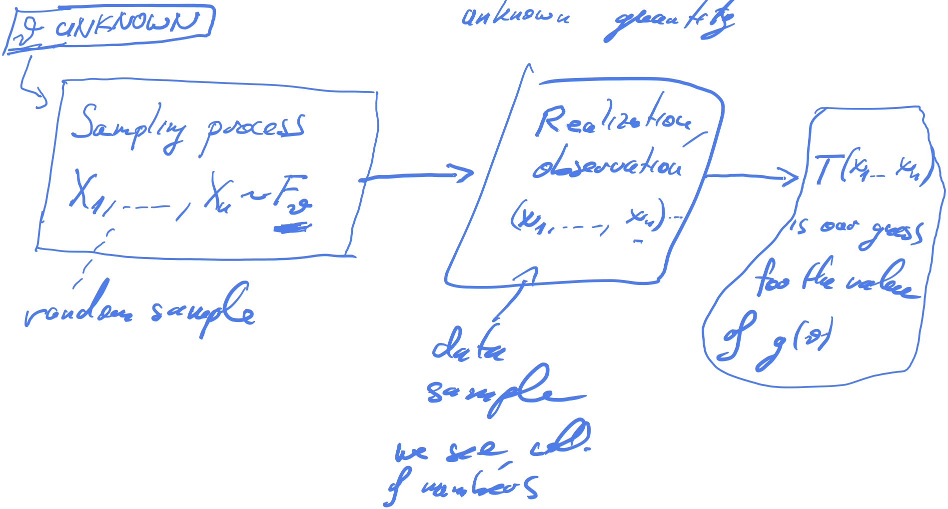
Estimator

Θ upper case theta
$\vartheta = \theta$ lower case theta

estimator ... T
estimate $T(x_1, \dots, x_n)$

Definition

Estimator is an arbitrary statistics. used to estimate an unknown quantity



Properties of estimators

$bias_\vartheta$ depends on ϑ

Definition

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

- WANT → *unbiased – pokud* $g(\vartheta) = \mathbb{E}(T_n)$ *(for each* ϑ*)* ⟸? $bias_\vartheta = 0$
- OR AT LEAST → *asymptotically unbiased – if* $g(\vartheta) = \lim_{n\to\infty} \mathbb{E}(T_n)$

  "T_n is eventually almost correct."
- *consistent – if* $T_n \xrightarrow{P} \vartheta$.

  $P(|T_n - g(\vartheta)| > \varepsilon) \xrightarrow{n\to\infty} 0$

  error of the estimator
- *bias is defined as* $bias_\vartheta := \mathbb{E}(T_n) - g(\vartheta)$
- *mean squared error, MSE is* $MSE := \mathbb{E}\left((T - g(\vartheta))^2\right)$

  WANT MSE low

$\mu = ET$

Theorem

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

$ET - \mu = 0$

$\mathbb{E}\left(\left((T-\mu) - (g(\vartheta) - \mu)\right)^2\right) = \mathbb{E}\left((T-\mu)^2 - 2(T-\mu)(g(\vartheta)-\mu) + bias_\vartheta^2\right)$

$-bias$

use linearity of $\mathbb{E}$ $= \mathbb{E}(T-\mu)^2 + bias_\vartheta^2 + 2\, bias_\vartheta\, \mathbb{E}(T-\mu)$, $\mathbb{E}(T-\mu) = 0$

Properties of sample mean and variance

For

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 $= g(\vartheta)$
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

(1) $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$

$\mathbb{E}\bar{X}_n = \frac{1}{n}(\mathbb{E}X_1 + \mathbb{E}X_2 + \ldots)$

$= \frac{1}{n}(\mu + \mu + \ldots) = \mu$

$\to \bar{X}_n$ is unbiased

$\bar{X}_n \xrightarrow{P} \mu$ .... weak Law of large Numbers

$var(\bar{X}_n) = var\left(\frac{1}{n}(X_1 + \ldots + X_n)\right)$

$= \frac{1}{n^2} var(X_1 + \ldots + X_n)$

$= \frac{1}{n^2}(\underbrace{var(X_1) + \ldots}_{n}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Properties of estimators

$$\mu = \mathbb{E}X_1 = \mathbb{E}X_2 = \dots$$

$$\sigma^2 = \mathrm{var}(X_i)$$

$$\mathbb{E}\bar{X}_n = \mu, \quad \mathrm{var}\,\bar{X}_n = \frac{\sigma^2}{n}$$

$$\mathbb{E}(X_i-\mu)^2 \overset{\mathrm{def}}{=} \mathrm{var}(X_i) = \sigma^2$$

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

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

$$= \mathbb{E}\frac{1}{n}\sum\left[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}_n-\mu) + (\bar{X}_n-\mu)^2\right]$$

$$= \mathbb{E}\frac{1}{n}\sum(X_i-\mu)^2 - \mathbb{E}\frac{2}{n}\sum_{i=1}^{n}(X_i-\mu)\underbrace{(\bar{X}_n-\mu)}_{\text{no } i \text{ here}} + \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\underbrace{(\bar{X}_n-\mu)^2}_{\text{no } i}$$

$$= \frac{1}{n}\sum\mathbb{E}(X_i-\mu)^2 - 2\mathbb{E}(\bar{X}_n-\mu)(\bar{X}_n-\mu) + \mathbb{E}(\bar{X}_n-\mu)^2$$

$$= \sigma^2 - \underbrace{\mathbb{E}(\bar{X}_n-\mu)^2}_{\mathrm{var}(\bar{X}_n) = \sigma^2/n} = \sigma^2 - \frac{\sigma^2}{n} = \left(1-\frac{1}{n}\right)\sigma^2$$

Properties of estimators

we proved $\mathbb{E}\bar{S}_n = \underline{\left(1-\frac{1}{n}\right)}\sigma^2 \xrightarrow{n\to\infty} \sigma^2$

---- $\underline{\bar{S}_n}$ is asymptot. unbiased

(3) $\hat{S}_n = \frac{n}{n-1}\bar{S}_n$

$\mathbb{E}\hat{S}_n = \frac{n}{n-1}\mathbb{E}\bar{S}_n = \frac{n}{n-1}\left(1-\frac{1}{n}\right)\sigma^2 = \sigma^2$

$\underline{\hat{S}_n}$ is unbiased

we skip consistency of $\bar{S}_n$, $\hat{S}_n$

Method of moments

- $m_r(\vartheta) := \mathbb{E}(X^r)$ 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 $X_1, \ldots, X_n$ z F_ϑ ... r-th sample moment

Theorem

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

(For r=1 we proved this
for r>1 similar.)

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

we see from the data

we know if
ϑ is given ⟶

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

Method of moments – examples

(1) $X_1 \dots X_n \sim \underline{Ber(p)}$

X_i -- i-th person is LH

$\vartheta = p \in [0,1]$

M.M.: $\boxed{\vartheta = \overline{X_n}}$

$m_1(\vartheta) = \mathbb{E}X_1 = \underline{\vartheta}$

this suggest to use

$\widehat{m_1(\vartheta)} = \frac{1}{n}(X_1 + \dots + X_n) = \overline{X_n}$

$T(X_1 \dots X_n) = \overline{X_n}$ as estimator for ϑ

(2) $X_1, \dots, X_n \sim U(0, \vartheta)$

$m_1(\vartheta) = \mathbb{E}X_1 = \frac{0+\vartheta}{2} = \frac{\vartheta}{2}$

$\frac{\vartheta}{2} = \overline{X_n}$

$\widehat{m_1} = \overline{X_n}$

$\vartheta = (\mu, \sigma)$

$T(X_1, \dots, X_n) = 2\overline{X_n}$

(3) $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

two equations

Maximal likelihood, ML

r.v.

- random sample $X = (X_1, \ldots, X_n)$ from a distribution with parameter ϑ
- possible realization $x = (x_1, \ldots, x_n)$ numbers
- … joint pmf $p_X(x; \vartheta)$ depends on par. ϑ
- … 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 $L(x;\vartheta)$. $= P_X(x_1, \dots, x_n; \vartheta)$
- for convenience we put $\ell(x;\vartheta) = \log(L(x;\vartheta))$ cont. par.
- by independence of X_1, X_2, etc. we have

$$L(x;\vartheta) = P_{X_1}(x_1;\vartheta) \cdot P(x_2;\vartheta) \cdots$$

$$\cdot L(x_1;\vartheta) \cdot L(x_2;\vartheta) \cdots$$

$$\ell(x;\vartheta) = \log L(x_1;\vartheta) + \log(L(x_2;\vartheta) \cdots$$

$x_1 = 1,\ x_2 = 0,\ x_3 = 0,\ x_3 = 1, \dots$

43 # 5 6 of them

$$\ell(x;\vartheta) = \sum_{i=1}^{43} \log \underbrace{L(x_i;\vartheta)}_{\text{prob of 1}}$$

$$= \log L(x_1) + \log L(x_2) + \cdots$$

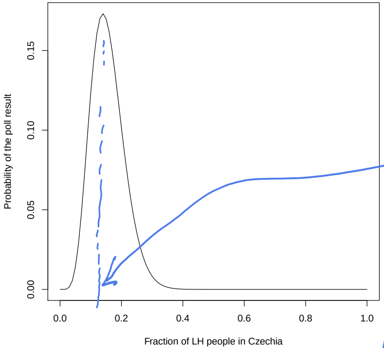
$$\log(p) + \log(1-p) + \log(1-p) \cdots$$

$$= \boxed{6 \log p + 37 \log(1-p)}$$

diff. $\dfrac{6}{p} - \dfrac{37}{1-p} = 0$

ML example – proportion of left-handed

$X_1, \dots, X_n \sim Ber(p)$ i.i.d. 6 LH out of 43 $Bin(43, p)$



$L(6 \text{ out of } 43) = \binom{43}{6} p^6 \cdot (1-p)^{37}$

what p s.t. ↑ is max'l

$(p^6(1-p)^{37})' =$

$6p^5(1-p)^{37} - 37p^6(1-p)^{36} = 0$

$\frac{6}{p} - \frac{37}{1-p} = 0$

$\frac{1-p}{p} = \frac{37}{6}$

$\frac{1}{p} - 1$

$\frac{1}{p} = \frac{37}{6} + 1 = \frac{43}{6}$

$\Rightarrow \boxed{p = \frac{6}{43}}$