NMAI 059 2019/2020 cvičenie 8 – Náhodný výber

1) $X_1, \dots, X_n$ náhodný výber z rozdelenia $X$, $X$ má hustotu $f$ a distr. fci $F$

a) odhad $EX$ je napríklad výberový priemer $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Že ide o dobrý odhad vieme z prednášky (napr. platí zákon veľkých čísiel)

b) $E\bar{X} = E\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{n}{n} EX_1 = EX_1$ odhad je nestranný

$$\operatorname{var}\bar{X} = \operatorname{var}\frac{1}{n}\sum X_i = \frac{1}{n^2}\cdot\operatorname{var}\sum X_i \underset{X_i \text{ nezávislé}}{=} \frac{n}{n^2}\operatorname{var}X_1 = \frac{\operatorname{var}X_1}{n} \xrightarrow[n\to\infty]{} 0.$$

c) $a$ je krajný bod, je možné ho odhadovať napr. $\hat{a} := \min_{i=1..n} X_i$

$\hat{a}$ je náhodná veličina. Pre jej distribučnú funkciu platí ($x \in \mathbb{R}$ pevné)

$$P(\hat{a} \le x) = P(\min_{i=1..n} X_i \le x) = 1 - P(\min X_i > x) = 1 - P(X_i > x \ \forall i) =$$

$$\underset{X_i \text{ nezávislé}}{=} 1 - \prod_{i=1}^{n} P(X_i > x) \underset{X_i \text{ stejne rozdelené}}{=} 1 - P(X_1 > x)^n = 1 - (1 - P(X_1 \le x))^n = \underbrace{1 - (1 - F(x))^n}_{\text{distr. fce } \hat{a} \text{ v } x} = F_{\hat{a}}(x)$$

Pre hustotu $\hat{a}$ derivujeme distr. fci podľa $x$:

$$\frac{\partial}{\partial x}\left(1 - (1 - F(x))^n\right) = n(1 - F(x))^{n-1}\frac{\partial F(x)}{\partial x} = n(1 - F(x))^{n-1} f(x) = \underset{\text{hustota } \hat{a} \text{ v } x}{f_{\hat{a}}(x)}$$

d) $X \sim R([a,b])$, tj. $F(x) = \frac{x-a}{b-a} I[x \in (a,b)] + 1 \cdot I[x \ge b]$

$f_X(x) = 1/(b-a)\, I[x \in (a,b)]$

$$E\hat{a} = \int_a^b f_{\hat{a}}(x)\cdot x \, dx = \int_a^b n\left(1 - \frac{x-a}{b-a}\right)^{n-1} \frac{1 \cdot x}{b-a}\, dx \underset{\frac{x-a}{b-a} = y}{=} \int_0^1 n(1-y)^{n-1}\left[y(b-a)+a\right] dy \underset{1-y=z}{=}$$

$$= \int_0^1 n z^{n-1}\left[(1-z)(b-a) + a\right] dz = \left[z^n(b-a) - n\frac{z^{n+1}}{n+1}(b-a) + z^n a\right]_{z=0}^{1} =$$

$$= (b-a) - \frac{n}{n+1}(b-a) + a = \underbrace{\frac{b}{n+1}}_{\xrightarrow[n\to\infty]{} 0} + \underbrace{\frac{an}{n+1}}_{\xrightarrow[n\to\infty]{} a}$$

odhad nie je nestranný, ale stredná hodnota konverguje k $a$ pre $n \to \infty$.

e) zaujíma nás $\theta = P(X \le 10)$ dobrý odhad je relatívna četnosť

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} I[X_i \le 10]. \quad E\hat{\theta} = E\frac{1}{n}\sum I[X_i \le 10] = \frac{n}{n} E\, I[X_1 \le 10] = P(X_1 \le 10) = \theta$$

nestranný odhad

$X_i$ nezávislé, stejne rozdelené $\Rightarrow I[X_i \le 10]$ nezávislé, stejne rozdelené

$$\operatorname{var}\hat{\theta} = \operatorname{var}\frac{1}{n}\sum I[X_i \le 10] = \frac{1}{n^2}\cdot n \cdot \operatorname{var} I[X_i \le 10] = \frac{\theta(1-\theta)}{n}$$

$$I[X_1 \le 10] = \begin{cases} 0 & \text{ak } X_1 > 10 \text{, tj. s pravd. } 1-\theta \\ 1 & \text{ak } X_i \le 10 \text{, tj. s pravd. } \theta \end{cases} \Rightarrow I[X_i \le 10] \sim Alt(\theta) \text{ a vieme, že } \operatorname{var} I[X_i \le 10] = \theta(1-\theta)$$

f) pre $F$ rovnomerného rozdelenia je $F(x) = \begin{cases} \frac{x-a}{b-a} & \text{ak } x \in (a,b) \\ 1 & \text{ak } x \ge b \end{cases}$

Môžeme odhadnúť $a$ pomocou $\hat{a} = \min X_i$ z c), podobne $\hat{b} = \max X_i$ a vykresliť $F$ pre $\hat{a}$ a $\hat{b}$ odhadnuté oproti $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \le x]$. ak sa funkcie "podobajú", ide o rozdelenie ktoré by malo byť rovnomerné na $[a,b]$.

2) úspech = píše ľavou rukou. Zrejme potom ide o $X \sim Alt(p)$, kde $p$ = pst toho, že náhodne vybraná osoba píše ľavou rukou.

$$P(X=1) = p$$
$$P(X=0) = 1-p$$

a) $EX = 1 \cdot p + 0 \cdot (1-p) = p$ $\qquad EX^2 = 1^2 \cdot p + 0^2 \cdot (1-p) = p$
$\operatorname{var} X = EX^2 - (EX)^2 = p - p^2 = p(1-p)$

b) $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} I[X_i = 1]$, tj. "relatívna četnosť úspešnosti". Ak $X_1, \dots, X_n$ je náh. výber z rozdelenia $X$, pôjde o dobrý odhad. Všimnime si, že $\hat{p} = \overline{X}$, lebo $X_i$ je iba 0 alebo 1.

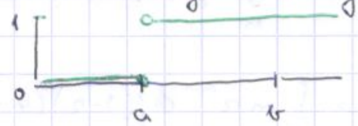
c) $E\hat{p} = E\overline{X} \overset{1b)}{=} EX_1 = p$ nestranný odhad

$\operatorname{var} \hat{p} = \operatorname{var} \overline{X} \overset{1b)}{=} \frac{\operatorname{var} X_1}{n} = \frac{p(1-p)}{n}$

3) V pr 1 sme nepredpokladali špeciálny tvar rozdelenia $X$ [tzv. neparametrický prístup]. V pr 2. sme z povahy experimentu usúdili, že $X \sim Alt(p)$ pre nejaké neznáme $p \in (0,1)$ a stačilo nám odhadnúť neznámy parameter $p$. [tzv. parametrický prístup].

4) Pre $F_{\hat{a}}$ platí $F_{\hat{a}}(x) = 1 - (1-F(x))^n = \begin{cases} 0 & \text{ak } x < a \\ 1 - \left(1 - \frac{x-a}{b-a}\right)^n & \text{ak } x \in [a,b] \\ 1 & \text{ak } x > b. \end{cases}$

$\lim_{n \to \infty} F_{\hat{a}}(x) = \begin{cases} 0 & \text{ak } x \leq a \\ 1 & \text{ak } x > a \end{cases}$ teda v limite je distrib. fce tvaru




To zodpovedá náhodnej veličine $Y = a$ skoro iste, tj "konštanta" $a$.

Z definície teda $\hat{a} \xrightarrow{P} a$, $\hat{a}$ "konverguje v pravdepodobnosti" k $a$.

pozri prednášku.