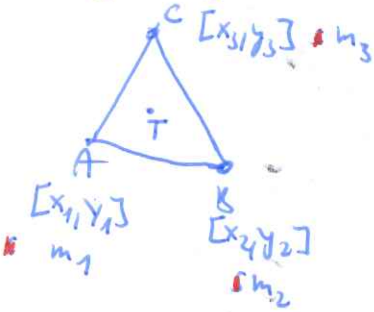
Geometrická interpretácia strednej hodnoty

Uvažujme trojuholník s váhami m_i, potom



súradnice ťažiska sú

$$T = \left[\overbrace{\underbrace{\frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}}}^{x_T} , \overbrace{\frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3}}^{y_T} \right]$$

$$x_T = \underbrace{\frac{m_1}{m_1+m_2+m_3}}_{p_{x_1}} \cdot x_1 + \underbrace{\frac{m_2}{m_1+m_2+m_3}}_{p_{x_2}} \cdot x_2 + \underbrace{\frac{m_3}{m_1+m_2+m_3}}_{p_{x_3}} \cdot x_3$$

"VÁŽENÝ PRIEMER HMOTNOSŤAMI"

TAKŽE STREDNÁ HODNOTA X je x-ová súradnica ťažiska, ak náhodná veličina X nadobúda x-ové súradnice vrcholov Δ s nenulovými pst $\frac{m_i}{m_1+m_2+m_3}$.

Zovšeobecnenie

Nech A je hmotná doštička s hustotou materiálu $\rho(x,y)$, (nejaká oblasť roviny s hmotnosťou rozloženou podľa $\rho(x,y)$)

potom hmotnosť A je

$$M = \iint_A \rho(x,y)\, dx\, dy \neq 0$$

$T = [x_T, y_T]$

$$x_T = \frac{1}{M} \iint_A \rho(x,y) \cdot x \, dx\, dy$$

ak A je vhodná oblasť roviny

$$= \int_{A_x} x \Big(\underbrace{\frac{1}{M} \int_{A_y} \rho(x,y)\, dy}_{\text{marginálna hustota}} \Big) dx = \underbrace{\int_{A_x} x \cdot \rho_X(x)\, dx}_{E(X)}$$

$A_x, A_y \subseteq \mathbb{R}$

$$\rho_X(x) = \frac{1}{M} \int_{A_y} \rho(x,y)\, dy$$

Ak sa pozrieme na spojitú N.V. X, potom jej rozdelenie pst. je popísané nezápornou hustotou $f_X(x)$ s vlastnosťou $\int_{\mathbb{R}} f_X(x)\, dx = 1$ a platí $E(X) = \int_{\mathbb{R}} x \cdot f_X(x)\, dx$

VZOREČKY: $E(g(X)) = \begin{cases} \sum_{-\infty}^{\infty} g(x)\, \pi_X(x) & X \text{ diskrétna N.V.} \\ \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ spojitá N.V} \end{cases}$

(ak existujú a sú konečné)

Transformácie náhodnej veličiny $X: \Omega \to \mathbb{R}$ spojitá N.V.

$g: \mathbb{R} \to \mathbb{R}$ vhodná funkcia, potom $Y = g(X) : \Omega \to \mathbb{R}$

Chceme popísať rozdelenie pravdepodobnosti pomocou hustoty f_X náhodnej veličiny X.
Predpokladajme, že g má inverziu g^{-1}.

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = \int_{\{g(x) \le y\}} f_X(x)\, dx \quad \left| \begin{array}{l} t = g(x) \\ x = g^{-1}(t) \\ dx = \frac{d}{dt}(g^{-1}(t))\, dt \end{array} \right|$$

$$= \int_{-\infty}^{y} f_X(g^{-1}(t)) \cdot \underbrace{\left| \frac{d}{dt} g^{-1}(t) \right|}_{|J|} dt$$

abs. hodnota Jakobiánu transformácie.

$$\Rightarrow \quad \boxed{f_Y(t) = f_X(g^{-1}(t)) \cdot \left| \frac{d}{dt} g^{-1}(t) \right|}$$

(6) Používame vzoreček $f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d g^{-1}(y)}{dy} \right|$.

a) $Y = e^X$, $X \geq 0$, vieme, že isto $g(x) = e^x \geq 0$ a $g^{-1}(y)$ existuje
$Y \geq 0$

$y = e^x \Rightarrow x = \ln(y) = g^{-1}(y)$ $\quad \frac{d g^{-1}(y)}{dy} = \frac{d \ln y}{dy} = \frac{1}{y}$

$f_Y(y) = f_X(\ln(y)) \cdot \left|\frac{1}{y}\right| = f_X(\ln(y)) \frac{1}{y}$ $\quad y > 0$; $f_Y(y) = 0$ $y \leq 0$

b) $Y = \sqrt{X} \geq 0$

$y = \sqrt{x} \Rightarrow x = y^2 = g^{-1}(y)$ $\quad \frac{d g^{-1}(y)}{dy} = 2y$
$x > 0 \longrightarrow$

$f_Y(y) = f_X(y^2) \cdot 2y$ $\quad y \geq 0$, $f_Y(y) = 0$ $y < 0$

c) $Y = \ln(X)$, $X > 0$

$y = \ln(x) \Rightarrow x = e^y = g^{-1}(y)$ $\quad \left| \frac{d g^{-1}(y)}{dy} = e^y \right.$

$f_Y(y) = f(e^y)\, e^y$ $\quad y \in \mathbb{R}$

d) $Y = \frac{1}{X} > 0$, lebo $X > 0$ $\quad g^{-1}(y) = \frac{1}{y}$ $\quad \frac{d g^{-1}(y)}{dy} = -\frac{1}{y^2}$

$f_Y(y) = f(1/y) \cdot \left| -\frac{1}{y^2} \right| = f(1/y)\, 1/y^2$ $\quad y > 0$, $f_Y(y) = 0$ $y \leq 0$

(7) $X \sim Rs(-\pi/2, \frac{\pi}{2})$

$f_X(x) = \begin{cases} c = \frac{1}{\pi} & x \in (-\pi/2, \pi/2) \\ 0 & \text{inak} \end{cases}$

$f_Y(y) = ?$, $f_Z(z) = ?$

$Y = \sin(X)$, $Z = tg(X)$

$Y \in [-1, 1]$, $Z \in \mathbb{R}$

$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d g^{-1}(y)}{dy} \right| = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}} & y \in [-1, 1] \\ 0 & \text{inak} \end{cases}$

$g(x) = \sin(x)$ pre $x \in (-\pi/2, \pi/2)$

$g^{-1}(y) = \arcsin(y)$

$h(x) = tg(x)$ $\quad x \in (-\pi/2, \pi/2)$

$h^{-1}(z) = \arctan(z)$

$f_Z(z) = f_X(h^{-1}(z)) \cdot \left| \frac{d h^{-1}(z)}{dz} \right| = \frac{1}{\pi} \frac{1}{1+z^2}$ $\quad z \in \mathbb{R}$

---

(8) $f(x) = \begin{cases} \cos x & x \in (0, \pi/2) \\ 0 & \text{inak} \end{cases}$ $\quad f_Y(y) = ?$, $Y = X^2 \Rightarrow \sqrt{Y} = X$, $X \geq 0$
čo je OK

$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2} \frac{1}{\sqrt{y}} = \frac{1 \cos(\sqrt{y})}{2\sqrt{y}}$ $\quad \sqrt{y} \in (0, \pi/2)$ t.j. $y \in (0, \pi^2/4)$

$E(Y) = \int_0^{\pi^2/2} y \cdot \frac{1}{2} \frac{1}{\sqrt{y}} \cdot \cos(\sqrt{y})\, dy = \left| \begin{array}{l} t = \sqrt{y} \\ y = t^2 \\ dy = 2t\,dt \end{array} \right| = \int_0^{\pi/2} t^2 \cos(t)\, dt =$ atď.

(musí vyjsť rovnako ako $E(X^2)$ na príklade 3.)

$E(Y^2) = \ldots$

Nech $\Phi(x)$ je DISTRIBUČNÁ FUNKCIA $N(0,1)$, potom platí

$$\boxed{\Phi(x)+\Phi(-x)=1 \quad \forall x\in\mathbb{R}.}$$!

Zadání 13. cvičení, podzim 2019
Popisná statistika, nerovnosti, limitní věty

Exponenciální rozdělení $\mathrm{Exp}(\lambda)$ je rozdělení s hustotou

$$f(x) = \begin{cases} 0 & \text{for } x < 0, \\ \lambda\,\mathrm{e}^{-\lambda x} & \text{for } x \geq 0. \end{cases}$$

Markovova nerovnost. Pro libovolnou nezápornou náhodnou veličinu X se střední hodnotou $\mathrm{E}(X)$ platí

$$P(X \geq \varepsilon) \leq \frac{\mathrm{E}(X)}{\varepsilon}$$

pro všechna $\varepsilon > 0$.

Čebyševova nerovnost. Pro libovolnou náhodnou veličinu X se střední hodnotou $\mathrm{E}(X)$ a roztylem $\mathrm{D}(X)$ platí

$$P(|X - \mathrm{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{D}(X)}{\varepsilon^2}$$

nebo ekvivalentně

$$P(|X - \mathrm{E}(X)| < \varepsilon) \geq 1 - \frac{\mathrm{D}(X)}{\varepsilon^2}$$

pro všechna $\varepsilon > 0$.

Centrální limitní věta. Uvažujme posloupnost nezávislých náhodných veličin X_i, které mají společnou střední hodnotu $EX_i = \mu$, společný rozptyl $\operatorname{var} X_i = \sigma^2 > 0$ a stejně omezený třetí absolutní moment $E|X_i| < C$. Pro rozdělení náhodných veličin

$$S_n = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n}\sigma}$$

$$Z_n = \frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{D(X_i)}} = \frac{\sum_{i=1}^{n}(X_i) - n\mu}{\sqrt{n}\cdot\sigma}$$

STANDARDIZOVANÁ NÁHODNÁ VELIČINA

platí limitní vztah

$$\lim_{n\to\infty} P(S_n \leq x) = \Phi(x),$$

kde Φ je distribuční funkce normovaného normálního rozdělení $\mathrm{N}(0,1)$.

Moivreova-Laplaceova věta je speciálním případem centrální limitní věty: Nechť X_n jsou náhodné veličiny s binomickým rozdělením $\mathrm{Bi}(n,p)$. Pak pro normované náhodné veličiny

$$S_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

platí vztah

$$\lim_{n\to\infty} P(S_n \leq x) = \Phi(x),$$

kde Φ je distribuční funkce normovaného normálního rozdělení $\mathrm{N}(0,1)$.

Příklad 1. *Byly naměřeny následující hodnoty nějakého znaku*

10; 7; 7; 8; 8; 9; 10; 9; 4; 9; 10; 9; 11; 9; 7; 8; 3; 9; 8; 7.

Určete aritmetický průměr, VÝBEROVÝ *medián,* VÝBEROVÉ *kvartily,* VÝBEROVÝ *rozptyl.*

Příklad 2. *Mějme nezápornou náhodnou veličinu X se střední hodnotou μ.*

(1) *Bez dalších informací o rozdělení X odhadněte $P(X > 3\mu)$.*
(2) *Víte-li, že $X \sim \mathrm{Ex}(\frac{1}{\mu})$, vypočtěte $P(X > 3\mu)$.* kde $\frac{1}{\mu} > 0$, lebo $\mu > 0$ (parameter Ex je vždy kladný!)

stredná hodnota X ($E(X)$ je $\frac{1}{\frac{1}{\mu}} = \mu$)

① 10 3x | 7 4x
11 1x | 4 1x
9 6x | 3 1x
8 4x

$\bar{X} = \dfrac{10\cdot 3 + 11\cdot 1 + 9\cdot 6 + 8\cdot 4 + 7\cdot 4 + 4\cdot 1 + 3\cdot 1}{20}$ ← výberový priemer

$\tilde{X}_{0,5} =$ výberový medián $= \dfrac{X_{(10)} + X_{(11)}}{2} = \dfrac{9+8}{2} = \dfrac{17}{2}$

z usporiadaného súboru dát

Kvantily: • DOLNÍ
• HORNÍ

$\tilde{X}_{0,25} = \dfrac{X_{(5)} + X_{(6)}}{2} = \dfrac{9+9}{2} = 9$ $\boxed{\frac{1}{4}\cdot 20 = 5}$

$\tilde{X}_{0,75} = \dfrac{X_{(15)} + X_{(16)}}{2} = \dfrac{7+7}{2} = 7$ $\boxed{\frac{3}{4}\cdot 20 = 15}$

Výberový rozptyl $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{20} (x_i - \bar{x})^2 = \dots$

$n = 20$

Definícia kvantilov: Nech dáta sú usporiadané vzostupne. Určíme index i_p vyhovujúci rovnici $np < i_p < np+1$, potom $\tilde{x}_p = x_{(i_p)}$.
Ak sú hodnoty np, $np+1$ celočíselné, potom $\tilde{x}_p = \dfrac{x_{(np)} + x_{(np+1)}}{2}$

---

② $X \geq 0$, $E(X) = \mu$

Markovova nerovnosť $\forall \varepsilon > 0$
$P(X \geq \varepsilon) \leq \dfrac{E(X)}{\varepsilon}$

Čiviko (1)

$P(\{X > 3\mu\} \cup \{X = 3\mu\})$

$P(X > 3\mu) \leq P(X \geq 3\mu) \leq \dfrac{E(X)}{3\mu} = \dfrac{1}{3}$

(2) $X \sim Ex(\frac{1}{\mu})$, určte $P(X > 3\mu)$ $\mu > 0$ (hustota je nezáporná f a X nadobúda nezáporné hodnoty integrál je plocha pod krivkou)

$P(X > 3\mu) = 1 - P(X \leq 3\mu) = 1 - F(3\mu) = 1 - (1 - e^{-3\mu\frac{1}{\mu}})$
$= \underline{\underline{e^{-3}}}$

Distribúcia $F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$

---

③ prednáška

④ $E(X) = 20$, $X \geq 0$ (uvažujeme nezápornú výšku celku)

a) $P(X \leq 60) = 1 - P(X > 60) \overset{(*)}{\geq} 1 - \frac{1}{3} = \underline{\underline{\frac{2}{3}}}$

(*) $P(X > 60) \leq P(X \geq 60) \leq \dfrac{E(X)}{\varepsilon} = \dfrac{20}{60} = \dfrac{1}{3}$

b) $P(a \leq X \leq b) \geq 0{,}9$, $DX = 1$ $a, b = ?$

Čebyševova nerovnosť: $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{DX}{\varepsilon^2}$ t.j. $P(|X-20| < \varepsilon)$

$P(|X-20| \geq \varepsilon) \leq \dfrac{1}{\varepsilon^2} \Rightarrow 1 - P(|X-20| \geq \varepsilon) \geq 1 - \dfrac{1}{\varepsilon^2}$

Ivanced Valves Solutions

$$1-0{,}9=\frac{1}{\varepsilon^2}$$

$$P(|X-20|<\varepsilon)\geq 1-\frac{1}{\varepsilon^2}=0{,}9 \qquad 0{,}1\,\varepsilon^2=1$$

$$\varepsilon^2=10 \Rightarrow \boxed{\varepsilon=\sqrt{10}}$$

$$P(|X-20|<\sqrt{10})\geq 0{,}9$$

$$P(-\sqrt{10}<X-20<\sqrt{10})\geq 0{,}9$$

$$P(20-\sqrt{10}<X<20+\sqrt{10})\geq 0{,}9$$

interval je $(20-\sqrt{10},\ 20+\sqrt{10})$

QUIZ

(5) 10% studentov s prospechom do 1,2 $\Rightarrow E\left(\sum_{i=1}^{n} X_i\right)=0{,}1n$

$n=?$

$X_i=\begin{cases}1 & \text{student s prospechom do } 1{,}2\\ 0 & \text{inak}\end{cases}$ $\quad P\left(0{,}08n\leq \sum_{i=1}^{n}X_i\leq 0{,}12n\right)=0{,}95$ $\quad D\left(\sum X_i\right)=n\cdot p\cdot(1-p)=n\cdot 0{,}1\cdot 0{,}9$

$$\boxed{\sum_{i=1}^{n}X_i\sim Bi(n;\underbrace{0{,}1}_{p})}$$

(A) Čebyševova nerovnosť

$$P\left(\underbrace{0{,}08n-E(X)}_{0{,}08n-0{,}1n}\leq \sum_{i=1}^{n}X_i-E(X)\leq \underbrace{0{,}12n-E(X)}_{0{,}12n-0{,}1n}\right)=0{,}95$$

$$P\left(-0{,}02n\leq \sum_{i=1}^{n}X_i-E(X)\leq 0{,}02n\right) \quad (*)$$

Podmienky dobrej aproximácie

$$n\cdot p(1-p)>9$$

$$\frac{1}{n+1}<p<\frac{n}{n+1}$$

$$P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|\leq 0{,}02n\right)=0{,}95$$

$$1-P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|\geq 0{,}02n\right)=0{,}05$$

$$P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|>0{,}02n\right)=0{,}05$$

A (Čebyšev) $\quad P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|\geq 0{,}02n\right)\leq \frac{0{,}1\cdot 0{,}9\cdot n}{0{,}02^2\cdot n^2}=\frac{0{,}09}{0{,}0004n}=\frac{900}{4n}=\frac{225}{n}$

$$P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|>0{,}02n\right)\leq P\left(\left|\sum_{i=1}^{n}X_i-E(X)\right|\geq 0{,}02n\right)\leq \frac{225}{n}=0{,}05$$

$$n\geq \frac{225}{0{,}05}\geq \frac{22500}{5}\geq 4500$$

B, CLV

$$\lim_{n\to\infty}P\Bigg(\underbrace{\frac{\sum_{i=1}^{n}X_i-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}}_{:=Z_n}\leq x\Bigg)=\Phi(x)$$

$\hookrightarrow$ DIST. F. $N(0,1)$

Použijeme $(*)$ $\quad P\left(-0{,}02n\leq \sum_{i=1}^{n}X_i-E(X)\leq 0{,}02n\right)=0{,}95$

ÚLOHA 6,
$\Phi(x)+\Phi(-x)=1$

$$P\left(\frac{-0{,}02n}{\sqrt{n\cdot 0{,}1\cdot 0{,}9}}\leq \frac{\sum_{i=1}^{n}X_i-E(X)}{\sqrt{n\cdot 0{,}1\cdot 0{,}9}}\leq \frac{0{,}02n}{\sqrt{n\cdot 0{,}1\cdot 0{,}9}}\right)=0{,}95$$

$$\Phi\left(\frac{2}{30}\sqrt{n}\right)-\boxed{\Phi\left(\frac{-1}{15}\sqrt{n}\right)} \stackrel{A}{=} P\left(-\frac{0{,}02\sqrt{n}\cdot 100}{30}\leq Z_n\leq \frac{0{,}02\sqrt{n}\cdot 10}{3}\right)=0{,}95$$

$$\Phi\left(\frac{1}{15}\sqrt{n}\right)-\left(1-\Phi\left(\frac{1}{15}\sqrt{n}\right)\right)$$

$$2\cdot\Phi\left(\frac{\sqrt{n}}{15}\right)-1=0{,}95$$

$Z_n \underset{A}{\sim} N(0,1)$ (Spojitá N.V., použijeme vetu pre dist. funkciu)

$2 \cdot \Phi\left(\frac{\sqrt{n}}{15}\right) = 1{,}95 \Rightarrow \Phi\left(\frac{\sqrt{n}}{15}\right) = \frac{1{,}95}{2} = 0{,}975$ | Tabuľky: $z = \Phi^{-1}(0{,}975) = \underbrace{1{,}96}_{1{,}9+0{,}06}$

$z = \frac{\sqrt{n}}{15} \Rightarrow \sqrt{n} = 15z \quad |^2$

$n = 15^2 \cdot z^2 = 15^2 \cdot 1{,}96^2 \approx 865{,}5$

	0,06
1,9	0,975

výsledok $\lceil 15^2 \cdot 1{,}96^2 \rceil$.

skoro 865 študentov.

CVIKO

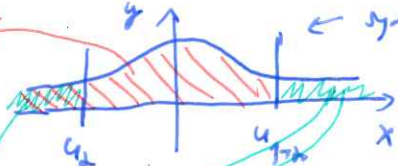
(6) $\boxed{-u_\alpha = u_{1-\alpha}}$

$\Phi(u_\alpha)$ = plocha pod krivkou

def: $P(U \leq u_\alpha) = \alpha$, $P(U \leq u_{1-\alpha}) = 1-\alpha = \Phi(u_{1-\alpha})$

$1-\alpha = \Phi(u_{1-\alpha})$

← symetrická podľa y



$\alpha = \Phi(u_\alpha)$

rovnaké

z obrázka:

$\underbrace{\Phi(u_{1-\alpha})}_{1-\alpha} + \Phi(u_\alpha) = 1$

$\Phi(u_\alpha) = \alpha$

dostávame tiež

$\Phi(u_\alpha) + \Phi(-u_\alpha) = 1$, ak $x = u_\alpha \in \mathbb{R}$

$\Phi(x) + \Phi(-x) = 1 \quad \forall x \in \mathbb{R}$

Zo symetrie hustoty $N(0,1)$,

kde $\boxed{u_{1-\alpha} = -u_\alpha}$