Pravdepodobnosť a štatistika

Na základe prednášok Daniela Hlubinky

Vladislav Vancák

hlubinka@karlin.mff.cuni.cz

PRAVDĚPODOBNOST:

Pravděpodobnostní prostor:

- d-stěnná kostka => výsledky $1, .. d$.
- $\forall$ prvek množiny $\{1, \dots d\}$ = ELEMENTÁRNÍ JEV
- Ω - MNOŽINA EL. JEVŮ ... neprázdná!
- $\mathcal{F}$ - SYSTÉM PODMNOŽIN Ω : vlastnosti:
  - I. $\Omega \in \mathcal{F}$ (prvky $\mathcal{F}$ jsou množiny).
  - II. $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$ (je tam i komplement)
  - III. Pro libovolné $A_1, A_2 \dots \in \mathcal{F}$ je $\bigcup_i A_i \in \mathcal{F}$ (spočetné & konečné) ... Uzavřenost na spočetná sjednocení.
- $\mathcal{F}$ nazveme σ-algebrou na Ω
- např. $\{\emptyset, \Omega\}$ je σ-algebra.
- pro Ω nejvýše spočetnou je $\mathcal{P}(\Omega)$ σ-algebrou.
  - ↳ i pro nespočetnou, ale to by dále vedlo k paradoxu.
- prvky $\mathcal{F}$ nazveme náhodné jevy.
- P množinová funkce na $\mathcal{F}$ se nazývá pravděpodobnost / pravděp. míra, pokud:
  - I. $\forall F \in \mathcal{F}$: $0 \leq P(F) \leq 1$ (P vezme $\mathcal{F}$ a přiřadí jí číslo)
  - II. Pro $F_1, F_2, \dots$ po 2 disjunktní: $P(\bigcup_{i=1}^{n/\infty} F_i) = \sum_{i=1}^{n/\infty} P(F_i)$
    ... σ-aditivita P
  - III. $P(\Omega) = 1$.

Def: Trojice $(\Omega, \mathcal{F}, P)$ se nazývá PRAVDĚPODOBNOSTNÍ PROSTOR.

např. ⚀ ... $\Omega = \{1,2,3,4,5,6\}$

$\mathcal{F} = \mathcal{P}(\Omega)$

$P(\{i\}) = p_i \quad i = 1,2 \dots 6 \;;\; p_i \geq 0 \wedge \sum_{i=1}^{6} p_i = 1$

- např. $p_i = 1/6 \quad i = 1..6$ pro pravidelnou.

p_i je „očekávaný podíl výsledku i v mnoha pokusech provedených za stejných podmínek.

Vlastnosti a počítání pravděpodobností:

$P(A) = 1 - P(A^c)$ ... věta o doplňkové pravděpodobnosti

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Dk: Pro $B \subset A$ platí $P(A \setminus B) = P(A) - P(B)$

· $B \cap (A \setminus B) = \emptyset$

· $B \cup (A \setminus B) = A$ disjunktní

· $E_1 \cup E_2 = E_1 \cup \underbrace{(E_2 \setminus E_1)}_{\subset E_2}$

$\Rightarrow P(E_1 \cup E_2) = P(E_1 \cup (E_2 \setminus E_1)) = P(E_1) + \cancel{P(E_2 \setminus E_1)}\, P(E_2 \setminus (E_1 \cap E_2$

$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

V: Princip inkluze a exkluze:

Buďte $E_1 .. E_n$ náhodné jevy. Pak $P(\bigcup_{i=1}^{n} E_i) =$

$$= \sum_{i=1}^{n} P(E_1) - \sum_{1 \le i < j \le n}\sum P(E_1 \cap E_j) + \sum\sum\sum_{1 \le i < j < k \le n} P(E_i \cap E_j \cap E_k) + \dots + (-1)^{n-1} P(\bigcap_{i=1}^{n} E_i)$$

Dk: MI:

$n=2$ věta platí.

$$P(\bigcup_{i=1}^{n} E_i) = P(\bigcup_{i=1}^{n-1} E_i \cup E_n) = P(\bigcup_{i=1}^{n-1} E_i) + P(E_n) - P((\bigcup_{i=1}^{n-1} E_i) \cap E_n) =$$

$$= \sum_{i=1}^{n} P(E_i) - \sum_{1 \le i < j \le n-1} P(E_i \cap E_j) + \sum_{1 \le i < j < k \le n} P(E_i \cap E_j \cap E_k) - \dots + (-1)^{n-2} P(\bigcap_{i=1}^{n-1} E_i)$$

$$- \sum_{i=1}^{n-1} P(E_i \cap E_n) + \sum_{i=1;\, j \le n-1,\, j \ne i} P(E_i \cap E_j \cap E_n) \dots + (-1)^{n-1} P(\bigcap_{i=1}^{n} E_i) =$$

$\rightarrow \sum_{j>i=1}^{n} P(E_i \cap E_j)$

= to, čo sme chceli.

Př: 20-ti stěnná kostka .. číslo dělit. 3? $P = 6/20$

... padlo č. dělit. 4, jaká je pravděp., že padlo č. dělit. 3: 4, 8, 12, 16, 20

↳ 1/5 ⇒ Dělit. 3 má menší pravděp., pokud padlo dělit co dřěma? 2, 4, $\underline{6}$, 8, 10, $\underline{12}$, 14, 16, $\underline{18}$, 20 $\rightsquigarrow$ 3/10 -> nezávislé

Def: (nezávislost jevů):

i. Náhodné jevy E_1, E_2 jsou nezávislé, pokud $P(E_1 \cap E_2) = P(E_1)P(E_2)$

- $\underbrace{P(D_2) \cdot P(D_3)}_{\text{nezávislé}} = P(D_2 \cap D_3)$, ale $\underbrace{P(D_4) \cdot P(D_3)}_{\text{závislé}} \neq P(D_4 \cap D_3)$.

ii. Náhodné jevy E_α; $\alpha \in A$ jsou vzájemně nezávislé, pokud pro každou konečnou $I \subset A$ platí $P(\bigcap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$.

Jevy E_α; $\alpha \in A$ jsou po dvou nezávislé, pokud $\forall i, j \in A$ platí $P(E_i \cap E_j) = P(E_i)P(E_j)$.

Pozn. Vzájemně nezávislé jevy jsou také po 2 nezávislé, ale NE NUTNĚ NAOPAK. např:

$\Omega = \{1, 2, 3, 4\}$ a $p_i = 1/4$ $\forall i$.

$E_1 = \{1, 2\}$ $\quad E_1 \cap E_2 = \{1\}$ $\quad P(E_1) = P(E_2) = P(E_3) = 1/2$.

$E_2 = \{1, 3\}$ $\quad E_1 \cap E_3 = \{1\}$ $\quad P(E_1 \cap E_2) = 1/4 = P(E_1)P(E_2)$

$E_3 = \{1, 4\}$ $\quad E_2 \cap E_3 = \{1\}$ $\quad P(E_1 \cap E_3) = 1/4$ $\quad \vdots$

JSOU PO 2 NEZ. $P(E_2 \cap E_3) = 1/4$

ALE: $E_1 \cap E_2 \cap E_3 = \{1\}$

$P(E_1 \cap E_2 \cap E_3) = 1/4 \neq P(E_1)P(E_2)P(E_3) \Rightarrow$ Nejsou vz. nezávislé.

PROČ?:

nastaly E_1 i E_2 -> padla 1. => nastal i E_3.

nastal E_1 ale ne E_2 => padla 2 => nenastal ani E_3.

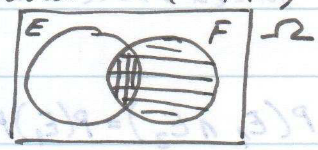
ne E_1, ani E_2 => padla 4 => nastal E_3.

=> To, jestli nastal E_1 a E_2 určuje E_3 => nejsou nezávislé.

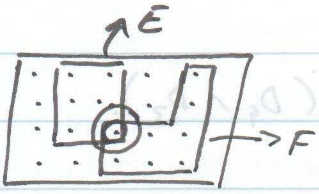
Def: (podmíněná pravděpodobnost):

Buďte E a F náhodné jevy, $P(F) > 0$. Podmíněná pravděp. E za podmínky F je dána $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$.

Platí: $P(F|F) = 1$.



$P(\cdot|F)$ je pravděpodobnost splňující axiomy I.–III. pro P.



V: Jsou-li E a F nezávislé jevy, $P(F)>0$, pak $P(E|F) = P(E)$.

Dk: z definice: $P(E|F) = \dfrac{P(E\cap F)}{P(F)} = \dfrac{P(E)P(F)}{P(F)} = P(E)$

Pozn. $\forall$ jev s nulovou $P(E)$ a každý jev s $P(E)=0$ je nezávislý s libovolným jevem.

Věty užitečné pro výpočet pravděp.:

Tvrzení: $E_1 .. E_n$ náhodné jevy takové, že $P(E_1\cap E_2 \dots \cap E_{n-1})>0$. pak $P(E_1\cap\dots\cap E_n) = P(E_n|E_1\cap\dots\cap E_{n-1}) * P(E_{n-1}|E_1\cap\dots\cap E_{n-2}) * \dots$
$\dots * P(E_2|E_1) . P(E_1)$

Dk: z definice: $P(E_n\cap(E_{n-1}\cap\dots\cap E_1)) = P(E_n|E_1\cap\dots\cap E_{n-1})\cdot$
$\cdot P(E_1\cap\dots\cap E_{n-1})$ a post. opak. dostaneme výsledok.

Tvrzení: (věta o úplné pravděp.) $E_1 .. E_n$ náhodné jevy:

i.) $E_i\cap E_j = \emptyset \quad \forall i\neq j$
ii.) $P(E_i)>0 \quad \forall i$
iii.) $\bigcup_{i=1}^{n} E_i = \Omega$

} Disjunktní rozklad.

a B náhodný jev. pak $P(B) = \sum_{i=1}^{n} P(B|E_i)P(E_i)$

Dk: $P(B) = P(B\cap\Omega) = P(B\cap(\bigcup_{i=1}^{n} E_i)) = P(\bigcup_{i=1}^{n}(B\cap E_i)) =$
$= \sum_{i=1}^{n} P(B\cap E_i) = \sum_{i=1}^{n} P(B|E_i)\cdot P(E_i)$ ↳ po 2 disjunktní

... POZN: věta platí i pro spočetný disj. rozklad.

Věta (Bayesova):
Za podmínek i) - iii) + $P(B)>0$ platí $P(E_i|B) = \dfrac{P(B|E_i)P(E_i)}{\sum_{j=1}^{n} P(B|E_j)P(E_j)}$

$P(E_i)$ ... apriorní rozdělení (pravděp.)

B ... informace

$P(E_i|B)$ aposteriorní pravděp.

Důkaz:

$P(E_i|B) = \dfrac{P(E_i \cap B)}{P(B)}$ ... v čitateli: $P(E_i \cap B) = P(B|E_i)P(E_i)$

... v menovateli $P(B) = \sum_{j=1}^{n} P(B|E_j)P(E_j)$ (o úplné pravděp.)

Př.

2 polynomy $R(x) = (x-a_1)(x-a_2)\dots(x-a_n)$; $a_i \in \mathbb{Z}$

$Q(x) = x^n + \dots$ má celoč. kořeny

Odhalte $Q \neq R$

$\rightsquigarrow$ dosadíme třeba 1 do $R(x) - Q(x)$. Pak jsou-li R a Q různé, jen s malou pravděp. dostaneme 0. $R-Q$ má nejvýše n různých kořenů. Vezmeme $100 \cdot n$ různých ~~kořenů~~ celých čísel a z nich 1 náhodně vybereme. Pravděpodobnost toho, že vybrané číslo je kořen $R-Q$ (jsou-li $R \neq Q$) je nejvýše $\frac{n}{100n} = 1/100$.

$\Rightarrow$ Vybereme-li náhodně k-čísel z oněch $100n$, pak jsou-li $R \neq Q$, pak $P(R(x_i) = Q(x_i) \ \forall i = 1 \dots k) \leq (1/100)^k$

? Co kdybychom vybírali bez vracení, t.j. pokud $R(x_i) = Q(x_i)$, tak x_i již nevrátíme (nevybíráme ho dál).

$P(R(x_i) = Q(x_i),\ i = 1,2 \dots k) = P(R(x_k) = Q(x_k) \mid R(x_i) = Q(x_i),\ i = 1 \dots k-1) * P(R(x_{k-1}) = Q(x_{k-1}) \mid R(x_i) = Q(x_i),\ i = 1 \dots k-2) * \dots$

$* \underbrace{P(R(x_2) = Q(x_2) \mid R(x_1) = Q(x_1))}_{\substack{n-1 \text{ kořenů} \\ \text{v } 100n-1 \text{ číslech}}} \leq \frac{n-1}{100n-1} < \frac{1}{100} \cdot \underbrace{P(R(x_1) = Q(x_1))}_{\leq 1/100}$

$\leq \prod_{i=1}^{k-1} \frac{n-i}{100n-i} \cdot \frac{1}{100} < (1/100)^k$

Náhodná veličina:

hod dvěma kostkami

$\Omega = \{1, 2, \dots, 6\}^2$

$\omega = (i, j) \quad i \in \{1, \dots, 6\}$

$\qquad\qquad\quad j \in \{1, \dots, 6\}$

Součet bodů na kostkách.

$\omega = (\omega_1, \omega_2) \rightsquigarrow \omega_1 + \omega_2$

$X : \Omega \to \mathbb{R}$ zde $X(\omega) = \omega_1 + \omega_2$

X je náhodná veličina,

$P(X = x)$?

$P(X = 4)$? $\Rightarrow P(X=4) = P(\{1,3\}, \{2,2\}, \{3,1\}) = 3/36 = 1/12$

$P(X=2) = 1/36 = P(X=12)$

$P(X=3) = 2/36 = P(X=11)$

$P(X=4) = 3/36 = P(X=10)$

$P(X=5) = 4/36 = P(X=9)$

$P(X=6) = 5/36 = P(X=8)$

$P(X=7) = 6/36$

$\Omega' = \{2, \dots 12\}$

$X' : \Omega' \to \mathbb{R}$

$X'(\omega) = \omega$

$P'(\omega) = \dots/36$

Definice: Zobrazení $X : \Omega \to \mathbb{R}$ takové, že $\forall a \in \mathbb{R} : X^{-1}(-\infty, a] = \{\omega : X(\omega) \leq a\} \in \mathcal{F}$ nazveme náhodná veličina. (M)

Poznámka: Pokud je X diskrétní, tedy nabývá nejvýše spočetně mnoho hodnot, můžeme vždy volit Ω spočetnou a $\mathcal{F} = \mathcal{P}(\Omega)$ ($\forall$ podmnožiny) a X pak vždy splňuje (M).

Definice: (Rozdělení náhodné veličiny): Buď $X : \Omega \to \mathbb{R}$ náhodná veličina a $(\Omega, \mathcal{F}, P)$ pravděp. prostor. Rozdělením náh. velič. X rozumíme pravděp. míru P_X na $\mathbb{R}$ takovou, že $P_X((a, b]) = P(\{\omega : X(\omega) \in (a, b]\}) = P(\{\omega : X(\omega) \leq b\}) - P(\{\omega : X(\omega) \leq a) \; \forall a <$

Poznámka: Takto zavedená pravděp. P_X může být rozšířena na borelovskou σ-algebru, t.j. nejmenší σ-algebru obsahující všechny otevřené podmnožiny $\mathbb{R}$, speciálně $\forall$ intervaly a jednobodové množiny a jejich konečná (i spočetná) sjednocení

Tvrzení: Pro diskrétní náhodnou veličinu X (t.j. má hodnoty v nějaké nejvýše spočetné podmnožině $S \subset \mathbb{R}$) je rozdělení jednoznačně dáno hodnotami $\{p_s, s \in S\}$ $p_s = P(X = s) = P(\{\omega : X(\omega) = s\})$

$P_X((a,b]) = \sum_{s \in S \cap (a,b]} p_s$ splňuje $\forall$ vlastnosti pravděp.

hodnoty $\{p_s, s \in S\}$

je hustotou rozdělení X vůči čítací míře na S.

Musí platit
$p_s \geq 0 \quad \forall s$
$\sum_{s \in S} p_s = 1$

$\mathcal{L}(A) = |A|$ .. počet bodů A
vhodné po konečné/spočetné množiny.

~> Klasická pravděp.:

$p(A) = \frac{|A|}{|\Omega|}$, zde $|\Omega| < \infty$

... ale co s $|\Omega| = \infty$, ale Ω spočetná?

$P(A) = \sum_{s \in A} p_s$ $\quad \mathcal{L}(A) = \sum_{s \in A} 1$

např. $\Omega = \mathbb{N}$; $p_s = 2^{-s}$

Definice (Distribuční funkce):

$F: \mathbb{R} \to [0,1]$ je def. jako $F(x) = P(X \leq x)$, což pro diskrétní náhodnou veličinu je $F(x) = \sum_{s \leq x} p_s$

jak distribuční funkce F, tak i hustota $\{p_s, s \in S\}$ plně char. rozdělení náh. veličiny X, tedy pravděp. P_X.

=> P_X je jednoznačně dána F.

RŮZNÉ NÁHODNÉ VELIČ. MOHOU MÍT STEJNÉ ROZDĚLENÍ.

Náhodná veličina a její rozdělení:

$(\Omega, \mathcal{F}, P)$, $\quad X: \Omega \to \mathbb{R}$

Chceme, aby $\{\omega : X(\omega) \leq a\} \in \mathcal{F}$

$X^{-1}((-\infty, a]) \in \mathcal{F}$, t.j. $[X \leq a]$ je náhodný jev a známe $P[X \leq a]$

Rozdělení n.v. je míra na $\mathbb{R}$.

Pravděpodobnost $P_X((-\infty,a]) = P(X \le a)$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\ \ \cap \quad\quad\quad\quad \cap$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\ \ \mathbb{R} \quad\quad\quad\quad \Omega$

Rozdělení náhodné veličiny

Rozdělení je vždy pravděp. na reálných číslech,
Ω je mat. nástroj po modelování & důkazy.

míra: Množinová funkce (tj. vezme množinu a přiřadí jí hodnotu ! nezápornou).

Příklad: Lebesgueova míra λ na $\mathbb{R}$

$\lambda(a,b) = b-a$ pro $b \ge a$ (interval)

$\lambda[a,b] = b-a$

$\lambda(\{a\}) = 0$

Čítací míra (aritmetická) $\mathcal{L}(A) = |A|$ hodí se pro nejvýše spočetné množiny, $\mathcal{L}_{\mathbb{N}}$ je čítací míra na přiroz. číslech

$B \subset \mathbb{R} \quad \mathcal{L}_{\mathbb{N}}(B) = \sum_{n=1}^{\infty} \underbrace{1(n \in B)}_{\text{indikátor}} = |B \cap \mathbb{N}|$

$1(n \in B) \Leftrightarrow 1_B(n) = \begin{cases} 1 & n \in B \\ 0 & n \notin B \end{cases}$

Úplný popis náhodnosti je dán hodnotami
$(P_X(B),\ B \in \sigma-$ $\underbrace{\text{algebry obsahující všechny otevřené intervaly}}_{\text{borelovská}})$

Toto je velice složité & nepřehledné.

Naštěstí stačí distribuční funkce

$F_X(a) = P[X \le a] = P_X(-\infty,a] \quad F_X\ \mathbb{R} \to [0,1]$

Znáte-li F_X, pak lze jednoznačně určit P_X.

Tvrzení (vlastnosti F_X): Mějme $(\Omega, \mathcal{F}, P)$ a náh. veličinu X a její distribuční funkci F_X. Pak

i) F_X je neklesající a zprava spojitá.

ii) $\lim_{a \to -\infty} F_X(a) = 0$, $\lim_{a \to \infty} F_X(a) = 1$

Dk $a<b$ $F_X(a) = P(X\le a) = P[\omega : X(\omega)\le a]$ $\{\omega : X(\omega)\le a\} \subset \{\omega : X(\omega)\le b\}$ $\le P[\omega : X(\omega)\le b] = F_X(b)$

Pozn. $\{\omega : X(\omega)\le a\} = \bigcap_{n=1}^{\infty} \{\omega : X(\omega)\le a_n\}$

$a_n \to a+$

$$F_X(a_n) = P[\omega : X(\omega)\le a_n] = P[\omega : X(\omega)\le a] = P[\bigcap_{n=1}^{\infty}\{\omega : X(\omega)\le a_n\}]$$

$$\underbrace{P(\{\omega : X(\omega)\le a\}]}_{F_X(a)} = P[\bigcap_{n=1}^{\infty}\{\omega : X(\omega)\le a_n\}] = \lim_{N\to\infty} P[\bigcap_{n=1}^{N}\{\omega : X(\omega)\le a_n\}] =$$

$$= \lim_{N\to\infty} P[\{\omega : X(\omega)\le a_N\}] = \lim_{N\to a} F_X(a_N)$$

Spojitost zleva neplatí, protože

$$\{\omega : X(\omega)<a\} \ne \bigcup_{n=1}^{\infty}\{\omega : X(\omega)\le a_n\} = \{\omega : X(\omega)<a\}$$

Takže $\boxed{F_X(a) - \lim_{a_n\to a^-} F_X(a_n)} = P[X=a]$ $\to$ velikost skoku F_X v a

Důkaz ii) $F_X(a) = P[X\le a]$

$$\lim_{a\to-\infty} F_X(a) = P[\bigcap_{n=1}^{\infty}\underbrace{\{\omega : X(\omega)\le -n\}}_{=\emptyset}] = 0$$

$$\lim_{a\to\infty} F_X(a) = P[\bigcap_{n=1}^{\infty}\underbrace{\{\omega : X(\omega)\le n\}}_{\Omega}] = 1$$

Věta: Buď F funkce splňující i a ii z 1. Pak existuje $(\Omega, \mathcal{F}, P)$ pravděp. prostor a náh. veličina X, tak, že F je její distribuční funkce.

Důkaz:

Volme $\Omega = \mathbb{R}$, $\mathcal{F}$ σ-algebru nad $\forall$ otevř. množ. a $P(-\infty, a] = F(a)$. Náhodná veličina $X(\omega)=\omega$ má distrib. funkci $P[\{\omega, X(\omega)\le a\}] = P[\{\omega : \omega\le a\}] =$
$= P(-\infty, a] = F(a)$.

Def.: Diskrétní náhodná veličina X je taková, pro kt. $\exists$ nejvýše spočetná množina $S \subset \mathbb{R}$ splňující $P[X \in S] = 1$
$\{\omega : X(\omega) \in S\} = \Omega$. X je celočíselná náh. veličina (CNV), pokud X je diskrétní a $S = \mathbb{Z}$

Tvrzení: Buď X CNV. Pak její rozdělení je plně popsáno HUSTOTOU $\{p_z, z \in \mathbb{Z}\}$, kde $p_z = P[X = z]$

Platí: X je CNV $\Rightarrow$ $p_z \geq 0 \ \forall z \in \mathbb{Z}$, $\sum_{z \in \mathbb{Z}} p_z = 1$

$$\Omega = [X \in S] = \bigcup_{s \in S} [X = s]$$

$$\downarrow \qquad\qquad \downarrow$$

$$P[\,] = 1 \qquad \sum P[\,] = 1$$

Příklad: Bernoulliho náhodná veličina

máme pokus a ten dopadne $\boxed{\text{úspěšně}}$ / $\boxed{\text{neúspěšně}}$

$$(\Omega, \mathcal{F}, P) \quad X(\omega) = \begin{cases} 1 \text{ (úspěch)} \\ 0 \text{ (neúspěch)} \end{cases}$$

$p_0 \in (0,1)$ & $p_1 = 1 - p_0$ & $p_z = 0$ pro $z \neq 0,1$

Rozdělení $p_1 = p$
$p_0 = 1 - p \quad p \in (0,1)$

p je parametr tohoto rozdělení se nazývá alternativní s parametrem p.

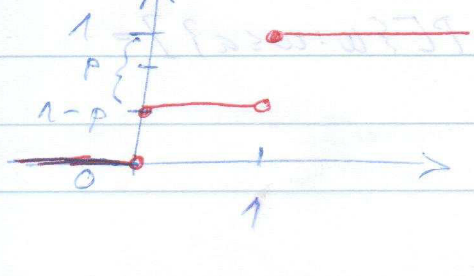
Distribuční funkce $F(a) = P[X \leq a]$

$a < 0 \quad P[X < a] = 0$

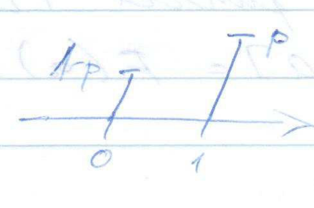
$0 \leq a < 1 \quad P[X < a] = P[X = a]$

$1 \leq a \quad P[X \leq a] = 1$

D.F.



HUSTOTA



Na jednom $(\Omega, \mathcal{F}, P)$ může být mnoho náh. veličin

$X: \Omega \to \mathbb{R}$ $\quad$ ~~X~~ Y: CNV

$Y: \Omega \to \mathbb{R}$

? sdružené rozdělení X a Y, resp.

$P[X=n, Y=k] \quad \forall n, k \in \mathbb{Z}$

$\parallel$

$P[\{\omega : X(\omega)=n\} \cap \{\omega : Y(\omega)=k\}]$

<u>Def:</u> CNV X a Y jsou nezávislé, pokud $\forall n, k \in \mathbb{Z}$ platí,

$$P[X=n, Y=k] = P[X=n]\, P[Y=k]$$

CNV $X_1 \dots X_n$ jsou vzájemně nezávislé, pokud $\forall I \subset \{1..N\}$ a pro $\forall n_i \in \mathbb{Z}$, $i \in I$ platí $P[\bigwedge_{i \in I} [X_i = n_i]] = \prod_{i \in I} P[X_i = n_i]$

Charakteristiky náhodné veličiny (MOMENTY náh. velič.):

STŘEDNÍ HODNOTA CNV:

<u>Def</u> CNV X má konečnou střední hodnotu, pokud

$$\sum_{n \in \mathbb{Z}} |n|\, P[X=n] = \sum_{n \in \mathbb{Z}} |n|\, p_n < \infty$$

a v tom případě $EX = \sum_{n \in \mathbb{Z}} n \cdot p_n$

Střední hodnota (Expectation)

<u>Věta</u> Pro CNV X a Y s konečnou EX, EY platí $E(X+Y) = E(X) + E(Y)$

<u>Důkaz</u>

$$E(X+Y) = \sum_{n \in \mathbb{Z}} n \,\cancel{p_n}\, P(X+Y=n) = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} (k+l)\, P(X=k, Y=l) =$$

$$= \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} k\, P(X=k, Y=l) + \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} l\, P(X=k, Y=l) =$$

$$= \sum_{k \in \mathbb{Z}} k \underbrace{\sum_{l \in \mathbb{Z}} P(X=k, Y=l)}_{\text{Podle věty o úplné pravděp.} = P[X=k]} + \sum_{l \in \mathbb{Z}} l \sum_{k \in \mathbb{Z}} P(X=k, Y=l) = \sum_k k\, P[X=k] + \sum_l l\, P[Y=l] =$$

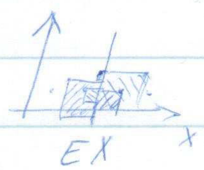
$$= EX + EY$$

indukcí $X_1 \dots X_n$ CNV $EX_i < \infty$ $\forall i$
$$E(\sum X_i) = \sum EX_i$$
Věta: Buď X CNV a $c \in \mathbb{R}$ $EX < \infty$ Pak $E(cX) = c \cdot EX$
Dk. $E(cX) = \sum_{n \in \mathbb{Z}} c \cdot n \cdot P[X = n] = c \cdot EX$

CNV = celočíselná náhodná veličina.
$P[X \in \mathbb{Z}] = 1$
Střední hodnota $EX = \sum_{i \in \mathbb{Z}} i P(X=i)$
Pokud $P[X \in A] = 1$
$P[X \in \{0, 1 \dots n\}] = 1$
Pokud $E|X_i| < \infty$ pro $i = 1 \dots n$, pak $E \sum_{i=1}^{n} v X_i = \sum_{i=1}^{n} v EX_i$
→ jde o váhy → EX je vážený průměr hodnot CNV X, vážený pravděpodobností.
Střední hodnota má tyto významy:
1. Pro rozdělení náh. velič. X jde o charakt. „polohy"
$E(X+a) = EX + a$
X a Y mohou mít stejnou stř. hodnotu, ale naprosto různá rozdělení.
→ Hod kostkou ⇝ $P[X=i] = 1/6$ $EX = 7/2$
Další charakteristiky náh. veličiny.
Def: Momentem náhodné veličiny rozumíme hodnotu $EX^p = \sum_{i \in \mathbb{Z}} i^p P[X=i]$ má-li tento součet smysl pro dané $p \in \mathbb{N}$
EX je první moment náh. veličiny X.
Def: Centrálním momentem náhodné veličiny rozumíme hodnotu
$E(X - EX)^p = \sum_{i \in \mathbb{Z}} (i - EX)^p P[X=i]$ má-li pravá strana smysl.
Def: Pokud $EX^2 < \infty$, pak $E(X - EX)^2$ nazveme rozptylem náhodné velič. X.

Je-li $EX^2 = \infty$, pak buď EX neexistuje$_{\nexists}$ → ani rozptyl $\nexists$,
$P[X=i] = k / i^{1+\varepsilon}$ $i = 1, 2, \dots$ $\sum_{i=1}^{\infty} 1 / i^{1+\varepsilon} < \infty$
$P[X=-i] = k / i^{1+\varepsilon}$ $i = 1, 2, \dots$ $EX = \underbrace{\sum_{i=1}^{\infty} \frac{k \cdot i}{i^{1+\varepsilon}}}_{\infty} - \underbrace{\sum_{i=1}^{\infty} \frac{k \cdot i}{i^{1+\varepsilon}}}_{-\infty} = \infty - \infty$

NEBO $E|X| < \infty \Rightarrow$ n.v. X má nekonečný rozptyl.



Rozptyl charakterizuje variabilitu náhodné veličiny okolo její střední hodnoty

~> STŘEDNÍ ČTVERCOVÁ ODCHYLKA

(E) $(\quad)^2$ $(X-EX)$

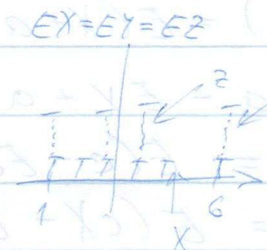
Př: $P[X=i] = 1/6$ $\quad i=1\dots 6$ $\quad EX = 7/2$ (1)

$P[Y=j] = 1/2$ $\quad i = 1, 6$ $\quad EY = 7/2$ (2)

(1) ROZPTYL = var $X = E(X-EX)^2 = \sum_{i=1}^{6} (i-7/2)^2 \cdot 1/6 = 1/3 \sum_{i=1}^{3} (i-(7/2))^2 =$

$= (1/3)((5/2)^2 + (3/2)^2 + (1/2)^2) = 35/12$

(2) $E(Y-EY)^2 = (\frac{5}{2})^2 = 75/12.$

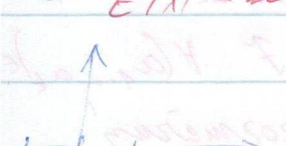


var $Z <$ var $X <$ var Y

Jensenova nerovnost.

Obecně neplatí $Ef(X) = f(EX)$

Věta: Buď X náh. veličina a f konvexní funkce takové, že $E|X| < \infty$ a $E|f(X)| < \infty$. Pak $Ef(X) \geq f(EX)$.



$P[X=-1] = 1/2$ $\quad EX = 0$

$P[X=1] = 1/2$ $\quad (EX)^2 = 0 \Rightarrow (EX)^2 \leq EX^2$

$EX^2 = 1$

Lemma: Je-li f dvakrát diferencovatelná, pak $f''(x) \geq 0$ $\forall x \Leftrightarrow f$ je konvexní

Dk věty

Z Taylorova rozvoje f kolem EX $\exists c$ takové, že

$f(x) = f(EX) + (x-EX) f'(EX) + \frac{(x-EX)^2}{2} \cdot f''(c) \geq f(EX) + (x-EX) \cdot f'(EX)$ ~> dosadit náh. veličinu + průměrovat.

$Ef(X) \geq E\boxed{f(EX)} + \underbrace{E(X-EX)}\boxed{f'(EX)}$

konstanta konstanta

$$E(X-EX) = EX - EX = 0$$

$\Rightarrow Ef(X) \geq Ef(EX)$.

... důkaz pro f obecnou vychází z faktu

$f = \sup\{a : a \leq f,\ a$ je afinní$\}$

$\operatorname{var} X = E(X-EX)^2 = E(X^2 - 2XEX + (EX)^2) = EX^2 - 2E(XEX) +$
$E(EX)^2 = EX^2 - 2(EX)(EX) + (EX)^2 = \underline{EX^2 - (EX)^2 \geq 0}$

Základní vlastnosti rozptylu

$\operatorname{var}(X+a) = E(X+a-E(X+a))^2 = E(X+a-E(X+a)^2) =$
$= E(X+a-EX+a)^2 = E(X-EX)^2 = \underline{\operatorname{var} X}$.

$\operatorname{var}(c) = E(c-Ec)^2 = \underline{0}$

$\operatorname{var}(cX) = E(cX - EcX)^2 = Ec^2(X-EX)^2 = c^2 \cdot \operatorname{var} X$.

$\operatorname{var}(X+Y) = E(X+Y-E(X+Y))^2 = E(X-EX+Y-EY)^2 = E[(X-EX)^2 + 2E(X-EX)$
$\cdot (Y-EY)] + E(Y-EY)^2 = \operatorname{var} X + \operatorname{var} Y + 2E(X-EX)(Y-EY)$

$\Rightarrow \operatorname{var}(X+Y) = \operatorname{var} X + \operatorname{var} Y + 2E((X-EX)(Y-EY))$

<u>Def</u>: Buď $(\Omega, \mathcal{F}, P)$ pravděp. prostor. Zobrazení $X: \Omega \to \mathbb{R}^k$ ($k \geq 2$) splňující $\{\omega : \bigcap\{\omega : X_i(\omega) \leq a_i\}\ i = 1..k\} \in \mathcal{F}\ \forall (a_1, \dots a_k) \in \mathbb{R}^k$ (Pozn. $\underset{\sim}{X}(\omega) = (X_1(\omega), \dots, X_k(\omega))$) nazveme k-rozměrný náhodný vektor.

Pozor.: $\forall$ složka náh. vektoru $\underset{\sim}{X}$ je náhodnou veličinou

<u>Def</u>: (ROZDĚLENÍ NÁH. VEKTORU)

Pravděpodobnost $P_{\underset{\sim}{X}}$ def. na $\mathbb{R}^k$ předpisem $P_{\underset{\sim}{X}}[\prod_{i=1}^{k}(-\infty, a_i]] =$

$= \sout{P\{\omega} \; P[\bigcap_{i=1}^{k}(X_i \leq a_i)] = P\{\omega : \bigcap_{i=1}^{k}\{\omega : X_i \leq a_i\}\}$

se nazývá SDRUŽENÉ ROZDĚLENÍ NÁH. VEKTORU X.

hod dvěma kostkami

$X_1 =$ součet ok

$X_2 =$ abs. hodnota rozdílu

$x_1 = 2 \quad x_2 = 0 \quad \rightsquigarrow 1/36$

$X_1 = 3 \quad X_2 = 1 \quad 2/36$

$X_1 = 4 \quad X_2 = 2 / 0$

$1/36 \quad 2/36$

Def:

Pro podvektor $Y = (X_{i_1} \dots X_{i_d})$, $d < k$ náh. vektoru $\underset{\sim}{X}$ definujeme marginální rozdělení $\underset{\sim}{X}$ jako

$$P_{\underset{\sim}{Y}}\left(\mathop{\times}_{j=1}^{d}(-\infty, a_{i_j}]\right) = P\{\omega : \bigcap_{j=1}^{d}\{\omega : X_{i_j} \le a_{i_j}\}\} =$$

$$= P\{\omega : \bigcap_{i=1}^{k}\{\omega : X_i \le a_i\}, \text{ kde } a_i = \infty \text{ pro } i \ne i_1 \dots i_d\}$$

Pozn. $\{\omega : X_i \le \infty\} = \Omega$. (=> nic nevbyde)

Speciálně $P_{X_i}(-\infty, a_i] = P\{\omega : X_i \le a_i\}$

SDRUŽENÉ ROZDĚLENÍ jednoznačně určuje marginální.

MARGINÁLNÍ rozdělení se do sdruženého dají složit nekonečně způsoby.

Náhodný vektor:

$\underset{\sim}{X} = (X_1 \dots X_k) : \Omega \to \mathbb{R}^k$

$\underset{\sim}{P}(X_1 \le a_1, \dots, X_k \le a_k) = P\{\omega; \bigcap_{i=1}^{k}(X_i(\omega) \le a_i)\}$

$P_X\left(\mathop{\times}_{i=1}^{k}(-\infty, a_i]\right) \dots$ SDRUŽENÉ ROZDĚLENÍ

Marginální rozděl: je rozdělení podvektoru náh. vektoru $\underset{\sim}{X}$, např. rozdělení $(X_1 \dots X_{k-1})$, a to dostaneme z $\underset{\sim}{X}$ volbou $a_k = \infty$

Nejdůležit. (nejpoužív.) jsou ~~roz~~ rozdělení jednotlivých složek $X_1, X_2 \dots X_k$.

Př. 1.) (X, Y) ... hážime 2ma kostkami:

X je ~~výsled~~ výsledek na 1. kostce

Y součet na oboch kostkách.

X \ Y	2	3	4	5	6	7	8	9	10	11	12	
1	•	•	•	•	•	•						1/6
2		•	•	•	•	•	•					1/6
3			•	•	•	•	•	•				1/6
4				•	•	•	•	•	•			1/6
5					•	•	•	•	•	•		1/6
6						•	•	•	•	•	•	1/6

$(6,12) = 1/36$

$\bullet = 1/36$

⤳ chcem marg. rozdel. $X \Rightarrow Y$ uvolním a sčítam: ($\forall$ 1/6)

$$P(X=j) = \sum_{i=2}^{12} P(X=j, Y=i) = 6 * 1/36 = 1/6 \quad (\text{pre } \forall j)$$

⤳ chcem marg. rozdel. $Y \Rightarrow X$ uvolním $\Rightarrow$

$$P(Y=j) = \sum_{i=1}^{6} P(Y=j, X=i) \quad \leadsto \text{podľa} \uparrow$$

Marginální ⤳ $P(X=5) = 1/6$, $P(Y=6) = 5/36$

Združené ⤳ $P(X=5, Y=6) = \underline{1/36}$.

2.) 3 kostky, X výsl. na 1., Y součet na 2. a 3.

	2	3	4	5	6	7	8	9	10	11	12
1	(1,1,1) •	(1,1,2),(1,2,1) ••	3•								
2	(2,1,1) •	••	3•								
3	•	••	3•	4•	5•	6•	5•	4•	3•	2•	1•
4	•	••	3•								
5	•	••	3•								
6	•	••	3•								

$\bullet = 1/216$

$\overset{!!!}{\underset{m}{\forall}}$ $P(X=m, Y=n) = P(X=m)\,P(Y=n)$ ⤳ Nezávislé.

$Y=3$ ⤳ $(1,1,2)$ alebo $(1,2,1)$

1/36 2/36 3/36 4/36 ...

+ stále $P(X=i) = 36/216 = 1/6$ ✓

ale $P(Y=i)$ je iná ⤳

3/36 2/36 1/36.

Značenie: CNV = celočíselný náhodný vektor = náhodný vektor, jehož složky jsou celočíselné náh veličiny.

Def: CNVeličiny X a Y jsou nezávislé, pokud $\forall m, n \in \mathbb{Z}$ platí

$$P(X=m, Y=n) = P(X=m)\,P(Y=n)$$

neboli sdružené rozdělení je součinem marginálních.

Náhodné veličiny X, Y jsou nezávislé, pokud sdružené rozděl. (X,Y) je součinem marg. rozdělení X a Y.

$\rightsquigarrow$ Všeobecně?

Def (Nezávislost CNV obecně):

CNVeličiny $X_1 \dots X_k$ se nazývají (vzájemně) nezávislé, pokud $\forall m_i \in \mathbb{Z}$; $i = 1..k$ platí

$$P(X_1 = m_1, \dots, X_k = m_k) = \prod_{i=1}^{k} P(X_i = m_i)$$

$\rightsquigarrow$ Teď lze zadefinovat podmíněné rozdělení.

Def: Buď (X,Y) CNV. Rozdělení (k pevné)

$$P[X=k \mid Y=l] = \frac{P[X=k, Y=l]}{P[Y=l]} \quad ; \; P[Y=l] > 0, \; k \in \mathbb{Z}$$

nazveme podmíněné rozdělení X za podmínky $Y=l$

$\rightsquigarrow$ spät k #2 kockám (vyrezané 2. a 5. stĺpec)

Př:

$X \backslash Y$	2	5
1	•	•
2		•
3		•
4		•
5		
6		

$$P[X=1 \mid Y=2] = \frac{P[X=1, Y=2]}{P[Y=2]} = \frac{1/36}{1/36} = 1$$

$$P[X=3 \mid Y=5] = \frac{P[X=3, Y=5]}{P[Y=5]} = 1/4$$

$$P[X=1 \mid Y=5] = 1/4$$

$$P[X=1 \mid Y=8] = 0/\tfrac{5}{36} = 0$$

+ toto umožňuje podmieniť samo seba $\rightsquigarrow$

$$P[X=k \mid X=l] \rightsquigarrow \delta_{kl} = \begin{cases} 1 & k=l \\ 0 & k \neq l \end{cases}$$

Lemma: Podmíněné rozdělení je opravdu rozdělení.

Dk: ? co vlastně máme ověřit?

- $\forall k \in \mathbb{Z}$ je $P[X=k \mid Y=l] \geq 0$
- $\sum_{k\in\mathbb{Z}} P[X=k \mid Y=l] = 1$

Prvé: $\left.\begin{array}{l}\dfrac{P[X=k,\, Y=l] \geq 0}{P[Y=l] \quad > 0}\end{array}\right\} \geq 0$ ✓

Druhé:

$$\sum_{k\in\mathbb{Z}} \frac{P[X=k,\, Y=l]}{P[Y=l]} = \left(\sum_{k\in\mathbb{Z}} P[X=k,\, Y=l]\right)\cdot \frac{1}{P[Y=l]} =$$

$$= \frac{1}{P[Y=l]}\cdot P\Big[\underbrace{\bigcup_k [X=k]}_{\Omega} \cap [Y=l]\Big] = \frac{1}{P[Y=l]} P[Y=l] = 1 \quad ✓ \quad \blacksquare$$

Připomeňme:

$$var(X+Y) = E(X+Y-E(X+Y))^2 = \underbrace{var X}_{E(X-EX)^2} + \underbrace{var Y}_{E(Y-EY)^2} + \underbrace{2E[(X-EX)(Y-EY)]}_{???}$$

Def: Buď (X,Y) náhodný vektor takový, že má konečné druhé momenty složek (t.j. $EX^2<\infty$, $EY^2<\infty$).

- Kovariace (znač. $cov(X,Y)$):

$$cov(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - EX\cdot EY$$

- Korelace (znač. $\rho_{X,Y}$)

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var X \, var Y}}$$

- Nenahradí sdružené rozdělení, ale alespoň o něm něco dá.

≈ říká, jak se chovají kladné/záporné a malé/velké X, Y.

? Proč potřebujeme náh. vektor? ⇝ Abychom měli sdružené rozdělení a uměli spoč. $E(XY)$

PRO CNV: $E(XY) = \sum_{r\in\mathbb{Z}}\sum_{s\in\mathbb{Z}} r\cdot s\cdot P[X=r, Y=s]$

Základní vlastnosti korelace a kovariace:

Věta: Buď (X,Y) náhodný vektor. Pak:

1. Jsou-li X a Y nezávislé, pak $cov(X,Y)=0$, čemuž říkáme, že X a Y jsou NEKORELOVANÉ.

! Z $cov(X,Y)=0$ <u>neplyne</u> nezávislost X a Y !

2. $cov(aX+b, Y) = a\cdot cov(X,Y)$.
3. $cov(X+Y, Z) = cov(X,Z) + cov(Y,Z)$

Dk: 1.) pro CNV:

$$E(XY) = \sum_r\sum_s rsP[X=r, Y=s] \overset{\text{Nezávislost}}{=} \sum_r\sum_s rs\,P[X=r]P[Y=s] =$$

$$= \underbrace{\sum_r r\,P[X=r]}_{EX}\ \underbrace{\sum_s s\,P[Y=s]}_{EY} = EX\cdot EY \Rightarrow cov(X,Y)=0$$

! Ale jen 1 směrem ! (Nezávislost není obousměrná)

Př: $P[X=i] = 1/3 \quad i=-1,0,1$

$Y = X^2$.

$P[X=0, Y=0] = 1/3$

$P[X=1, Y=1] = 1/3$

$P[X=-1, Y=1] = 1/3$

$EXY = 0\cdot 0\cdot 1/3 + 1\cdot 1\cdot 1/3 + (-1)\cdot 1\cdot 1/3 = 0 = EX\cdot EY$

$\Rightarrow EXY - EXEY = 0$, ale nejsou nezávislé.

2.) $cov(aX+b, Y) = E[(aX+\cancel{b} - \underbrace{E(aX+b)}_{aEX+\cancel{b}})\cdot(Y-EY)] = E(a(X-EX)(Y-EY)) = a\cdot E[(X-EX)(Y-EY)]$

3.) obdobně jako 2.)

Pozorování:

- $cov(X,X) = var X \quad \ldots \quad E(X-EX)^2 = E[(X-EX)(X-EX)]$
- $var(X+X) = var X + var X + 2cov(X,X) = 4 var X \{= var(2X)\}$
- je-li $cov(X,Y) > 0$, pak $var(X+Y) > var X + var Y$
  < 0, pak $var(X+Y) < var X + var Y$
- $var(X-Y) = E(X-Y-E(X-Y))^2 = var X - 2cov(X,Y) + var Y > 0.$
- jsou-li X a Y nekorelované, pak $var(X+Y) = var(X) + var(Y)$.
- $var\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} var X_i + \sum_{i \neq j} cov(X_i, X_j)$
- $cov(X,Y) = cov(Y,X)$.

Opak:

sdružené a marginální rozdělení:

↳ Podmíněné rozdělení

→ Nezávislost

→ Kovariance, korelace.

(X,Y) CNV

$cov(X,Y) = E[(X-EX)(Y-EY)]$, pokud $EX^2 < \infty$, $EY^2 < \infty$

základní vlastnosti:

- $cov(aX+b, Y) = a \cdot cov(X,Y)$
- $cov(X,Y) = cov(Y,X)$
- $cov(X,X) = var X$
- $cov(X+Y, Z) = cov(X,Z) + cov(Y,Z)$
- $cov(a, X) = 0$ (lebo $(a - Ea) = 0$ pre $\forall a \in \mathbb{R}$).

Korelace ≡ normovaná ~~koreltac~~ kovariace

$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var X \cdot var Y}}$ $\quad 0 < var X, var Y < \infty$ (≈ aby šlo dělit)

Tvrzeni: (zákl. vlastn. $\rho_{X,Y}$)

- $\rho_{aX+b,Y} = \rho_{X,Y} \cdot \operatorname{sgn}(a)$
- $\rho_{X,Y} = \rho_{Y,X}$
- $\rho_{X,X} = 1$
- $|\rho_{X,Y}| \leq 1$
- $|\rho_{X,Y}| = 1 \iff \exists a \neq 0, b \in \mathbb{R} : P\{\omega : Y(\omega) = a X(\omega) + b\} = 1$

+ zjednodušene $P(Y = aX+b) = 1$. (neformálne).

+ dá sa upresniť $\exists a > 0$ a $\exists a < 0 \Rightarrow$ priama/nepriama úmer.

Viac-menej korelace $\equiv$ míra (lineární) závislosti.

Dk:

1. $\rho_{aX+b,Y} = \dfrac{\operatorname{cov}(aX+b,Y)}{\sqrt{\operatorname{var}(aX+b)\operatorname{var}Y}} = \dfrac{a \cdot \operatorname{cov}(X,Y)}{\sqrt{a^2 \operatorname{var}X \operatorname{var}Y}} = \rho_{X,Y} \cdot \operatorname{sgn}(a)$

2. (zjavné, $\forall$ je symetrické), 3. po rozp. def. $= \dfrac{\operatorname{var}X}{|\operatorname{var}X|}$ & $\operatorname{var}X \geq 0$

4. $\rho_{X,Y} = \dfrac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}X \operatorname{var}Y}} = \dfrac{\sum_{i,j\in\mathbb{Z}} (i-EX)(j-EY) \cdot P[X=i, Y=j]}{\sqrt{\sum_{i\in\mathbb{Z}} (i-EX)^2 P[X=i] \cdot \sum_{j\in\mathbb{Z}} (j-EY) P[Y=j]}} =$

$=$ (ak majú $\forall \sum$ zmysel) $\Rightarrow$ (Cauchy-Schwarz)

$$\left(\sum\sum (i-EX)(j-EY) P[X=i, Y=j]\right)^2 \leq \underbrace{\sum_i \sum_j (i-EX)^2 P[X=i, Y=j]}_{= \sum_i (i-EX)^2 P[X=i] = \operatorname{var}X} \cdot$$

$$\cdot \underbrace{\sum_i \sum_j (j-EY)^2 P[X=i, Y=j]}_{= \sum_j (j-EY)^2 P[Y=j] = \operatorname{var}Y}$$

5. ($\rho_{X,Y} = 1$) $\iff \exists a, b : Y = aX+b$. ($\Leftarrow$)

Nechť $Y = aX+b$.

pak
- $\operatorname{cov}(X,Y) = \operatorname{cov}(X, aX+b) = a \cdot \operatorname{cov}(X,X)$
- $\operatorname{var}Y = \operatorname{var}(aX+b) = a^2 \cdot \operatorname{var}(X)$

$\Rightarrow \rho_{X,Y} = \dfrac{a \cdot \operatorname{cov}(X,X)}{\sqrt{a^2 (\operatorname{var}X)^2}} = \dfrac{a}{|a|} = \operatorname{sgn}(a)$

Nechť $\rho_{X,Y} = 1$ $(\Rightarrow)$

Vezmeme matici

$$\begin{pmatrix} \operatorname{var} X & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & \operatorname{var} Y \end{pmatrix} = \begin{pmatrix} \operatorname{var} X & \sqrt{\operatorname{var} X \operatorname{var} Y} \\ \sqrt{\operatorname{var} X \operatorname{var} Y} & \operatorname{var} Y \end{pmatrix}$$

$\rightsquigarrow$ determinant: $0 \Rightarrow$ singulární $\Rightarrow$ 1 řádek lin. komb. 2.

~~($\exists \sqrt{\operatorname{var} X} = \sqrt{\operatorname{var} Y} \cdot$ koeficient)~~

resp. $\exists$ konstanty c, d takové, že $(c,d) \neq (0,0)$.

$$(\cancel{\exists}\, c, d) \begin{pmatrix} \operatorname{var} X & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & \operatorname{var} Y \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = 0.$$

$$\Rightarrow c^2 \operatorname{var} X + 2cd \operatorname{cov}(X,Y) + d^2 \operatorname{var} Y = \operatorname{var}(cX + dY) = 0$$

$\rightsquigarrow$ je-li rozptyl náh. veličiny 0, pak je tato veličina konstantní (s pravděp. 1) $\Rightarrow \exists k: cX + dY = k$

$\Rightarrow Y = aX + b$ ∎

Podmíněná střední hodnota

(X,Y) CNV podm. za podmínky $Y = j$

$$P[X = i \mid Y = j] = \frac{P[X=i, Y=j]}{P[Y=j]} \quad i \in \mathbb{R} \quad \text{je-li } P[Y=j] > 0.$$

Def: Buď (X,Y) CNV, pro ~~?~~ který existuje EX. Def.

$$E[X \mid Y = j] = \sum_{i \in \mathbb{Z}} i \cdot P[X = i \mid Y = j], \text{ pokud } P[Y=j] > 0.$$

Věta: (vztah EX a $E[X \mid Y=j]$):

Nechť (X,Y) CNV, EX existuje & $< \infty$. Pak

$$EX = \sum_{j \in J} P[Y=j]\, E[X \mid Y=j], \text{ kde}$$

J je množina těch j, pro než $P[Y=j] > 0$.

D.k: výpočtem:

$$EX = \sum_i i \cdot P[X=i]$$

$$P[X=i] = \sum_j P[X=i, Y=j] = \sum_j P[X=i \mid Y=j] \cdot P[Y=j]$$

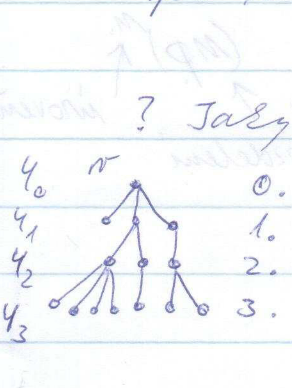
$$EX = \sum_i i \sum_j P[X=i \mid Y=j]\, P[Y=j] = \sum_{j\in J} P[Y=j] \underbrace{\sum_i i\, P[X=i \mid Y=j]}_{E[X \mid Y=j]} =$$

$$= \sum_{j \in J} P[Y=j]\, E[X \mid Y=j] \quad \blacksquare$$

Př:

výpočet v, kt. spustíme a on v svém průběhu volá X-krát sebe, kde X jsou nezávislé náh. veličiny, pro něž $P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$ (Binom. rozdel.)
$k = 0, 1 \dots n$

? Jaký je střední počet spuštěných výpočtů v?



Y_i = počet v i-úrovni

$E Y_i$ nie je jednoduché spočítať, ale dá sa určiť $E[Y_3 \mid Y_2 = j]$

$Y_n \dots$ jaká je $E[Y_n \mid Y_{n-1} = x_{n-1}] = ?$

$$E[Y_n \mid Y_{n-1} = x_{n-1}] = \sum_i i\, P[Y_n = i \mid Y_{n-1} = x_{n-1}]$$

$\hookrightarrow$ # procesů o 1 $\uparrow$

$$+ \quad E Y_1 = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

x_{n-1} je skutečný: $\rightsquigarrow$ Podmínka říká, že Y_{n-1} NENÍ V TUTO CHVÍLI NÁHODNÉ, ale jde o pevné č. $\approx x_{n-1}$

$P_1 \; P_2 \dots P_{Y_{n-1}}$

$$\Rightarrow \;=\; \sum_{i=0}^{\substack{n\cdot X_{n-1}\\ \cancel{n(n-1)}}} i\, P\Big[\sum_{j=1}^{X_{n-1}} Z_j = i \;\Big|\; Y_{n-1} = X_{n-1}\Big]$$

kde Z_j jsou nezávislé náh. veličiny s rozdělením
$P[Z_j = k] = \binom{n}{k} p^k (1-p)^{n-k}$ pro $k = 0, 1, \dots, n$.

+ jestli Z_j jsou nezávislé s Y_{n-1} } podmíněná pravděp.
Σ takových Z_j se stává nepodmíněnou

$$E[Y_n \mid Y_{n-1} = X_{n-1}] = \sum_{i=0}^{nX_{n-1}} i\, P\Big[\sum_{j=1}^{X_{n-1}} Z_j = i\Big] = E\Big(\sum_{j=1}^{X_{n-1}} Z_j\Big) =$$

$$= \sum_{j=1}^{X_{n-1}} E Z_j = \boxed{X_{n-1}\cdot n\cdot p}$$

Počítáme $E Y_n = \sum_j P[Y_{n-1} = j]\, E[Y_n \mid Y_{n-1} = j] = \sum_j P[Y_{n-1} = j]\cdot jnp =$

$= np \sum_j j\, P[Y_{n-1} = j] = np\, E Y_{n-1} = np\, np\, E Y_{n-2} = (np)^n$ ← úroveň, ↑ v rozdělení

$\sum_{k=0}^{\infty} (np)^k \;\cancel{=}\; \dots \quad np < 1 \quad 1/(1-np)$

$\qquad\qquad\qquad\qquad np \geq 1 \quad \infty$

Opakování: $E[X \mid Y = y] = \sum_k k\, P[X = k \mid Y = y]$

X, Y nezávislé $\Rightarrow E[X \mid Y = y] = EX$.

$P[X = x \mid Y = y] = P[X = x] \;\; \forall x, y \Leftrightarrow X, Y$ nezávislé

$E[X \mid Y = y] = EX \;\; \forall y \Leftrightarrow X, Y$ nezávislé

Všimněme si, že $E[X \mid Y = y]$ je funkcí y

Def: $E[X \mid Y]$ je náhodná veličina, kt. na množině $\{\omega : Y(\omega) = y\}$ nabývá hodnoty $E[X \mid Y = y]$

Tedy $E[X \mid Y] : \Omega \to \mathbb{R}$ náh. veličina

$E[X \mid Y](\omega) = E[X \mid Y = y]$ pokud $Y(\omega) = y$

$E[X \mid Y] = f(Y)$ pro nějakou $f : \mathbb{R} \to \mathbb{R}$.

Věta: $E[E[X|Y]] = EX$, mají-li obě strany smysl.

Dk: $E(E[X|Y]) = E f(Y) = \sum_k f(k) P[Y=k] = \sum_k E[X|Y=k] P[Y=k] =$
$= EX$, viz věta z min. přednášky.

Modely Rozdělení Diskrétních Náh. Veličin

1. Alternativní rozdělení: ($Alt(p)$)

Bernoulliho pokus:

N.V. X nabývá pouze 2 hodnot $X \in \{\underset{\text{neúspěch}}{0}, \underset{\text{úspěch}}{1}\}$

$P[X=1] = p$ pro parametr $p \in (0,1)$.

$P[X=0] = 1-p$

$\boxed{\begin{aligned} EX &= p \\ EX^2 &= p \end{aligned}}$ $\Rightarrow \operatorname{var} X = p - p^2 = p(1-p)$

2. Binomické rozdělení ($Bi(n,p)$)

- Začneme opakovat (nezávisle*) bernoulliho pokusy
- Dostaneme posloupnost 0 a 1. $\Rightarrow$ 0101101

$(1-p)p(1-p)pp(1-p)p$

$\Rightarrow$ Otázka $\leadsto$ Jaký je počet 1? (úspěchů v n pokusech)

$Y = \sum_{i=1}^{n} X_i =$ počet úspěchů

$P[Y=k] = \binom{n}{k} p^k (1-p)^{n-k}$ ... BINOMICKÉ ROZDĚLENÍ.

(pozn. $1 = (p + (1-p))^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$)

$\boxed{\begin{aligned} EY &= np \\ \operatorname{var} Y &= np(1-p) \end{aligned}}$

3. Geometrické rozdělení ($Geom(p)$)

$\boxed{EY = \frac{1-p}{p}}$

... Rozdělení čekání na 1. úspěch

$X_1, X_2, X_3 \dots$ nezávislé n.v. ~~Alternující~~ $Alt(p)$

$Y = \min\{i : X_i = 1\} - 1$

$P[Y=k] = (1-p)^k \cdot p$ $k = 0, 1, 2, \dots$

$\frac{1}{p} = \frac{1}{1-(1-p)} = \sum_{k=0}^{\infty} (1-p)^k$ & $1 = \sum_{k=0}^{\infty} (1-p)^k \cdot p$

4. Negativně binomické $NB(k, p)$ $\boxed{EY = k \cdot \frac{1-p}{p}}$

\# neúspěchů před k-tým úspěchem

$Y = \sum_{i=1}^{k} Y_i$, kde Y_i jsou nezávislé n.v. Geom.(p)

$P[Y=j] = \binom{j+k-1}{j} (1-p)^j p^k \rightsquigarrow j \geq 0, \; k \geq 1, \; j+k$ pokusů

· U binomického rozdělení známe předem # pokusů a neznámý je # úspěchů.

U neg. binom. víme, kolik bude úspěchů ale náhodný je počet pokusů

5. Poissonovo rozdělení $Po(\lambda)$

$$\boxed{\begin{array}{l} EY = \lambda \; (= \lim n p_n) \\ \operatorname{var} Y = \lambda \; (= \lim n p_n (1-p_n)) \end{array}}$$

mějme $Bi(n, p_n)$ tak, že $n p_n \xrightarrow[n \to \infty]{} \lambda > 0$

motivace: 1 atom, za určitou dobu se rozpadne s p & počítám, kolik se rozpadne za $1/k$ sekundy (na zač. n)

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$$

$$\frac{n!}{(n-k)!\,k!} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} p_n^k \cdot (1-p_n)^n \cdot (1-p_n)^{-k} =$$

$\Rightarrow$ pozor. $\left.\begin{array}{l} k \text{ pevné} \\ n \to \infty \end{array}\right\}$ jestliže $n p_n \to \lambda \Rightarrow p_n \to 0 \Rightarrow$

$$= \frac{1}{k!} \underbrace{\left[\frac{n}{n} \cdot \frac{(n-1)}{n} \cdot \ldots \cdot \frac{n-k+1}{n}\right]}_{\downarrow \; 1} \underbrace{\left[n^k \cdot p_n^k\right]}_{\downarrow \; \lambda^k} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\downarrow \; e^{-\lambda}} \cdot \underbrace{\left(\frac{1-p_n}{1-\frac{\lambda}{n}}\right)^n}_{\downarrow \; 1} \cdot \underbrace{(1-p_n)^{-k}}_{\downarrow \; 1}$$

$$\Rightarrow P[X_n = k] \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{pro } k = 1, 2, \ldots$$

$$P(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!} \Rightarrow Y \text{ má Poissonovo rozdělení } Po(\lambda)$$

6. Hypergeometrické rozdělení

⇝ Binomické (s racionálním p) $p = l/m$

l bílých (úspěchů) $m-l$ černých (neúspěch) + vytahujeme a vždy vrátíme & zamícháme.

⇝ Hypergeometrické ⇝ nevracíme:

N míčků, M bílých a vytáhneme n míčků, bez vracení. X je počet vytažených bílých míčků.

$$P(X=k) = \frac{\text{příznivé}}{\text{všechny}} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad \begin{array}{l} 0 \le k \le n \\ 0 \le k \le M \\ 0 \le n-k \le N-M \end{array}$$

Pokud $N \to \infty$, $\frac{M}{N} \to p \in (0,1)$ a n je konstantní, pak

$$P(X=k) \to \binom{n}{k} p^k (1-p)^{n-k} \qquad \boxed{EX = n \cdot \frac{M}{N}}$$

jde k binom. rozdělení.

Rozdělení náh. vektorů:

7. Multinomické rozdělení:

Základ: obecné rozdělení s více možnými výsledky.

k možných výsledků $\{a_1, \dots a_k\}$

$P(X_i = a_j) = p_j$, kde $p_j > 0$ & $\sum_{j=1}^{k} p_j = 1$

⇝ náh. vektor $Y = (Y_1, \dots Y_k)$

Zavedme $Z_i = (Z_{i1}, \dots Z_{ik})$, kde $Z_{ij} = 1 \Leftrightarrow X_i = a_j$ (jinak 0)

$Y = \sum_{i=1}^{n} Z_i$ … vždy platí $Y_k = n - \sum_{j=1}^{k-1} Y_j$

$$P[Y = (n_1, \dots n_k)] = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

pro $\forall (n_1 \dots n_k)$ takové, že $n_i \ge 0$ & $\sum_{j=1}^{k} n_j = n$

Y_j má $Bi(n, p_j)$ ⇒
$$\boxed{\begin{array}{l} EY_j = np_j \\ \operatorname{var} Y_j = np_j(1-p_j) \\ \operatorname{cov}(Y_i, Y_j) = -np_ip_j \end{array}}$$

Zobecnění Hypergeometrického

N koulí

M_i koulí barvy i ; $\sum_{i=1}^{k} M_i = N$

n koulí vytáhn. bez vracení

$$P(Y=(m_1 \dots m_k)) = \frac{\binom{M_1}{m_1} \cdots \binom{M_k}{m_k}}{\binom{N}{n}}$$

↳ # vytaž. barvy i

Spojité náhodné veličiny a vektory:

X nabývá hodnoty v nějakém otevřeném intervalu / sjednocení otevřených intervalů.

Def: (Absolutně spojitá náh. velič.)

Náh. veličina $X: (\Omega, A, P) \to (\mathbb{R}, \mathcal{B})$ (kde $\mathcal{B}$ je σ-algebra obsahující $\forall$ otevřené intervaly (borelovská σ-algebra)) se nazývá absolutně spojitá, existuje-li nezáporná funkce $f: \mathbb{R} \to [0, \infty)$ taková, že $P[X \leq x] = \int_{-\infty}^{x} f(x)\,dx$.

Pozn. + jistě musí platit $\int_{-\infty}^{\infty} f(x) = 1$

Př: $f(x) = 1 \quad x \in (0,1)$

$= 0 \quad x \notin (0,1)$

$\Rightarrow P[X \leq x] = \int_0^x 1\,dx = x$ pro $x \in (0,1)$

$= 0$ pro $x \leq 0$ ⇝ Distribuční funkce.

$= 1$ pro $x \geq 1$

Def: Absolutně sp. n.v. je taková, jejíž distrib. funkce je integrálem nějaké nezáporné funkce, tedy platí:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$$

a pro „skoro všechny" x platí $F_X'(x) = \frac{d}{dx} F_X(x) = f_X(x)$

↳ ~~derivace~~ derivace podle x

+ funkci f_X říkáme HUSTOTA ROZDĚLENÍ X.

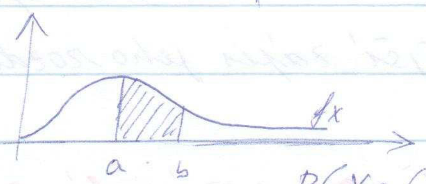
... skoro všechny ≈ je konečně mnoho výjimek.

$f_X(x) = 1 \quad x \in (0,1)$
$\quad = 0 \quad$ jinak

$P(X \in (a,b)) = b - a$
$\quad$ = délka intervalu.

• Skripta: Zvára Štěpán: Pravděpodobnost a statistika.
• ∃ funkce, kt. nejde zapsat jako integrál => ∃ jiné definice.

po obecnou f_X:



$P(X \in (a,b)) = \int_a^b f_X(x)\,dx.$

+ Pro „skoro všechny" body x tedy platí:

$$f_X(x) = \lim_{h \to 0+} \frac{P(X \in [x, x+h])}{h}$$

ABSOLUTNĚ SPOJITÝ NÁH. VEKTOR:

Def: Náhodný vektor $X = (X_1 \dots X_d)$ je absolutně spojitý, existuje-li nezáp. funkce $f_X \; \mathbb{R}^d \to \mathbb{R}$ takové, že

$$F_X(x_1 \dots x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_d} f_X(t_1, t_2 \dots t_d)\, dt_d\, dt_{d-1} \dots dt_1.$$

! X abs. spoj. náh.v., vektor $(x, -x)$ má abs. spojité složky, ale není abs. sp. náh. vektor!

Marginální distr. funkce:

$$F_{X_1}(x_1) = \lim_{x_2 \to \infty \dots x_d \to \infty} F_X(x_1 \dots x_d) = \int_{-\infty}^{x_1} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(t_1 \dots t_d)\, dt_d \dots dt_2 \right) dt_1$$

$= \int_{-\infty}^{x_1} f_{X_1}(t_1)\, dt_1$ ...., opět uvolnené ∀ složky mimo 1.

$$\Rightarrow f_{X_1}(x_1) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{d-1 \text{ integrálů}} f_X(x_1, t_2, \dots t_d)\, dt_d \dots dt_2$$

⇝ f_{X_1} je MARGINÁLNÍ HUSTOTA.

Def: Marginální hustota $f_{X_1 \dots X_k}(x_1 \dots x_k) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{d-k \text{ int.}} f_X(x_1 \dots x_k, t_{k+1}, \dots t_d)\, dt_d \dots dt_{k+1}$

~~Pozor~~:

Rozdělení abs. spoj. n. vekt. je dáno jednoznačně hustotou

Pozn. Náhodný vektor může sdružovat abs. spojité a diskrétní složky. Potom je obvykle složitější zápis jeho rozdělení.

· Absolutně spojitý náh. vektor má nezávislé složky $\Leftrightarrow$

$$f_X(x_1, \dots x_d) = \prod_{i=1}^{d} f_{X_i}(x_i) \quad \forall (x_1 \dots x_d) \in \mathbb{R}^d$$

„Def:"

Pro absolutně spojitý náh. vektor X definujeme

$$E\varphi(X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x_1 \dots x_d)\, f_X(x_1 \dots x_d)\, dx_d \dots dx_1$$

má-li integrál napravo smysl, a $\varphi: \mathbb{R}^d \to \mathbb{R}$ (φ je funkce náh. vektoru)

Pozn.

Korektní definice: $E\varphi(X) = \int_{\Omega} \varphi(X(\omega)) \cdot P(d\omega)$

Příklady φ: 1. $\varphi(x) = x_1$

$$EX_1 = E\varphi(X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1 f_X(x_1 \dots x_d)\, dx_d \dots dx_1 =$$

$$= \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(x_1, x_2 \dots x_d)\, dx_d \dots dx_2 \right) dx_1 = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\, dx.$$

2. $\varphi(X) = (X_1 - EX_1)^2 \rightsquigarrow E\varphi(X) = E(X_1 - EX_1)^2 = \operatorname{var} X_1$.

3. $\varphi(X) = (X_1 - EX_1)(X_2 - EX_2) \rightsquigarrow E\varphi(X) = \operatorname{cov}(X_1, X_2)$.

Příklady rozdělení abs. spoj. náh. veličin:

1. ROVNOMĚRNÉ ROZDĚLENÍ:

· konečný interval (a,b): $a < b$. + Hustota musí být konstantní.

$\Rightarrow f_X = \frac{1}{b-a}$ (aby $\int_a^b f(x)\,dx = 1$).

$$\boxed{EX = \frac{a+b}{2} \qquad \operatorname{var} X = \frac{(b-a)^2}{12}}$$

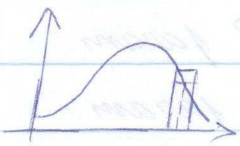
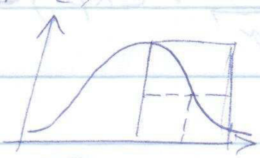
· Používá se jako „neinformativní rozdělení"

· „zaokrouhlovací chyba": 2. str.

1. 12. 15 / 3

PAS

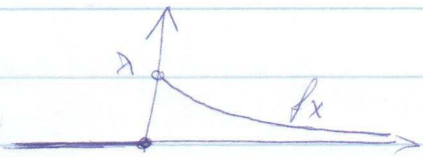
X změříme s přesností na cm. „Skutečné" X má rovnoměrné rozdělení na int. $(x-\frac{1}{2}, x+\frac{1}{2})$

 , pozor na  , velká chyba!

2. EXPONENCIÁLNÍ ROZDĚLENÍ

– je dáno hustotou $f_X(x) = \lambda e^{-\lambda x}$ $\quad x > 0$

$= 0$ $\quad x \leq 0$



$$\boxed{\begin{aligned} EX &= 1/\lambda \\ \operatorname{var} X &= 1/\lambda^2 \end{aligned}}$$

⇝ Základní rozdělení pro modelování „doby mezi událostmi"

+ je to ROZDĚLENÍ BEZ PAMĚTI : $P(X > x+y \mid X > y) = P(X > y) =$
→ reálně ne vždy odpovídá skut. $\quad = P(X > y \mid X > 0)$.

3. GAMA ROZDĚLENÍ

$$f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \qquad x > 0$$

- $\lambda > 0$ je PARAMETR, $n \in \mathbb{N}$ je taky parametr.
- Vezmeme-li n nezávislých náh. veličin s rozdělením $Exp(\lambda)$ pak jejich součet má Gama(n, λ): $\boxed{\begin{aligned} EX &= \frac{n}{\lambda} \\ \operatorname{var} X &= n/\lambda^2 \end{aligned}}$
- Pro $n \in \mathbb{R}$ $\quad f_X(x) = \dfrac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)}$, kde $\Gamma(n) = \int_{-\infty}^{\infty} x^{n-1} e^{-x} \, dx$

je GAMA FUNKCE: $n \in \mathbb{N}$ $\quad \Gamma(n) = (n-1)!$

$\forall n \quad \Gamma(n) = (n-1)\,\Gamma(n-1)$

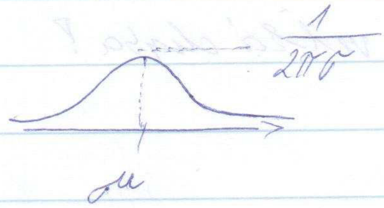
(zobecnění faktoriálu).

4. NORMÁLNÍ (GAUSSOVO) ROZDĚLENÍ

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad x \in \mathbb{R}$$

$\mu \in \mathbb{R}$ param.

$\sigma > 0$ param.



menší $\sigma \rightsquigarrow$ užší (Ω)

větší $\sigma \to$ širší ($\frown$)

$$\boxed{\begin{aligned} EX &= \mu \\ \operatorname{var} X &= \sigma^2 \end{aligned}}$$

5. NORMOVANÉ NORMÁLNÍ ROZDĚLENÍ

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad x \in \mathbb{R} \quad \dashrightarrow \sigma^2 = 1, \ \mu = 0$$

$N(0,1)$

Z má $N(0,1)$

$X = \sigma Z + \mu$ má právě rozdělení $N(\mu, \sigma^2)$ s hustotou f_X gaussova rozdělení.

6. STANDARDIZOVANÉ DVOUROZMĚRNÉ NORMÁLNÍ

$N_2(0, I_2) =$ (napr.) $N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1&0\\0&1\end{pmatrix}\right)$ má hustotu

$$f_Z = \frac{1}{2\pi} \exp\left\{\frac{-x_1^2 + x_2^2}{2}\right\} \qquad (x_1, x_2) \in \mathbb{R}^2$$

- jde o součin 2 $N(0,1)$ hustot $\equiv$

$Z = (Z_1, Z_2)$; Z_1, Z_2 mají $N(0,1)$ rozdělení a jsou nezávislé.

A matice 2×2 s hodností 2, b vektor z $\mathbb{R}^2$. $AZ + b$ má obecné dvourozměrné rozdělení $N(b, AA^T)$.

$\rightsquigarrow$ Hustota $N_2(\mu, \Sigma)$: $f_X(x_1, x_2) = \frac{1}{2\pi(\det \Sigma)^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$

1.12.15/4
8.12.15/1

PAS

Pozor. : $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ $\left.\begin{matrix} X_1 \text{ má } N(\mu_1, \sigma_{11}) \\ X_2 \text{ má } N(\mu_2, \sigma_{22}) \end{matrix}\right\}$ $\operatorname{cov}(X_1, X_2) = \sigma_{12}$

MARGINÁLNÍ ROZDĚLENÍ JE OPĚT NORMÁLNÍ.

Hustota normovaného normálního rozdělení

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}$$

Distribuční funkce:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt.$$

SOUČTY NEZÁVISLÝCH NÁHODNÝCH VELIČIN:

? Proč se zajímat o součty?

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$; X_i jsou nezávislé a stejně rozdělené n.v.

jde o náhodnou veličinu

- Paralela ke střední hodnotě:

EX - teor. stř. hodnota

$\overline{X}_n$ - můžeme spočítat z pozorování. Pro $n \to \infty$ $\overline{X}_n \to EX$.

Věta (rozdělení součtu 2 nezávislých náh. veličin.)

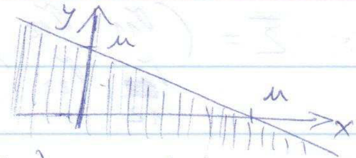
1. Buďte X, Y celoč. náh. velič. Pak $Z = X+Y$ je CNV a platí,

$$P[Z=z] = \sum_{n=-\infty}^{\infty} P[X=n] \, P[Y=z-n]$$

2. Buďte X, Y nezávislé absolutně spojité náh. veličiny, pak $Z = X+Y$ je absolutně spojitá n.v. a platí

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \, f_Y(z-x) \, dx.$$

Dk: 2.) Víme $F_Z(u) = \int_{-\infty}^{u} f_Z(z)\,dz$



$$F_Z(u) = P(X+Y \le u) = \iint\limits_{\{x,y:\, x+y\le u\}} f_X(x)\, f_Y(y)\, dy\, dx =$$

$$= \int_{-\infty}^{\infty} \boxed{\int_{-\infty}^{u-x} f_X(x) f_Y(y)\, dy}\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{u} f_X(x)\, f_Y(z-x)\, dz\, dx =$$

substituce $y = z - x$

Môžem prehodiť integrály podľa nejakej vety.

$$= \int_{-\infty}^{u} \underbrace{\int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx}_{f_Z(z)}\, dz = \int_{-\infty}^{u} f_Z(z)\, dz \qquad \blacksquare$$

Př: X má hustotu $f_X(x) = \lambda e^{-\lambda x}$ $x>0$
Y má hustotu $f_Y(y) = \lambda e^{-\lambda y}$ $y>0$ } a jsou nezávislé!

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx =$$

$$= \lambda^2 \int_0^z e^{-\lambda z}\, d\cancel{x} = \boxed{\lambda^2 \cdot z \cdot e^{-\lambda z}}$$ $\rightsquigarrow$ hustota gama rozdělení s parametry $(2, \lambda)$

Pozn.

1. Součet diskrétní a absol. spojité n.v. je absolutně spojitá n.v.
2. Obdobně lze odvodit i rozdělení součinu 2 nezávislých n.v. $\rightsquigarrow$ problém se znaménky $\Rightarrow$ $P[X>0] = P[Y>0] = 1$ a OK.

$Z = XY$

celočís: $P[Z=z] = \sum_{x=1}^{\infty} P[X=x] \cdot P[Y=\frac{z}{x}]$

abs. spoj: $f_Z(z) = \int_0^{\infty} f_X(x)\, f_Y\left(\frac{z}{x}\right) dx.$

ČEBYŠEVOVA NEROVNOST

- Obecným způsobem s využitím rozptylu odhadneme pravděpodobnost odchylky náh. veličiny od její stř. hodnoty

Věta (Č.N.)

Buď X náhodná veličina s konečnou střední hodnotou EX. Pak

$$P[|X-EX|>\varepsilon] \leq \frac{\operatorname{var} X}{\varepsilon^2} \quad \forall \varepsilon>0.$$

Dk $P[|X-EX|>\varepsilon] = P\left[\frac{|X-EX|}{\varepsilon}>1\right] = P\left[\frac{(X-EX)^2}{\varepsilon^2}>1\right] =$

$= $ (abs. spojitá, celoč. má součet $\sum P[X=x]$ a vyjde stejně.) $\displaystyle\int_{\{x: \frac{(x-EX)^2}{\varepsilon^2}>1\}} 1\cdot f(x)\,dx < \int_{\{x: \frac{(x-EX)^2}{\varepsilon^2}>1\}} \overbrace{\frac{(x-EX)^2}{\varepsilon^2}}^{\geq 0} \overbrace{f(x)}^{\geq 0}\,dx$

$$\leq \int_{-\infty}^{\infty} \frac{(x-EX)^2}{\varepsilon^2} f(x)\,dx = \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-EX)^2 f(x)\,dx = \frac{E(X-EX)^2}{\varepsilon^2} =$$

$$= \frac{\operatorname{var} X}{\varepsilon^2}$$ ∎

(slabý) zákon velkých čísel:

Vezměme si $Y = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (=\bar{X}_n)$

- $EY = E\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} EX_i$ a jsou-li X_i nezávislé & stejně rozdělené,

$\Rightarrow EX_i = EX_1 \; \forall i \Rightarrow EY = EX_1$

- $\operatorname{var} Y = \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\left(\sum_{i=1}^{n} \underbrace{\operatorname{var} X_i}_{\operatorname{var} X_1} + \sum_{i\neq j}\sum \underbrace{\operatorname{cov}(X_i,X_j)}_{0}\right)\right)$

$= 1/n \cdot \operatorname{var} X_1$

Věta: (Čebyševův (slabý) zákon velkých čísel)
Jsou-li $X_1 \ldots$ nezávislé & stejně rozdělené n.v. s konečným rozptylem ($\operatorname{var} X_i < \infty$), ~~tak platí~~: a $EX_1 = \mu$, pak:

$$P[|\bar{X}_n - \mu| < \varepsilon] \xrightarrow[n\to\infty]{} 0 \qquad \forall \varepsilon > 0.$$

Dk: označíme $Y = \bar{X}_n$. Z Čebyševovy nerovnosti plyne

$$P[|Y - EY| > \varepsilon] = P[|Y-\mu| > \varepsilon] < \frac{\operatorname{var} Y}{\varepsilon^2} = \frac{1}{n} \frac{\operatorname{var} X_1}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

$\blacksquare$

Př: Buďte $X_1 \ldots$ nezávislé $\operatorname{alt}(p)$

$P[X=1] = p = 1 - P[X=0]$.

$\frac{1}{n}\sum_{i=1}^{n} X_i$ = podíl úspěchů v n pokusech.

$EX = p$, $\operatorname{var} X = p(1-p)$.

$$P\left[\left|\frac{1}{n}\sum_{i=0}^{n} X_i - p\right| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0 \qquad \forall \varepsilon > 0.$$

Také píšeme $\frac{1}{n}\sum_{i=0}^{n} X_i \xrightarrow{P} p$ ... konverguje v pravděpodobnosti.

Pozor. ~~Platí~~: Někdo se domnívá, že platí $\left(\sum_{i=1}^{n} X_i - np\right) \xrightarrow{P} 0$,
ale TO NENÍ PRAVDA!
(např. ač padl 3x ~~orel~~ rub, není vyšší pravděp., že padne líce).

Def: Posloupnost $X_1, X_2 \ldots$ náhodných veličin splňuje slabý zákon velkých čísel, existuje-li konstanta b (konečná) taková, že

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - b\right| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0.$$

Podle Čeb. zákona tedy posloupnost nez. stejně rozdělených n.v. s $\operatorname{var} X < \infty$ splňuje ZVČ, kde $b = EX_1$.

• $X_1 \dots$ takové, že $\frac{1}{n}\sum_{i=1}^{n} EX_i \to b$

nezávislé $\frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var} X_i \to 0$

podle Č.N. opět $P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - b\right| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0$.

Protože $\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var} X_i + \frac{1}{n^2}\sum_{i\neq j}\sum \operatorname{cov}(X_i, X_j)$,

pokud $\operatorname{var} X_i \le K \;\forall i$ a pokud $\frac{1}{n^2}\sum_{i\neq j}\sum \operatorname{cov}(X_i, X_j) \to 0$, ~~pak~~ a $\frac{1}{n}\sum_{i=1}^{n} EX_i \to b$, pak $\{X_i\}_{i=1}^{\infty}$ také splňuje ZVČ.

Př: $P[X_i = 1] = \frac{1}{2} = P[X_i = -1]$

$P[X_i = 1$ pro 5_i a $X_i = -1$ pro $0_i] = \frac{1}{2^5} = \underline{1/32}$

$\frac{1}{n}\sum_{i=1}^{n} X_i = \sum \frac{X_i}{n} \xrightarrow{P} 0$

$\cancel{\sum_{i=1}^{\infty} \frac{1}{i} = +\infty}$ $\cancel{\sum_{i=1}^{\infty} \frac{(-1)^i}{i} < \infty}$ (fail.)

CENTRÁLNÍ LIMITNÍ VĚTA.

? Co říká ZVČ? ⇝ Pokud jsou X_i nezávislé, st. rozdělené, $EX_1 = \mu$ $\operatorname{var} X_1 = \sigma^2 < \infty$, pak $\frac{1}{n}\sum_{i=1}^{\infty}(X_i - \mu) \xrightarrow{P} 0$.

? Co když vezmeme jen $\frac{1}{\sqrt{n}}$? (Ukáže se, že v limitě konverg. k $\Phi(X)$).

Věta: (CLV): Buďte $X_1, X_2 \dots$ nezávislé, stejně rozdělené m.v., $EX_1 = \mu$, $0 < \operatorname{var} X_1 = \sigma^2 < \infty$, PAK:

$$\cancel{P\left[\frac{1}{\sqrt{n}} \cdot \frac{\sum_{i=1}^{n}(X_i-\mu)}{\sqrt{\sigma^2}}\right]} \quad P\left[\frac{1}{\sqrt{n}} \cdot \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{\sigma^2}} \le x\right] \longrightarrow \Phi(x)$$

pro $\forall x \in \mathbb{R}$ (stejnoměrně). Distrib. fce $N(0,1)$

Distribuční funkce n.v.

$$z = \frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}}$$ konverguje k d.f. $N(0,1)$.

$X_1 \dots X_n$ nezávislé $N(0,1)$

$EX_1 = 0$ $\Rightarrow$ $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ má opět $N(0,1)$

$\operatorname{var} X_1 = 1$

$$\frac{\sum_{i=1}^{n} X_i - n \cdot 0}{\sqrt{n \cdot 1}} = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \frac{X_i - 0}{\sqrt{1}}$$

Jiné formy CLV:

$$P\left[\frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sqrt{\sigma^2}} \le x\right] \to \Phi(x)$$

$$P\left[\sqrt{n} \cdot \frac{(\bar{X}_n - \mu)}{\sqrt{\sigma^2}} \le x\right] = P\left[\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}(X_i - \mu)\right) \le x\sqrt{\sigma^2}\right] \to \Phi(x)$$

X_i $\operatorname{alt}(p)$; $EX_1 = p$, $\operatorname{var} X_1 = p(1-p)$

$$P\left[\sqrt{n}\,\frac{(\bar{X}_n - p)}{\sqrt{p \cdot (1-p)}} \le x\right] \to \Phi(x)$$

$$P\left[(\bar{X}_n - p) \le x \cdot \sqrt{\frac{p(1-p)}{n}}\right] \to \Phi(x)$$

$$\Phi(1{,}96) = 0{,}975$$

$P\left[|\bar{X}_n - p| \le 1{,}96 \cdot \sqrt{\frac{p(1-p)}{n}}\right] \doteq$ ~~0,975 0,975~~ 0,95, kvůli abs. hodnote $\Rightarrow$ $(0{,}975)^2$

15.12.15/1

Opakování: PAS

<u>ZVČ</u>:

$X_1, X_2 \dots$ nezávislé & stejně rozdělené

$\bar{X}_n = \frac{1}{n} \sum X_i \xrightarrow{P} EX$ pokud $\operatorname{var} X_1 < \infty$

$$P\left(\left|\frac{1}{n} \sum_i X_i - \frac{1}{n} \sum EX_i\right| \geq \varepsilon\right) \leq \frac{\operatorname{var}\left(\frac{1}{n} \sum X_i\right)}{\varepsilon^2}$$

<u>CLV</u>:

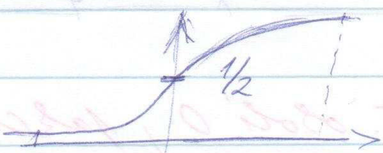
Pokud $X_1, X_2 \dots$ nez. & st. rozděl. $0 < \underset{\sigma^2}{\operatorname{var} X_1} < \infty$, $EX_1 = \mu$,

$$P\left(\sqrt{n}\, \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2}} \leq x\right) \to \Phi(x), \text{ resp. } \sqrt{n}\, \frac{(\bar{X}_n - \mu)}{\sqrt{\sigma^2}} \xrightarrow{\mathcal{D}} N(0,1)$$

„v distribuci"

! ZVČ netvrdí $|\sum X_i - n\mu| \xrightarrow{P} 0$, ale $\left|\frac{1}{n} \sum X_i - EX\right| \xrightarrow{P} 0$.

$\Phi(x)$:



$-1{,}645 \rightsquigarrow 0{,}05$ $1{,}645 \rightsquigarrow 0{,}95$

$$P\left(\left|\sum_i X_i - nEX_1\right| \leq \sqrt{n\sigma^2}\, x\right) \to \Phi(x) - \Phi(-x)$$

$\rightsquigarrow$ symetria $\Phi(x) = 1 - \Phi(-x)$ $\;\hookrightarrow = \underline{2\Phi(x) - 1}$

$$P\left(\left|\sum X_i - nEX_1\right| \leq 1{,}645 \cdot \sqrt{n\sigma^2}\right) \doteq 0{,}9$$

$\hookrightarrow$ roste s n.

$$P\left(\left|\sum_i X_i - nEX\right| \geq 0{,}67\sqrt{n\sigma^2}\right) \doteq 0{,}5$$

$\hookrightarrow$ roste nade $\forall$ meze

$n = 10^4$, $\sigma = 1$

$$\Rightarrow \quad P\left[\left|\sum_{i=1}^{10^4} X_i - 10^4 EX\right| \geq 67\right] \doteq 0{,}5$$

ale

$$P\left[\left|\frac{1}{n} \sum X_i - EX_1\right| \leq 3\sqrt{\frac{\sigma^2}{n}}\right] \doteq 0{,}999$$

$$P\left[\left|\bar{X}_{10000} - EX\right| \leq 0{,}03\right] \doteq 0{,}999$$

Chernoffovy meze

Def: Buď X náh. veličina. Funkci $\psi_X(t) = Ee^{tX}$ nazveme momentová vytvořující funkce X.

Pozn. Pro CNV je $\psi_X(t) = \sum_{k \in \mathbb{Z}} e^{tk} P[X=k]$

Pro ASV je $\psi_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$

$\psi_X(t)$ může být pro mnoho t nekonečná

$\psi_X(0) = 1$

~~Pokud $\psi_X(t)$ existuje konečná na otevřeném ~~intervalu~~ okolí ($0 - \delta, 0 + \delta$), pak charakt. rozdělení X, Y a platí $\psi_X(t) = \psi_Y(t) < \infty \ \forall t \in (-\delta, \delta) \Rightarrow F_X = F_Y$~~

Věta:

Pokud $\psi_X(t)$ ex. konečná na otevř. okolí 0, pak charakt. rozdělení X, Y a $\psi_X(t) = \psi_Y(t) < \infty \ \forall t \in (-\delta, \delta)$

$\Rightarrow F_X = F_Y$

Věta: X, Y nez. n.v.

$\psi_{X+Y}(t) = \psi_X(t)\,\psi_Y(t)$, speciálně

$\psi_{\sum X_i}(t) = \psi_{X_1}^n(t)$, pokud $X_1 \dots X_n$ jsou stejně rozděl.

Dk:

$\psi_{X+Y}(t) = Ee^{t(X+Y)} = E\,e^{tX}e^{tY} \overset{\text{nezáv.}}{=} Ee^{tX}\,Ee^{tY} = \psi_X(t)\,\psi_Y(t)$ ∎

Př.

$Bi(n,p) = P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$

$\psi_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \dots = (1-p+e^t p)^n$

$\left.\begin{array}{l} X \sim Bi(n_1, p) \\ Y \sim Bi(n_2, p) \end{array}\right\}$ nezáv. $\Rightarrow \psi_{X+Y}(t) = (1-p+e^t p)^{n_1} * (1-p+e^t p)^{n_2} =$

$= (1-p+e^t p)^{n_1+n_2} = \psi_X(t)\,\psi_Y(t)$.

Věta Nechť X náh. veličina a $\psi_X(t) < \infty$ $\forall t \in (-t_0, t_0)$ pro $t_0 > 0$. Pak
$E|X|^r < \infty$ $\forall r \in \mathbb{N}$ a platí $EX^r = \psi_X^{(r)}(0)$
(= r-tá derivace v 0).

Dk:
Za podmínky věty lze zaměnit derivaci a integrál resp. ~~sumu~~ sumu .... $\left(\int_{-\infty}^{\infty} e^{tx} f(x)\,dx\right)_0^{(r)} = \int_{-\infty}^{\infty} (e^{tx})_0^{(r)} f(x)\,dx =$
$= \int_{-\infty}^{\infty} x^r f(x)\,dx = EX^r$; přičom

$(e^{tx})_0^{(r)} = x^r e^{tx}\Big|_{t=0}$ ... r-tá deriv. podľa t v 0. $\blacksquare$

Chernoffovy meze - Věta:
Buď X náh. veličina a $\psi_X(t)$ její vytv. funkce. Pak
$$P(X \geq a) \leq \inf_{t>0} \frac{\psi_X(t)}{e^{ta}} \quad \text{a} \quad P(X \leq a) \leq \inf_{t<0} \frac{\psi_X(t)}{e^{ta}}$$

Dk: $\forall t > 0$: $P(X \geq a) = P(e^{tX} \geq e^{ta}) = P\left(\frac{e^{tX}}{e^{ta}} \geq 1\right) =$
rostoucí v X a v a

$= \int_{\{x: \frac{e^{tx}}{e^{ta}} \geq 1\}} f(x)\,dx \leq \int_{\{x: \frac{e^{tx}}{e^{ta}} \geq 1\}} \frac{e^{tx}}{e^{ta}} f(x)\,dx \leq \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \cdot \frac{1}{e^{ta}} =$

CNV: $\sum_{x: \frac{e^{tx}}{e^{ta}} \geq 1} P(X = x)$
a obdobně

$= \frac{\psi_X(t)}{e^{ta}} \leq 1$

$\Rightarrow P(X \geq a) \leq \inf_{t>0} \frac{\psi_X(t)}{e^{ta}}$ druhý podobně $\blacksquare$

Vybrané statistické úlohy a jejich řešení pomocí ZVČ a CLV
- Uvažujme $X_1, X_2 \ldots X_n$ náh. velic. se st. rozděl. (TZV. NÁHODNÝ VÝBĚR)
$P(X_1 = 1) = p = 1 - P(X_1 = 0)$. kolik je p?
1. bodový odhad
2. intervalový odhad

μ neznáme, jediné, co víme, je, že $(x_1 \dots x_n)$ jsou hodnoty náh. výběru.

B.O. chci 1 číslo (pokud možno „blízko" μ).

I.O. chci interval, kt. obsáhne neznámé μ s velkou pravděpodobností (+ co nejmenší).

Bodový odhad – metoda momentů

$EX = \mu$ (neznáme)

$P(|\bar{X}_n - EX| \geq \varepsilon] \longrightarrow 0 \quad \forall \varepsilon > 0$ nezávisle na EX, čili $\forall \mu$.

výběrový průměr

$|\bar{X}_n - \mu| \xrightarrow{P} 0$ ... konzistence odhadu

$X_1 \dots X_n$ ... rozděl. s hustotou $\lambda e^{-\lambda x}, x > 0$, $\lambda > 0$ param.

λ neznáme

$EX = 1/\lambda$ (1× per partes)

$P[|\bar{X}_n - \frac{1}{\lambda}| \geq \varepsilon] \xrightarrow{n \to \infty} 0$

$\varphi(z) = \frac{1}{z}$ spojitá funkce na $z > 0$.

$P[|\frac{1}{\bar{X}_n} - \lambda| \geq \varepsilon] \xrightarrow{n \to \infty} 0$

$P[|\varphi(\bar{X}_n) - \varphi(\lambda^{-1})| \geq \varepsilon] \to 0$

... MOMENTOVÁ METODA ... $EX = h(\theta)$ h spojité (prosté)

- odhadneme EX pomocí $\bar{X}_n$ (pomocí ZVČ)
- položíme $\bar{X}_n = h(\hat{\theta})$, kde ^ značí bodový odhad
- odtud $\hat{\theta} = \varphi(\bar{X}_n)$

$\theta \in R^k$ $\quad EX = h_1(\theta), \dots, EX^k = h_k(\theta)$

toto určíme z modelu. Pokud $E|X|^k < \infty$, pak platí

$$\underbrace{\frac{1}{n}\sum X_i^r}_{\overline{X_n^r}} \xrightarrow{P} EX^r \qquad r = 1, 2, \dots k$$

$$\left.\begin{array}{l} \overline{X_n^1} = h_1(\hat{\theta}) \\ \vdots \\ \overline{X_n^k} = h_k(\hat{\theta}) \end{array}\right. \rightsquigarrow \hat{\theta} = (\hat{\theta}_1, \dots \hat{\theta}_k)$$

Nestrannost odhadu

$\hat{\theta}$ je nestranným odhadem θ, pokud $\underset{\theta}{E}\hat{\theta} = \theta \ \forall \theta$

↳ stř. hodnota, je-li parametr opravdu θ

např. u odhadu

$$\hat{p} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad P[X_i = 1] = p = 1 - P[X_i = 0]$$

platí

$$E_p X_1 = 1 \cdot p + 0 \cdot (1-p) = p$$

$$E_p \hat{p} = E_p\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum \overbrace{E_p X_i}^{p} = p \Rightarrow \text{je Nest. odh.}$$

$\hat{p}$ je nestranný odhad pravděp. úspěchu v alternativním rozdělení i konzistentní

$$\boxed{\begin{array}{l} E(\overline{X}_n) = EX_1 \\ \operatorname{var} \overline{X}_n = \frac{1}{n} \operatorname{var} X_1 \end{array}}$$

Bodové a intervalové odhady:

Def: Parametrickou třídou rozdělení rozumíme množinu $\mathcal{P} = \{ P_\theta ; \theta \in \Theta \}$, kde P_θ jsou rozdělení na prostoru $\mathbb{R} (\mathbb{R}^d)$, Θ je známý prostor parametrů (Θ je typicky podmnožina $\mathbb{R}^k$) a platí $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$
(identifikovatelnost).

Třída BINOMICKÁ

- lze pro ~~~~ pevné N (# pokusů) jako jednoparametrickou s parametrem $p \in (0,1) = \Theta$.

Třída GAUSSOVA

- $\forall$ rozdělení typu $N(\mu, \sigma^2)$; $\Theta = \mathbb{R} \times \mathbb{R}^+$
(s hustotou $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$)

Def: Náhodným výběrem (o rozsahu n) s rozdělením P_θ rozumíme soubor nezávislých a stejně rozdělených náh. veličin $X_1 \dots X_n$, kde X_i má rozd. P_θ $\forall i$.

Pozn. $X_1 \dots X_n$ náhodný výběr (náh. veličiny)
$x_1 \dots x_n$ realizace náh. výběru (naměřené hodnoty)

Def: Bodový odhad je funkce $g: \mathbb{R}^n \to \Theta$

$$g(X_1, \dots, X_n) = \hat{\theta}_n$$ (stříška značí odhad, index rozsah výběru)

! · Předpis funkce nesmí záviset na θ (což je neznámá hodnota)

· $g(X_1, \dots, X_n)$ je náhodná veličina s hodnotami v Θ

$g(x_1 \dots x_n)$ hodnota odhadu zjištěná z realizace náh. výběru není náhodná (číslo).

Def Odhad $\hat{\theta}_n$ je KONZISTENTNÍM odhadem θ, pokud $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta \quad \forall\, \theta \in \Theta$

T.j. po $\forall\, \theta$ má platit $g(X_1, \dots, X_n) \xrightarrow[n\to\infty]{P} \theta$, je-li $(X_1, \dots, X_n)$ náh. výběr z rozdělení P_θ.

Jak takový odhad najít? $\rightsquigarrow$ Nejjúčinnější je metoda momentů:

	Teoret. moment	Výběrový moment
	$EX = \mu_1(\theta)$	$1/n \sum X_i = \overline{X}_n$
$\theta \in \Theta \subset \mathbb{R}^k$	$EX^2 = \mu_2(\theta)$	$1/n \sum X_i^2 = \overline{X^2_n}$
	$\vdots$	$\vdots$
	$EX^k = \mu_k(\theta)$	$1/n \sum X_i^k = \overline{X^k_n}$

Př: $N(\mu, \sigma^2)$

$EX = \mu \qquad \frac{1}{n} \sum X_i = \hat{\mu}_n$

$EX^2 = \mu^2 + \sigma^2 \qquad \frac{1}{n} \sum X_i^2 = \hat{\sigma}_n^2 + \hat{\mu}_n^2 = \hat{\sigma}_n^2 + \left(\frac{1}{n} \sum X_i\right)^2$

Jde o konzistentní odhad? $\rightsquigarrow$ Ano, což plyne ze ZVČ

$1/n \sum Y_i \xrightarrow{P} EY$, pokud $E|Y| < \infty$

Př posunuté expon. rozdělení

hustota $f(x) = e^{-\lambda(x-a)}$, $x > a$, $\lambda > 0$, $a \dots$ posunutí.

$EX = \frac{1}{\lambda} + a$

$EX^2 = \frac{1}{\lambda^2} + \left(\frac{1}{\lambda} + a\right)^2 = \frac{2}{\lambda^2} + \frac{2a}{\lambda} + a^2$

Zde není ~~možné~~ vhodné odhadovat „a" pomocí metody momentů, zkombinujeme odhad „a" pomocí minima a momentovou metodou po λ.

Tedy $X_1 > a$, $X_2 > a, \dots, X_n > a$, nabízí se vzít

$\hat{a}_n = \min(X_1, \dots, X_n) = X_{1:n} = X_{(1)}$

Spočítat rozdělení minima není tak těžké:

$$P[X_{(1)} \leq x] = 1 - P[X_{(1)} > x] = 1 - P[\bigcap_{i=1}^{n}(X_i > x)] =$$

$$= 1 - \prod_{i=1}^{n} P[X_i > x] = 1 - (1 - F(x))^n$$ a platí $\hat{a}_n \searrow a$

Hledání „nejlepšího" odhadu pomocí ztrátové funkce:

ztrátová funkce $L(\hat{\theta}_n, \theta) \geq 0$, např. $(\hat{\theta}_n - \theta)^2$ a chceme, aby odhad minimalizoval $EL(\hat{\theta}_n, \theta)$, např. $E(\hat{\theta}_n - \theta)^2 \ \forall \theta$

! to ale nejde obecně, pouze s omezením na další vlastnosti $\hat{\theta}_n$ (např. nestrannost).

Def Odhad je NESTRANNÝ, pokud $E\hat{\theta}_n = \theta \ \forall \theta$

(t.j. pokud $X_1 \ldots X_n$ je náh. výběr z P_θ, pak $Eg(X_1, \ldots, X_n) = \theta$).
Jde tedy o odhad s nejmenším rozptylem.

Intervalový odhad:

- Buď $\alpha \in (0,1)$ povolená pravděpodobnost omylu. Intervalovým odhadem parametrem θ se spolehlivostí $1-\alpha$ rozumíme interval

$$(g_L(X_1, \ldots, X_n), g_u(X_1 \ldots, X_n))$$ takový, že

$$P_\theta(g_L(X_1 \ldots X_n) \leq \theta \leq g_u(X_1 \ldots X_n)) \geq 1 - \alpha$$

... cílem je najít g_L a g_u tak, aby interval byl co nejužší
g_L a g_u nesmí záviset na θ, jde o náhodné veličiny - funkce náhodného výběru.

Jak zkonstruovat intervalový odhad?

→ najdeme funkci $H(X_1, \dots, X_n, \theta)$ tak, aby * rozdělení H bylo známé (Δ) a nezáviselo na θ

· pak existují c_L, c_U takové, že

$$P(c_L \le H(X_1, \dots, X_n, \theta) \le c_U) \ge 1-\alpha$$

$$\Updownarrow$$

$$P(g_L(X_1, \dots, X_n) \le \theta \le g_U(X_1, \dots, X_n)) \ge 1-\alpha$$

* co kdybychom přidali ASYMPTOTICKÉ?

a na místo (Δ) normální

(t.j. najdeme funkci $H(X_1, \dots, X_n, \theta)$ tak, aby asymptotické rozdělení H bylo známé normální a nezáviselo na θ)

→ To nás navádí k použití CLV

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - EX}{\sqrt{\operatorname{var} X}} \cdot \sqrt{n} \xrightarrow{P} N(0,1)$$

↳ nezávisí na θ

Intervalový odhad založený na CLV

- $X_1 \dots X_n$ je náh. výběr z rozdělení s distribuční funkcí F_θ, $\theta \in \Theta$ je neznámý parametr.
- Hledáme dvojici funkcí $g_D(X_1 \dots X_n)$ a $g_H(X_1 \dots X_n)$ tak, aby pro $\forall\, \theta \in \Theta$ platilo $P_\theta(g_D(X_1 \dots X_n) \le \theta \le g_H(X_1 \dots X_n)) \ge 1-\alpha$, kde $\alpha \in (0,1)$ je obvykle malé (nejčastěji $\alpha = 0{,}05$).
- Hledáme funkci $H(\theta, X_1 \dots X_n)$ jejíž rozdělení nezávisí na θ, a které ZNÁME $\to P(c_L \le H(\theta, X_1 \dots X_n) \le c_U) \ge 1-\alpha$

nezávisí na θ & známe $\Rightarrow$ víme najít obě konstanty.

? Kdy to jde snadno?

1.) Je-li $\theta = \Psi(EX)$, kde Ψ je známé prosté zobrazení $\Rightarrow$ $EX = \Psi^{-1}(\theta)$. ↳ + monotónní ~~alternat.~~

2.) Platí CLV.

- Za platnosti CLV:

$$\sqrt{n} \cdot \frac{\bar{X}_n - EX}{\sqrt{\sigma^2}} \xrightarrow{\mathcal{D}} N(0,1) \text{ … známé, nezávislé na } \theta.$$

$\bar{X}_n = \frac{1}{n} \sum X_i$ výběrový průměr.

$\sigma^2 = \operatorname{var} X$ neznáme $\rightsquigarrow$ rušivý parametr (nuisance)

$$\Rightarrow \sqrt{n}\, \frac{\bar{X}_n - \Psi^{-1}(\theta)}{\sqrt{\sigma^2}} \text{ má cca } N(0,1) \text{ rozdělení}$$

Normální rozdělení $N(0,1)$ je symetrické kolem 0. Platí $P(Z \le x) = P(Z \ge -x)$ pro $Z \sim N(0,1)$

označme u_α: hodnota taková, že $P(Z \le u_\alpha) = \alpha$

t.j. $u_\alpha = \Phi^{-1}(\alpha)$ pro Φ dist. fce $N(0,1)$.

$\to$ KVANTIL.

$$\Rightarrow P\left(\sqrt{n}\,\frac{\bar{X}_n-\Psi^{-1}(\theta)}{\sqrt{\sigma^2}}\le u_{\alpha/2}\right)\doteq \alpha/2$$

$$\& \quad P\left(\sqrt{n}\,\frac{\bar{X}_n-\Psi^{-1}(\theta)}{\sqrt{\sigma^2}}\ge u_{1-\alpha/2}\right)\doteq \alpha/2$$

$$\Rightarrow P\left(\underset{-u_{1-\alpha/2}}{\underset{\parallel}{u_{\alpha/2}}}\le\sqrt{n}\,\frac{\bar{X}_n-\Psi^{-1}(\theta)}{\sqrt{\sigma^2}}\le u_{1-\alpha/2}\right)\doteq 1-\alpha$$

$\Rightarrow$

$$P\left(\bar{X}_n-u_{1-\alpha/2}\cdot\sqrt{\frac{\sigma^2}{n}}\le\Psi^{-1}(\theta)\le\bar{X}_n+u_{1-\alpha/2}\cdot\sqrt{\frac{\sigma^2}{n}}\right)\doteq 1-\alpha$$

$\Rightarrow$ pro rostoucí Ψ:

$$P\left(\Psi\left(\bar{X}_n-u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)\le\theta\le\Psi\left(\bar{X}_n+u_{1-\alpha/2}\sqrt{\sigma^2/n}\right)\right)\doteq 1-\alpha$$

? co se σ^2 $\to$ odhadneme, ale je třeba 1 tvrzení navíc:

<u>Věta</u> ((Cramérova-) Sluckého):

Nechť U_n je posl. náh. veličin taková, že $U_n \xrightarrow{\mathcal{D}} N(0,1)$

i) Je-li $Y_n \xrightarrow{P} 1$, pak $U_n\cdot Y_n \xrightarrow{\mathcal{D}} N(0,1)$.

ii.) Je-li $Y_n \xrightarrow{P} 0$, pak $U_n+Y_n \xrightarrow{\mathcal{D}} N(0,1)$.

<u>Dk:</u> (Plán: Najdeme-li odhad S_n^2 rozptylu σ^2 takový, že

$$\frac{S_n^2}{\sigma^2}\xrightarrow{P}1,\ \text{pak}\ \underbrace{\sqrt{n}\,\frac{\bar{X}_n-EX}{\sqrt{S_n^2}}}_{\downarrow\mathcal{D}\ N(0,1)}=\underbrace{\sqrt{n}\,\frac{\bar{X}_n-EX}{\sqrt{\sigma^2}}}_{\downarrow\mathcal{D}\ N(0,1)}\cdot\underbrace{\sqrt{\frac{\sigma^2}{S_n^2}}}_{\downarrow P\ 1}$$

[$\Rightarrow$ ↑ + Sluckého věta:

$$P\left(\Psi\left(\bar{X}_n-u_{1-\alpha/2}\sqrt{\frac{S_n^2}{n}}\right)\le\theta\le \text{to isté s } +\right)\doteq 1-\alpha$$]

Takovým odhadem je za předpokladu $EX^2<\infty$, $\operatorname{var}X>0$

$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$ ... VÝBĚROVÝ ROZPTYL) (samotný důkaz):

$P(U_n \le x) \xrightarrow[n\to\infty]{} \Phi(x) : \forall x$ & $P(|Y_n - 1| > \varepsilon) \xrightarrow[n\to\infty]{} 0 : \forall \varepsilon > 0$

? $P(U_n \cdot Y_n \le x) \xrightarrow[n\to\infty]{} \Phi(x) \ \forall x$

$P([U_n Y_n \le x] \cap [|Y_n - 1| \le \varepsilon]) + P([U_n Y_n \le x] \cap [|Y_n - 1| > \varepsilon])$

$\wedge|$

$P(|Y_n - 1| > \varepsilon) \to 0.$

$P([U_n Y_n \le x] \cap [|Y_n - 1| \le \varepsilon]) \le P([U_n \cdot (1-\varepsilon) \le x] \cap [|Y_n - 1| \le \varepsilon])$

$\ge P([U_n (1+\varepsilon) \le x] \cap [|Y_n - 1| \le \varepsilon])$

↳ volíme min. a max. Y_n (podle odhad)

→ spolu $P(U_n Y_n \le x) \le (1+\delta) P(U_n \le \frac{x}{1-\varepsilon})$

$\ge (1-\delta) P(U_n \ge \frac{x}{1+\varepsilon})$

$\forall \delta \ \exists \varepsilon$ umíme najít takové n_0, že po $\forall n \ge n_0$ platí uvedené nerovnosti

$(1-\delta)(1-y)\Phi\left(\frac{x}{1+\varepsilon}\right) \le P(U_n Y_n \le x) \le (1+\delta)(1+y)\Phi\left(\frac{x}{1-\varepsilon}\right)$

Příklad: $X_1 \dots X_n$ buď výběr z rozdělení s hustotou $f(x) = \lambda e^{-\lambda x}$ $x > 0$

$= 0$ $x \le 0$

kde $\lambda > 0$ je neznámý parametr.

$EX = \int_0^\infty \cancel{\lambda e} \ x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$ $X > 0$ => platí ZVČ.

$EX^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$ => $\operatorname{var} X = \frac{1}{\lambda^2} \in (0, \infty)$

=> platí CLV.

$\bar{X}_n$, $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ spočítáme a použijeme

CLV: $\alpha = 0{,}05$ => $u_{1-\alpha/2} \doteq 1{,}96$

$$P\left(-1{,}96 \le \sqrt{n}\,\frac{\overline{X}_n - 1/\lambda}{\sqrt{S_n^2}} \le 1{,}96\right) \doteq 0{,}95$$

$$P\left(\overline{X}_n - 1{,}96\sqrt{\frac{S_n^2}{n}} \le 1/\lambda \le \overline{X}_n + 1{,}96\sqrt{\frac{S_n^2}{n}}\right) \doteq 0{,}95$$

+ otočia sa zlomky a OK.

II. možnost -> využít $\operatorname{var} X = 1/\lambda^2$

$$P\left(-1{,}96 \le \frac{\overline{X}_n - 1/\lambda}{1/\lambda}\sqrt{n} \le 1{,}96\right) = P\left(-1{,}96 \le (\lambda \overline{X}_n - 1)\cdot\sqrt{n} \le 1{,}96\right) =$$

$$= \cancel{P(\lambda}$$

$$= P\left(\left(1 - (1{,}96/\sqrt{n})\right) \le \lambda \overline{X}_n \le \left(1 + \frac{1{,}96}{\sqrt{n}}\right)\right) =$$

$$= P\left(\frac{1 - \frac{1{,}96}{\sqrt{n}}}{\overline{X}_n} \le \lambda \le \frac{1 + \frac{1{,}96}{\sqrt{n}}}{\overline{X}_n}\right) \doteq 0{,}95$$

~> Bodový odhad metodou momentů -> $\frac{1}{\hat{\lambda}} = \overline{X}_n \Rightarrow \hat{\lambda} = \frac{1}{\overline{X}_n}$.

? Jak porovnat střední hodnotu 2 náh. veličin ?:

X... rozdělení se stř. hodnotou μ_X, rozptylem $\sigma_X^2 \in (0,\infty)$

Y... ——— " ——— μ_Y, rozptylem $\sigma_Y^2 \in (0,\infty)$ platí CLV

chceme intervalový odhad pro $\mu_X - \mu_Y$

=> Použití: tvrzení $\mu_X - \mu_Y = \Delta$. Pokud $\Delta \notin$ interval spolehlivosti se spolehlivostí $1-\alpha$, pak řekneme, že dané tvrzení ZAMÍTÁME s hladinou statist. významnosti α.

mějme $X_1 \dots X_n$ výběr z rozdělení F_X + nezávislé
$Y_1 \dots Y_m$ F_Y výběry

tak, aby m a n byly dostatečně velké $\left(n \to \infty, \frac{m}{n} \to \lambda \in (0,\infty)\right)$

z CLV: $\dfrac{\bar{X}_n - \mu_x}{\sqrt{\frac{S_n^2}{n}}} \xrightarrow{\mathcal{D}} N(0,1)$ a $\dfrac{\bar{Y}_m - \mu_y}{\sqrt{\frac{S_y^2}{m}}} \xrightarrow{\mathcal{D}} N(0,1)$

$$\left.\begin{array}{l} \operatorname{var} \bar{X}_n = \dfrac{\sigma_x^2}{n} \\ \operatorname{var} \bar{Y}_m = \dfrac{\sigma_y^2}{m} \end{array}\right\}$$ z nejakej vety čo sme nemali

$\Rightarrow \operatorname{var}(\bar{X}_n - \bar{Y}_m) = \dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}$ z nezávislosti výběrů

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \xrightarrow{\mathcal{D}} N(0,1)$$

$$P\left(-u_{1-\alpha/2} \le \frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \le u_{1-\alpha/2}\right) =$$

$$= P\left(\bar{X}_n - \bar{Y}_m - u_{1-\alpha/2} \cdot \sqrt{\quad} \le \mu_x - \mu_y \le \bar{X}_n - \bar{Y}_m + \dots\right) \doteq 1-\alpha$$

$\Rightarrow$ zamítneme nebo ne.

THE END.