

Axiomy ($\Omega, \mathcal{F} \subseteq \mathcal{P}(\Omega), P: \mathcal{F} \rightarrow [0, 1]$)

- 1) $\emptyset, \Omega \in \mathcal{F}$
- 2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- 3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \cup A_i \in \mathcal{F}$
- 4) $P(\Omega) = 1$
- 5) $P(\cup A_i) = \sum_i P(A_i)$

Podmíněná pravd.

$$P(A \cap B) = P(B) \cdot P(A|B)$$

$$P(A_1 \wedge A_2 \wedge A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \wedge A_2)$$

$$P(A) = \sum_i P(B_i) \cdot P(A|B_i), B_i \text{ rozdělají } \Omega$$

$$P(B_j|A) = \frac{P(B_j \wedge A)}{P(A)} = \frac{P(B_j) \cdot P(A|B_j)}{P(A)}$$

$$A \perp B \equiv P(A \wedge B) = P(A) \cdot P(B)$$

$$A \perp c B \equiv P(A \wedge B|C) = P(A|C) \cdot P(B|C)$$

Diskrétní

- $X \sim \text{Ber}(p)$
 $f_X(x) = p^x (1-p)^{1-x}$
 $E[X] = p$ Var = $p(1-p)$

- $X \sim \text{Geom}(p)$
 $f_X(x) = (1-p)^{x-1} p$
 $E[X] = \frac{1}{p}$ Var = $\frac{1-p}{p^2}$

- $X \sim \text{Bin}(n, p)$
 $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
 $E[X] = np$ Var = $np(1-p)$

- $X \sim \text{Pois}(\lambda) \approx \text{Bin}(n, \frac{\lambda}{n})$
 $f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $\lambda > 0$
 $E[X] = \text{Var} = \lambda$

Rozdělení

- $X \sim U(a, b)$
 $f_X(x) = \frac{1}{b-a}$, $x \in [a, b]$
 $E[X] = \frac{a+b}{2}$ Var = $\frac{(b-a)^2}{12}$

- $X \sim \text{Exp}_\mu(\lambda) \quad \lambda > 0$
 $F_X(x) = 1 - e^{-\lambda x}$, $x > 0$
 $E[X] = \frac{1}{\lambda}$ Var = $\frac{1}{\lambda^2}$

- $X \sim N(0, 1)$
 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
 $E[X] = \mu$ Var = σ^2
- $X \sim N(\mu, \sigma^2) \Rightarrow X = \mu + \sigma Z$
 $f_X(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$
 $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

$$X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \sum X_i \sim N\left(\sum \mu_i, \sum \sigma_i^2\right)$$

Správka

Diskrétní m.r. $X: \Omega \rightarrow \mathbb{R}$

$$1) \text{Im}(X) \text{ je spojelná}$$

$$2) \forall x \in \mathbb{R}: \{x=x\} \in \mathcal{F}$$

$$\Rightarrow f_X(x): x \mapsto P[X=x]$$

Vlastnosti E - linearita

$$E[X] = \sum_{m=0}^{\infty} P[X=m], \text{Im}(X) \subseteq \mathbb{N}$$

$$X \perp Y \Rightarrow E[XY] = EX \cdot EY$$

Vlastnosti $\text{Var } X := E[(X-EX)^2]$

$$\text{Var } X = E[X^2] - (EX)^2$$

$$= E[X(X-1)] + EX - (EX)^2$$

$$\text{Var}(aX+bY) = a^2 \text{Var } X$$

$$X \perp Y \Rightarrow \text{Var}(X+Y) = \text{Var } X + \text{Var } Y$$

$$CV_x := \frac{\text{SD } X}{EX} \dots \text{ var. coef}$$

$$\text{Var}(X+Y) = \text{Var } X + \text{Var } Y + 2 \text{Cor}(X, Y)$$

$$\rightarrow \text{Cor}(X, Y) := E[(X-EX)(Y-EY)] = EXY - EX \cdot EY$$

$$\rightarrow \text{Cor}(X, aX+bY+c) = a \text{Cor}(X, Y) + b \text{Cor}(Y, X)$$

$$\rightarrow \text{Cor}(X, Y) := \frac{\text{Cov}(X, Y)}{\text{SD } X \cdot \text{SD } Y} \dots \text{ korelace}$$

Obecné m.r. $X: \Omega \rightarrow \mathbb{R}$

$$\forall x \in \mathbb{R}: \{X \leq x\} \in \mathcal{F}$$

$$F_X(x): x \mapsto P[X \leq x]$$

$\rightarrow F_X$ je: nelinejná, s pravou správ.

$$\rightarrow F_X(\infty) = 1, F_X(-\infty) = 0$$

Diskrétní

$$E[X] = \sum_{x \in \text{Im } X} x \cdot f_X(x)$$

$$E[X] = \sum_{w \in \Omega} X(w) \cdot P(\{w\})$$

$$Y := g(X) \Rightarrow E[Y] = \sum_{x: g(x)=y} f_X(x)$$

$$E[g(X)] = \sum_x g(x) f_X(x)$$

$$E[X|B] = \sum_x x \cdot P[X=x|B]$$

$$E[X] = \sum_i P(B_i) \cdot E[X|B_i]$$

$$f_{X,Y}(x, y) = P[X=x, Y=y]$$

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

$$P[g(X, Y) = z] = \sum_{x, y: g(x, y)=z} f_{X,Y}(x, y)$$

$$X \perp Y \equiv f_{XY} = f_X \cdot f_Y$$

$$P[X+Y = z] = \sum_x P[X=x] f_Y(z-x)$$

$$E[g(X, Y)] = \sum_{x, y} g(x, y) f_{X,Y}(x, y)$$

Kvantilová funkce

$$Q_X(p) = \min \{x \in \mathbb{R} \mid P[X \leq x] \geq p\} \rightarrow \text{median} := Q_X\left(\frac{1}{2}\right)$$

Normativy:

$$\text{Markov: } X \geq 0 \Rightarrow P[X \geq a] \leq \frac{EX}{a} \Leftrightarrow P[X \geq t \cdot EX] \leq \frac{1}{t}$$

$$\text{Chebyšev: } \lambda > 0 \Rightarrow P[|X-EX| \geq \lambda \cdot \text{SD } X] \leq \frac{1}{\lambda^2} \rightarrow \text{normativ } \lambda \cdot \text{SD } X$$

$$X \sim \text{Bin}(m, p), Y \sim \text{Bin}(n, p) \Rightarrow X+Y \sim \text{Bin}(m+n, p)$$

Základy velkých čísel

silný: $P[\lim \bar{X}_m = \mu] = 1$

slabý: $P[|\bar{X}_m - \mu| > \varepsilon] = 0$

$$\hookrightarrow \bar{X}_m \xrightarrow{P} \mu \quad \text{význam}$$

$$E\bar{X}_m = \frac{1}{m} (\sum E X_i) = \frac{m}{m} \mu = \mu$$

$$\text{Var } \bar{X}_m = \frac{1}{m^2} \sum \text{Var } X_i = \frac{m}{m^2} \sigma^2 = \frac{\sigma^2}{m}$$

$$\Rightarrow \sigma(\bar{X}_m) = \sigma/\sqrt{m} \xrightarrow{\text{nez. } X_i}$$

Centrální limitemí věta

X_1, \dots, X_m stejně r. m. n.

$$Y_m := \frac{\bar{X}_m - E\bar{X}_m}{\sigma(\bar{X}_m)} = \frac{\bar{X}_m - \mu}{\sigma/\sqrt{m}} \approx N(0, 1)$$

Intervalové odhady

\hookrightarrow maximální odhad $\hat{\theta}$ \hookrightarrow dolní a horní odhad

$$[D, H] \text{ je } (1-\alpha)\text{-CI} \equiv P[D \leq \theta \leq H] = 1-\alpha$$

Věta: Nechť $\hat{\theta}$ je pro θ , $\hat{\theta} \sim N(\mu, \sigma^2)$
 $\Rightarrow \hat{\theta} \pm z_{1-\alpha/2} \sigma(\hat{\theta})$ je $(1-\alpha)\text{-CI}$
 $\hookrightarrow z_{1-\alpha/2} := \Phi^{-1}(1 - \frac{\alpha}{2})$

Věta: $X_1, \dots, X_m \sim N(\mu, \sigma^2)$, nezávislé, chl. μ
 $\Rightarrow \bar{X}_m \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{m}}$ je $(1-\alpha)\text{-CI}$
 $\hookrightarrow z_{1-\alpha/2} := \Psi_{m-1}^{-1}(1 - \frac{\alpha}{2})$

Bodové odhady ... odhadují parametr θ

$$\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta \quad \dots \hat{\theta} = \text{statistika} \Leftrightarrow \text{odhadují}$$

\hookrightarrow neutrální $\equiv \text{bias} = 0$

$$\hookrightarrow \text{asympt. m.} \equiv \lim_{m \rightarrow \infty} E[\hat{\theta}_m] = \theta$$

$$\hookrightarrow \text{konvergencií} \equiv \hat{\theta}_m \xrightarrow{P} \theta \quad \rightarrow \text{ZVČ: } \bar{X}_m \text{ je konv. m.}$$

$$\text{MSE}(\hat{\theta}) := E[(\hat{\theta} - \theta)^2] = \text{Var } \hat{\theta} + \text{bias}(\hat{\theta})^2$$

$\rightarrow \bar{X}_m$... neutrální k. pro μ

$$\rightarrow \hat{\sigma}^2 := \frac{1}{m-1} \sum_i (X_i - \bar{X}_m)^2 \dots \text{neutrální k. pro } \sigma^2$$

Metoda momentů

$$m_r := E[X^r] \quad \left\{ \begin{array}{l} \text{polární rovnost} \Rightarrow \text{constitutive} \\ \hat{m}_r := \frac{1}{m} \sum_i X_i^r \end{array} \right. \quad \left\{ \begin{array}{l} m_1 = E[X], \quad m_2 = \sigma^2 + \mu^2 \\ \text{metoda} \end{array} \right.$$

Maximum Likelihood

$$\rightarrow \max P[X_1 = x_1, \dots, X_m = x_m] \stackrel{\text{mer.}}{=} \prod_i P[X_i = x_i; \theta]$$

$$\Rightarrow L(\theta, \bar{x}) = \prod_i f(x_i; \theta) = \prod_i f_{X_i}(x_i), \quad \ell(\theta, \bar{x}) = \log L$$

Testování hypotéz

$$H_0 \times H_1$$

\cdot vybrat statistický model: $X \sim N(\mu, \sigma^2)$

\cdot statistický test $T = h(X_1, \dots, X_m)$

\cdot kritický obor: $W \subseteq \mathbb{R}$

$$\Rightarrow \text{chyba 1. druhu} \quad \alpha = P[T \in W | H_0]$$

$$\Rightarrow \text{chyba 2. druhu} \quad \beta = P[T \notin W | H_1]$$

\hookrightarrow malý $\alpha \rightarrow$ slabý test - nerazítka testovací hodnota

$\hookrightarrow 1 - \beta = \text{sila testu}$

$\Rightarrow p\text{-value} = \min \alpha, \text{ pro kterou bych } H_0 \text{ razil}$

Jednostranný test ... $X_1, \dots, X_m \sim N(\mu, \sigma^2) \rightarrow H_0: \mu = \mu_0$

Dvoustranný test ... $X_1, \dots, X_m \sim N(\mu_X, \sigma^2)$
 $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma^2) \rightarrow H_0: \mu_X = \mu_Y$

Parity test ... $X_1, \dots, X_m \sim N(\mu_X, \sigma^2)$
 $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma^2) \rightarrow \text{výběr: } [\mu_X - \mu_Y] = 0$

\hookrightarrow chci statistiku $\bar{X}_m \Rightarrow$ normalizace

$$z\text{-test} \quad Z = \frac{\bar{X}_m - E\bar{X}_m}{\sigma/\sqrt{m}} \sim N(0, 1) \Rightarrow W = \{x \in \mathbb{R} \mid x_1 > z_{1-\alpha/2}\}$$

$$t\text{-test} \quad T = \frac{\bar{X}_m - E\bar{X}_m}{\hat{\sigma}/\sqrt{m}} \sim T(m-1) \Rightarrow W = \{x \in \mathbb{R} \mid |x_1| > t_{1-\alpha/2}\}$$

\hookrightarrow nezávislé $\sigma \Rightarrow$ odhad

$$\hat{\sigma}^2 = \Psi_{m-1}^{-1}(1 - \alpha/2)$$

Testy dobrého shodi

\rightarrow kategorická data

Observed: O_1, \dots, O_k

Expected: E_1, \dots, E_k

$$G\text{-test: } G = 2 \cdot \sum_i O_i \log \frac{O_i}{E_i}$$

$$\chi^2\text{-test: } \chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Kritický obor: $W = \{x \in \mathbb{R} \mid x > \gamma\}$

\hookrightarrow chci $P[T > \gamma | H_0] = \alpha$

Pravdepodobnosť

Def: Pravdepodobnosť prostoru je (Ω, \mathcal{F}, P) , kde

- Ω ... množina el. evení (skôr nesprávne)
- $\mathcal{F} \subseteq P(\Omega)$... množina evení (even = množiny jedin ktorých el. evení)

$\hookrightarrow \mathcal{F} = P(\Omega)$ funguje dobré ju pre správne Ω

$$i, \emptyset, \Omega \in \mathcal{F}$$

$$ii, A \in \mathcal{F} \Rightarrow A^c := \Omega \setminus A \in \mathcal{F}$$

$$iii, A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

- $P: \mathcal{F} \rightarrow [0, 1]$... pravdepodobnosť

$$i, P(\Omega) = 1$$

$$ii, P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{pre } A_i \dot{\cup} A_j \rightarrow \text{fro dom disj. even}$$

⊗ $P(\emptyset) = 0$ $\because P\left(\bigcup_{i=1}^{\infty} \emptyset\right) = \sum_{i=1}^{\infty} P(\emptyset) \dots$ jinak: $\forall \omega \in \Omega, i \in [0, 1] = \infty$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad \text{pre } A_1 \dot{\cup} A_2 \quad (A_3, A_4, \dots = \emptyset)$$

⊗ $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F} \quad \because A_3, A_4, \dots = \emptyset$
 $\Rightarrow A_1 \cap A_2 \in \mathcal{F} \quad \because A_1 \cap A_2 = (A_1^c \cup A_2^c)^c$



Tworzenie: Nechť (Ω, \mathcal{F}, P) je f.f. a $A, B \in \mathcal{F}$. Potom

$$1, P(A) + P(A^c) = 1$$

$$2, A \subseteq B \Rightarrow P(A) \leq P(B) \quad \& \quad P(B \setminus A) = P(B) - P(A).$$



$$3, P(A \cup B) = P(A) - P(B) + P(A \cap B)$$



$$4, P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \quad \dots \text{subadditívna}$$

Dôk:
 1, $\Omega = A \dot{\cup} A^c \Rightarrow P(\Omega) = P(A \dot{\cup} A^c) = P(A) + P(A^c)$

$$2, B = A \dot{\cup} (B \setminus A) \Rightarrow P(B) = P(A) + P(B \setminus A) \quad \& \quad P(B \setminus A) \geq 0 \quad \because \text{je pos.}$$

$$3, A \cup B = A \dot{\cup} (B \setminus (A \cap B)) \Rightarrow P(A \cup B) = P(A) + P(B \setminus (A \cap B)) = P(A) + P(B) - P(A \cap B)$$

$$4, B_i \subseteq A_i: B_1 := A_1, B_i := A_i \setminus \bigcup_{j < i} A_j \Rightarrow B_1, B_2, \dots \text{fro dom disj.}$$

$B_1 \subseteq A_1 \quad \dots \rightarrow \text{premeny pres}$

$B_2 \subseteq A_2$

$\Rightarrow P(UA) = P(UB) = \sum_i P(B_i) \leq \sum_i P(A_i) \quad \text{minaf i. a l.}$

$B_1 \subseteq A_1$

$B_2 \subseteq A_2$

$\Rightarrow P(UA) = P(UB) = \sum_i P(B_i) \leq \sum_i P(A_i) \quad \text{minaf i. a l.}$

Příklady p.f.:

• štatistiky: Ω konečná & uniformní $P \Rightarrow P(A) = \frac{|A|}{|\Omega|}$

• diskrétní: Ω s početnou

$$\mu: \Omega \rightarrow [0,1], \quad \sum_{\omega \in \Omega} \mu(\omega) = 1, \quad P(A) = \sum_{\omega \in A} \mu(\omega)$$

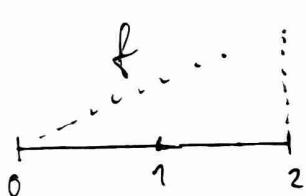
• geometrický: $\Omega \subseteq \mathbb{R}^d \dots \text{---} \text{---} \text{---} \subseteq \mathbb{R}^2$

$$\mathcal{F} = \{A \in \mathcal{P}(\Omega) \mid \text{Vol}(A) \text{ je definován}\}. \dots \text{různé cíhly a tisk}$$

$$P(A) = \frac{\text{Vol}(A)}{\text{Vol}(\Omega)}$$

• spojity: $\Omega \subseteq \mathbb{R}^d \dots [0,1] \subseteq \mathbb{R}$

$$f: \Omega \rightarrow [0, \infty), \quad \int_{\Omega} f = 1, \quad P(A) = \int_A f$$



$$f(x) = \frac{1}{2}x$$

$$\Omega = [0, 2]$$

$$\int_0^2 f(x) dx = 1 \quad \checkmark$$

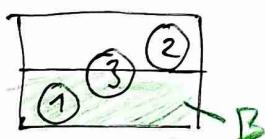
$$\Rightarrow P([1, 2]) = \int_1^2 f(x) dx = \frac{x^2}{4} \Big|_1^2 = 0.75$$

Podmíněná pravděpodobnost

Def: Při jevu A je podmínky, že následuje B je

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0.$$

Význam:



$$P(B) = \frac{1}{2}$$

$$P(1) = P(2) = P(3) = \frac{1}{10}$$

$$P(1|B) = \frac{P(1)}{P(B)} = \frac{\frac{1}{10}}{\frac{1}{2}} = \frac{1}{5}$$

$$P(2|B) = \frac{P(2)}{P(B)} = 0$$

$$P(3|B) = \frac{P(3 \cap B)}{P(B)} = \frac{1}{10}.$$

Tvorcení: Nechť je (Ω, \mathcal{F}, P) pr. f., $B \in \mathcal{F}$ a $P(B) > 0$. Pro $A \in \mathcal{F}$ definujeme $Q(A) := P(A|B)$. Tvrzme, že (Ω, \mathcal{F}, Q) je také pr. f.

Def: Finice Q musí byt pravděpodobnost, tedy musí splňovat

$$i) Q(\Omega) = 1 \dots Q(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1 \quad \checkmark$$

$$ii) Q\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Q(A_i), \quad A_i \cup A_j \xrightarrow{\text{de-Morgan}} \text{disj.}$$

$$\rightarrow Q\left(\bigcup A_i\right) = P\left(\bigcup A_i | B\right) = \frac{P\left(\bigcup A_i \cap B\right)}{P(B)} = \frac{1}{P(B)} \cdot P\left(\bigcup (A_i \cap B)\right) = \\ = \frac{1}{P(B)} \cdot \sum_i P(A_i \cap B) = \sum_i P(A_i | B) = \sum_i Q(A_i). \quad \blacksquare$$

Příklad: $A_i \sim i$ -ta karta v balíčku 32 karet je \heartsuit .

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1) = \frac{8}{32} \cdot \frac{7}{31}$$

\hookrightarrow Edgbyc hohl čísel z definice, kde jsou sice jen rozličné \Rightarrow vlastně k rozdílu a novem f. f.

Věta (o řetězencovém počítání): Nechť $A_1, \dots, A_m \in \mathcal{F}$, $P(A_1 \cap \dots \cap A_{m-1}) > 0$.

$$P(A_1 \cap \dots \cap A_m) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \dots \cdot P(A_m | A_1 \cap \dots \cap A_{m-1}).$$

Def: indukci.

Def: Rozklad možností Ω je systém disj. možností $B = \{B_1, B_2, \dots\}$ t. k. $\xrightarrow{\text{speciálně mnoho}}$

$$\forall i: B_i \subseteq \Omega \quad \& \quad \bigcup B = \Omega$$

Věta (o rozložení podle možností): Nechť B je rozklad Ω a $A \in \mathcal{F}$. Potom platí

$$P(A) = \sum_i P(B_i) \cdot P(A | B_i), \quad \text{kde pro } P(B_i) = 0 \text{ dodefinujeme } P(A | B_i) = 0.$$

Def:



$$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup \dots \quad \xrightarrow{\text{disj.}}$$

$$\Rightarrow P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i) \cdot P(A | B_i). \quad \blacksquare$$

Věta (Bayesova): Nechť D je rozklad Ω a $A \in \mathcal{F}$. Platí

$$P(B_j | A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(B_j) \cdot P(A | B_j)}{\sum_i P(B_i) \cdot P(A | B_i)}.$$

$$\rightarrow \text{speciálne } P(D|A) = \frac{P(A|D) \cdot P(D)}{P(A)}.$$

Příklad: Gambler's ruin.

→ fiktivní hráč: máže \$1 a je 50% šance vyhrát \$1 každ.

$$f(a) := P[\text{odejde s } \$N, \text{ když zacal s } \$a] \quad 0 \leq a \leq N$$

$$A \sim f(a)$$

$$B \sim \text{vyhra v 1. kole}$$

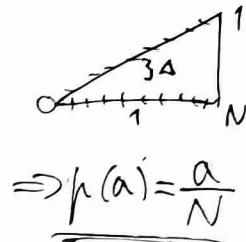
$$f(0) = 0$$

$$f(N) = 1$$

$$\left. \begin{array}{l} P(A) = P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c) \\ \Rightarrow f(a) = \frac{1}{2} \cdot f(a+1) + \frac{1}{2} \cdot f(a-1) \end{array} \right\}$$

$$2f(a) = f(a+1) + f(a-1)$$

$$\Delta = f(a+1) - f(a) = f(a) - f(a-1)$$



$$\Rightarrow f(a) = \frac{a}{N}$$

Příklad:

$$a = m+1 - \text{bitové číslo}$$

$$b = m - \text{bitové číslo}$$

$$\begin{matrix} 1 & 0 & 1 & \dots & 0 & | & 1 \\ 0 & 1 & 1 & \dots & \dots & 1 \end{matrix}$$

$$P[|a|_1 > |b|_1] = ? \quad \Rightarrow X := \text{fikt. řada } 2n \text{-bit čísel k sítě} \#1$$

→ poslední bit a je luhn

✓ fikt. řada n je prim a nich je několik

$$\bullet 0 \Rightarrow P[|a|_1 > |b|_1 | 0] = \frac{1-x}{2}$$

$$\bullet 1 \Rightarrow P[|a|_1 > |b|_1 | 1] = x + \frac{1-x}{2} \rightarrow \text{sítě několik n je několik}$$

$$\Rightarrow P[|a|_1 > |b|_1] = \frac{1}{2} \cdot \frac{1-x}{2} + \frac{1}{2} \left(x + \frac{1-x}{2} \right) = \frac{1}{2} \cdot (x+1-x) = \frac{1}{2}$$

Měřitelnost ještě

Def: Lety $A, B \in \mathcal{F}$ jsou měřitelné, $A \perp B \equiv P(A \cap B) = P(A) \cdot P(B)$

$$\circlearrowleft A \perp B \Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A) \Leftrightarrow P(A|B) = P(A)$$

Def: Pokud $P(A|B) > P(A)$... A, B jsou pozitivně korelované,
 $P(A|B) < P(A)$... A, B jsou negativně korelované

Def: Nechť I je indexová množina. Lety $\{A_i \mid i \in I\}$ jsou měřitelné \equiv

$$\forall J \subseteq I : P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

Def: Lety A, B jsou měřitelné na formálně C , $A \perp_C B \equiv$

$$P(A \cap B | C) = P(A|C) \cdot P(B|C)$$

Tvrdění: $A \perp B \Rightarrow A \perp B^c \quad \& \quad A^c \perp B^c$



$$\text{Důkaz: } P(A \cap B^c) = P(A \setminus (A \cap B)) = P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(B^c)$$

$$A^c \perp B^c \quad \because A \perp B^c \Rightarrow A^c \perp B^c.$$

Diskrétní náhodné veličiny

Def: Nechť (Ω, \mathcal{F}, P) je pr. f. Funkce $X: \Omega \rightarrow \mathbb{R}$ je diskrétní m.v. \Leftrightarrow

- 1, $\text{Im}(X)$ je nejvyšší spočetný ... obor hodnot
- 2, $\forall x \in \mathbb{R}: \{X=x\} := \{\omega \in \Omega | X(\omega) = x\} \in \mathcal{F}$ $\xrightarrow{-1} X^{-1}(x)$

\hookrightarrow tedy chceme, abychom se mohli plát na $P[X=x]$

Poznámka: Náhodná veličina X určuje p.p. $(\Omega', \mathcal{F}', \mathbb{P})$, kde

$$\Omega' = \text{Im}(X), \quad \mathcal{F}' = \mathbb{P}(\Omega'), \quad \mathbb{P}(x) = P[X=x]$$

\rightarrow tento prostor, resp. fci \mathbb{P} , nazýváme rozdělení / distribuce X .

Def: Pravděpodobnostní fce (probability mass function = pmf) d.m.v. X je $p_x: \mathbb{R} \rightarrow [0, 1]$, $x \mapsto P[X=x]$

$$\textcircled{1} \quad \sum_{x \in \text{Im}X} p_x(x) = 1 \quad \because \Omega = \bigcup_{x \in \text{Im}X} \{X=x\} \quad \begin{matrix} \Omega \\ \xrightarrow{P} \end{matrix} \quad \begin{matrix} \Omega' \\ \xrightarrow{\text{rozdělení}} \end{matrix}$$

Bernoulliho rozdělení

$$X \sim \text{Ber}(p) \quad \dots \quad P[X=1] = p_x(1) = p \\ P[X=0] = p_x(0) = 1-p$$

Indikátor jevu $A \in \mathcal{F}$

$$I_A(\omega) = \begin{cases} 1 & \dots \omega \in A \\ 0 & \dots \omega \notin A \end{cases} \quad \begin{matrix} \dots A \text{ nastal} \\ \dots A \text{ nezaslal} \end{matrix} \quad \textcircled{2} \quad I_A \sim \text{Ber}(P(A))$$

Geometrické rozdělení

X = na kolikátky fokus se poprvé něco povedlo (fokus je meravisek s pravd. p)

$$X \sim \text{Geom}(p)$$

$$\begin{aligned} p_x(1) &= p \\ p_x(2) &= (1-p)p \\ p_x(3) &= (1-p)^2 p \end{aligned} \quad \Rightarrow p_x(k) = \begin{cases} (1-p)^{k-1} p & , \text{ pro } k \in \mathbb{N}^+ \\ 0 & , \text{ jinak} \end{cases}$$

\rightarrow ještě potřebujeme $\sum_{x \in \text{Im}X} p_x(x) = 1$

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

Binomické rozdělení

$X = \#\text{náspěchů při } n \text{ nezávislých pokusech s pravd. } p$

$$X \sim \text{Bin}(n, p), \quad p_X(k) = \binom{n}{k} \cdot p^k (1-p)^{n-k} \quad \Rightarrow \sum_{k=0}^{\infty} p_X(k) = (p + (1-p))^n = 1$$

⊗ Součet nes. bin. rozd. je bin. rozdělení

$$X \sim \text{Bin}(m, p), \quad Y \sim \text{Bin}(n, p), \quad X+Y \Rightarrow Z = X+Y \sim \text{Bin}(m+n, p)$$

$$\begin{aligned} p_Z(k) &= P[X+Y=k] = \sum_{i=0}^k P[X=i \ \& \ Y=k-i] = \sum_{i=0}^k p_X(i) \cdot p_Y(k-i) \\ &= \sum_{i=0}^k \binom{m}{i} p^i (1-p)^{m-i} \binom{n}{k-i} \cdot p^{k-i} (1-p)^{n-k+i} = p^k (1-p)^{m+n-k} \underbrace{\sum_{i=0}^k \binom{m}{i} \binom{n}{k-i}}_{\binom{m+n}{k}} \end{aligned}$$

Poissonovo rozdělení

→ aproximace bin. r. pro velké n a malou p

$$X \sim \text{Pois}(\lambda) \quad \dots \quad p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0$$

$$\hookrightarrow \sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1 \quad \checkmark$$

⊗ $\text{Pois}(\lambda) \approx \text{Bin}(n, \frac{\lambda}{n})$

\hookrightarrow nechť $X \sim \text{Pois}(\lambda)$ a $X_m \sim \text{Bin}(m, \frac{\lambda}{m}) \Rightarrow$ Budeme $\lim_{m \rightarrow \infty} p_{X_m}(k) = p_X(k)$

$$\begin{aligned} p_{X_m}(k) &= \binom{m}{k} \cdot \left(\frac{\lambda}{m}\right)^k \cdot \left(1 - \frac{\lambda}{m}\right)^{m-k} = \frac{m^k}{k!} \cdot \frac{\lambda^k}{m^k} \cdot \left(1 - \frac{\lambda}{m}\right)^m \cdot \left(1 - \frac{\lambda}{m}\right)^{-k} \xrightarrow[m \rightarrow \infty]{1} i \quad k \in \mathbb{N}_0 \\ &= \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{m}\right)^m \cdot \cancel{\frac{m^k}{m^k}} = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = p_X(k) \quad \dots \quad \left(1 + \frac{\lambda}{m}\right)^m \xrightarrow{m \rightarrow \infty} e^\lambda \end{aligned}$$

Sřední hodnota

Def: Sřední hodnota d.m.r. X je $\mathbb{E}X := \sum_{x \in \text{Im } X} x \cdot p_X(x)$.

↳ pokud řada nekonverguje absolutně, tak není definována

⊗ Pokud je X d.m.r. na diskrétním p.f. (tedy Ω je nejvýše spočetná), tak

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

$$\text{Dle: } \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{x \in \text{Im } X} \sum_{\omega: X(\omega)=x} X(\omega) P(\{\omega\}) = \sum_{x \in \text{Im } X} x \sum_{\omega \in X^{-1}(x)} P(\{\omega\}) = \sum_{x \in \text{Im } X} x \cdot P[X=x]$$

$$X^{-1}(x)$$

Pravidlo:

$$\textcircled{1} \quad X \sim \text{Ber}(p) \Rightarrow \mathbb{E}X = 1 \cdot p + 0 \cdot (1-p) = \underline{p}$$

$$\textcircled{2} \quad X \sim \text{Bin}(n, p) \Rightarrow \mathbb{E}X = \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} =$$

$$\mathbb{E}X = \underline{p \cdot n}$$

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

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

Lemma o dílčinu:

$$\textcircled{3} \quad X \sim \text{Geom}(p) \Rightarrow \mathbb{E}X = \sum_{k=1}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = p \cdot \sum_{k=1}^{\infty} k (1-p)^{k-1}$$

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \Rightarrow \frac{d}{dx} \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2} \quad \left| \quad = p \cdot \frac{1}{(1-(1-p))^2} = \underline{\frac{1}{p}}$$

Pravidlo mainovního statistika - PNS

⊗ Pokud je X d.m.r. a $g: \mathbb{R} \rightarrow \mathbb{R}$, potom je $Y := g(X)$ také d.m.r.

Dle: Chceme Y d.m.r., tedy

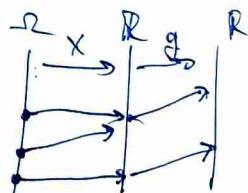
$$1, \quad Y: \Omega \rightarrow \mathbb{R} \quad \checkmark$$

$$2, \quad \text{Im}(Y) \text{ je nejvýše spočetný} \quad \because |\text{Im}(X \circ g)| \leq |\text{Im}(X)|$$

$$3, \quad \text{pro } y \in \mathbb{R}: \quad Y^{-1}(y) = \{Y=y\} \in \mathcal{F}$$

$$Y^{-1}(y) = \bigcup_{\substack{x \in \text{Im } X \\ g(x)=y}} X^{-1}(x) \quad \& \quad X^{-1}(x) \in \mathcal{F} \Rightarrow Y^{-1}(y) \in \mathcal{F}$$

→ speciálně sjednocením ↳ jedná se o axiomy pravidelností



Věta (PNS): Pokud X je d.m.r. a $g: \mathbb{R} \rightarrow \mathbb{R}$, potom y je určen jednoznačně,

$$\mathbb{E}[g(X)] = \sum_{x \in \text{Im } X} g(x) p_X(x)$$

$$\sum_x \sum_{y: g(x)=y} g(x) p_X(x) = \sum_x g(x) p_X(x)$$

Dle: Označme $Y := g(X)$. Z definice

$$\mathbb{E}[Y] = \sum_{y \in \text{Im } Y} y \cdot P[Y=y] = \sum_y y \cdot \sum_{x: g(x)=y} P[X=x] = \sum_y \sum_{x: g(x)=y} g(x) p_X(x)$$

$$\rightarrow \text{jako v minulém dílčaru: } \{Y=y\} = Y^{-1}(y) = \bigcup_{x: g(x)=y} X^{-1}(x) = \bigcup_{\substack{x \in \text{Im } X \\ g(x)=y}} \{X=x\}$$



Věta (vlastnosti E): X, Y jsou d.m.r.

$$\textcircled{1} \quad P[X \geq 0] = 1 \quad \& \quad \mathbb{E}X = 0 \Rightarrow P[X=0] = 1$$

$$\textcircled{2} \quad \mathbb{E}X \geq a \Rightarrow P[X \geq a] > 0$$

$$\textcircled{3} \quad \mathbb{E}[aX+b] = a\mathbb{E}X + b$$

$$\textcircled{4} \quad \mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y \quad \dots \text{ zatím pouze pro diskrétní f-f.}$$

Dle: $\textcircled{1} \quad \mathbb{E}X = \sum_x x \cdot p_X(x) = 0$, všechny členy $\geq 0 \Rightarrow$ všechny členy = 0

$$\textcircled{2} \quad \mathbb{E}X = \sum_x x \cdot p_X(x) \geq a. \text{ Když } P[X \geq a] = 0, \text{ tak } \forall x \geq a: P[X=x] = 0 \\ \Rightarrow \text{tedy } \sum_x x \cdot p_X(x) < \sum_x a \cdot p_X(x) = a \sum_x p_X(x) = a \Rightarrow a \leq \sum x < a$$

$$\textcircled{3} \quad \mathbb{E}[aX+b] = \mathbb{E}[g(X)] = \sum_x (ax+b) p_X(x) = a \sum_x x p_X(x) + b \sum_x p_X(x) = a\mathbb{E}X + b.$$

$$\textcircled{4} \quad \mathbb{E}[X+Y] = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot P\{\{\omega\}\} = \sum_{\omega} X(\omega) P(\omega) + \sum_{\omega} Y(\omega) P(\omega) = \mathbb{E}X + \mathbb{E}Y$$



Linearity střední hodnoty: $\mathbb{E}\left[\sum_i a_i X_i\right] = \sum_i a_i \mathbb{E}X_i$

→ aplikace: $X \sim \text{Bin}(n, p) \rightarrow \mathbb{E}X = ?$ (víme n-p)

$X = \# \text{ úspěchů} \text{ v } n \text{ pokusech} \Rightarrow X_i = \text{Indikátor, že } i\text{-tý pokus úspěš}$

$$\mathbb{E}X = \sum_{i=1}^n \mathbb{E}X_i = n \cdot p \quad \hookrightarrow X_i \sim \text{Ber}(p) \Rightarrow \mathbb{E}X_i = p$$

Dle (podmíněná E): Nechť je X d.m.r., B j.e.v., $P(B) > 0$. Definujeme

$$\mathbb{E}[X|B] := \sum_{x \in \text{Im } X} x \cdot P[X=x|B] \rightarrow \text{opět pokud je absolutně soudružející}$$

Věta (o celkové E): Nechť je B_1, B_2, \dots rozdělení Σ a X je d.m.r. Potom

$$\mathbb{E}X = \sum_i P(B_i) \cdot \mathbb{E}[X|B_i]$$

$$\text{Dle: } \sum_x x P[X=x] = \sum_x x \cdot \sum_i P(B_i) \cdot P[X=x|B_i] = \sum_i P(B_i) \cdot \underbrace{\sum_x x \cdot P[X=x|B_i]}_{\mathbb{E}[X|B_i]}$$

↪ věta o celkové f.d.f.



Příklady:

① $X = \text{celkové skóre při házení kostkou, očekávají řadu } 6$

$$\mathbb{E}X = P(\text{:}) \cdot \mathbb{E}[X|\text{:}] + (1-P(\text{:})) \cdot \mathbb{E}[X|\text{:}^c]$$

$$= \frac{1}{6} \cdot (6 + \mathbb{E}X) + \frac{5}{6} \cdot 3 \quad \hookrightarrow \frac{1}{5}(1+2+3+4+5) = 3$$

$$= \frac{1}{6} \mathbb{E}X + \frac{7}{2} \quad \Rightarrow \quad \mathbb{E}X = \frac{42}{10} = \underline{\underline{4.2}}$$

② m-bit číslo, $P(1) = p$

$$X = \#\text{podílůrčí 101} \Rightarrow \mathbb{E}X = ?$$

1234
101100

$$A_i := \text{ma } i\text{-té pozici racíma 101} \rightarrow P[A_i] = p(1-p)^{i-1} = p^2(1-p)$$

$$\rightarrow X = \sum_{i=1}^{m-2} I_{A_i} \Rightarrow \mathbb{E}X = \sum_{i=1}^{m-2} \mathbb{E}I_{A_i} = \underline{(m-2) \cdot p^2(1-p)} \quad \Rightarrow \mathbb{E}I_{A_i} = p^2(1-p)$$

③ Nechť A je pro α I_A jeho indikátor. Potom $A = A_1 \cup A_2 \cup \dots \cup A_m$, kde

$$1 - \mathbb{E}I_A = \prod_{i=1}^m (1 - \mathbb{E}I_{A_i}) \quad \dots \text{ celkově obrovský}$$

$$\mathbb{E}[1 - \mathbb{E}I_A] = 1 - P(A) = \mathbb{E}\left[\prod_{i=1}^m (1 - \mathbb{E}I_{A_i})\right]$$

\hookrightarrow rovnásobit + lim E \Rightarrow princip i a vše

Větice: Pokud je X d.m.v. a $\text{Im}(x) \subseteq \mathbb{N}$, potom

$$\mathbb{E}X = \sum_{m=0}^{\infty} P[X > m] \rightarrow s(t) := P[X > t] \dots \text{survival function}$$

$$\text{Dle: } \mathbb{E}X = 1 \cdot p_x(1) + 2 \cdot p_x(2) + 3 \cdot p_x(3) + \dots$$

$$= p_1 + p_2 + p_3 + p_4 + \dots = P(X > 0)$$

$$+ p_2 + p_3 + p_4 + \dots = P(X > 1)$$

$$+ p_3 + p_4 + \dots = P(X > 2)$$

$$+ p_4 + \dots = P(X > 3)$$

Příklad: $X \sim \text{Geom}(p)$

$$\mathbb{E}X = \sum_{m=0}^{\infty} P(X > m) = \sum_{m=0}^{\infty} (1-p)^m = \frac{1}{1-(1-p)} = \frac{1}{p}$$

Def: Rozptyl d.m.v. X je $\text{var}(X) := \mathbb{E}[(X - \mathbb{E}X)^2]$

Smeřodatná odchylka $\sigma_X := \sqrt{\text{var}X}$

Variacní koeficient $CV_X := \frac{\sigma_X}{\mathbb{E}X}$... pouze pro $\mathbb{E}X > 0$

Věta: $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2$

Důk: Označme $\mu := \mathbb{E}X$

$$\text{var}X = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2X\mu + \mu^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}X + \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

$$\Rightarrow \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mu \Rightarrow \mathbb{E}[X(X-1)] + \mu - (\mathbb{E}X)^2 = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \text{var}X$$

⊗ $\text{var}(X) \geq 0 \Rightarrow \mathbb{E}[X^2] \geq (\mathbb{E}X)^2$

Různobokost: nerovnost mezi aritmetickým a kvadratickým průměrem

$$\frac{x_1^2 + x_2^2 + \dots + x_m^2}{m} \geq \left(\frac{x_1 + x_2 + \dots + x_m}{m} \right)^2 \Rightarrow \frac{1}{m} \sum_i x_i \leq \sqrt{\frac{1}{m} \sum_i x_i^2}$$

$$\rightarrow \text{dodatek plati i pro vážený průměr: } \sum_i x_i w_i \leq \sqrt{\sum_i x_i^2 w_i}, \sum_i w_i = 1$$

⊗ $\text{var}(X) = 0 \Leftrightarrow X$ je konstantní fá

Příklady:

① $X \sim \text{Ber}(\mu) \Rightarrow \mathbb{E}X = \mu$

$$\hookrightarrow \text{Var}(X) = \mathbb{E}[(X-\mu)^2] = \mu \cdot (1-\mu)^2 + (1-\mu) \cdot \mu^2 = \underline{\mu(1-\mu)}$$

$$\hookrightarrow X^2 = X \Rightarrow \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \mu - \mu^2$$

dodatek kružnice

② $X \sim \text{Bin}(n, p) \Rightarrow \mathbb{E}X = np$

a, z definice: PNS: $g(x) = (x-np)^2 \Rightarrow \text{Var}X = \sum_{\ell=0}^n (\ell-np)^2 \binom{n}{\ell} p^\ell (1-p)^{n-\ell}$

b, $\text{Var}X = \mathbb{E}[X(X-1)] + np - (np)^2$

$$\begin{aligned} \mathbb{E}[X(X-1)] &\stackrel{\text{PNS}}{=} \sum_{\ell=0}^n \ell(\ell-1) \binom{n}{\ell} p^\ell (1-p)^{n-\ell} = \sum_{\ell=2}^n (\ell-1) \ell \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \\ &= (n-1)n \sum_{\ell=2}^n \binom{n-2}{\ell-2} p^\ell (1-p)^{n-\ell} = n(n-1)p^2 \sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^\ell (1-p)^{n-2-\ell} \\ &= n(n-1)p^2 \cdot (p + (1-p))^{n-2} = n(n-1)p^2 \end{aligned}$$

$$\Rightarrow \text{Var}X = p^2 n(n-1) + np - np^2 n^2 = \underline{n \cdot p \cdot (1-p)} \rightarrow n \text{-bohat než Ber}(p)$$

③ $X \sim \text{Pois}(\lambda)$

$$\rightarrow \text{obdobně } \mathbb{E}X = \text{Var}X = \lambda \rightsquigarrow \text{Pois}(\lambda) \approx \text{Bin}(n, \frac{\lambda}{n}) \rightsquigarrow \mathbb{E}X = \lambda = n \cdot \frac{\lambda}{n}.$$

$$\rightsquigarrow \text{Var}X = \lambda \approx \lambda \cdot (1 - \frac{\lambda}{n})$$

$$\textcircled{4} \quad X \sim \text{Geom}(p) \Rightarrow \mathbb{E}X = \frac{1}{p}$$

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \mathbb{E}[X^2] - \frac{1}{p^2}$$

$$\Rightarrow \begin{cases} B_1 = 1. \text{ počet nářadí} \\ B_2 = B_1^c \end{cases} \quad \left. \begin{aligned} \mathbb{E}[X^2] &= P(B_1) \cdot \mathbb{E}[X^2 | B_1] + P(B_2) \cdot \mathbb{E}[X^2 | B_2] = \\ &= p \cdot 1^2 + (1-p) \cdot \mathbb{E}[(1+X)^2] = \\ &= p + (1-p) \cdot \left(1 + 2 \cdot \frac{1}{p} + \mathbb{E}[X^2]\right) \end{aligned} \right.$$

$$\Rightarrow \mathbb{E}[X^2] (1 - (1-p)) = p + 1 - p + \frac{2(1-p)}{p} \Rightarrow \mathbb{E}[X^2] = \frac{2-p}{p^2}$$

$$\Rightarrow \text{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \underline{\underline{\frac{1-p}{p^2}}}$$

$$\textcircled{5} \quad Y = aX + b$$

$$\Rightarrow \text{Var}(Y) = a^2 \text{Var}(X) \quad \left. \begin{aligned} &\text{imamu' vici posunum} \\ &\sigma(Y) = |a| \sigma(X) \end{aligned} \right\}$$

$$CV_Y = CV_X, \text{ pro } a \geq 0 \Rightarrow CV \text{ je imamu' vici čálováni}$$

• Náhodné veličiny → na stejném f.p.

Def: Pro d.m.r. X, Y definujeme souřazenou podmífkou (joint pmf) jako

$$p_{X,Y}(x,y) := P[X=x \& Y=y]$$

Aby to byla f.s., musí pro $\forall x, y \in \mathbb{R}$: $\{X=x \& Y=y\} \in \mathcal{F}$

Výzva:

$$\begin{array}{c|cc|c} x \backslash y & 0 & 1 & 2 \\ \hline 0 & \frac{1}{12} & \frac{1}{6} & \frac{1}{12} \\ 1 & \frac{1}{6} & \frac{1}{4} & \frac{1}{12} \end{array} \rightarrow p_X(0) = P[X=0] = \frac{1}{6} + \frac{1}{6} + \frac{1}{12} \\ p_Y(2) = P[Y=2] = \frac{1}{12} + \frac{1}{12}$$

Def: Rozdělení $p_{X,Y}$ říkáme souřazené rozdělení. Rozdělení jednotlivých složek p_X a p_Y nazýváme marginalní rozdělení

$$\textcircled{1} \quad p_X(x) = \sum_{y \in \text{Im } Y} p_{X,Y}(x,y) \quad p_Y(y) = \sum_{x \in \text{Im } X} p_{X,Y}(x,y)$$

⚠️ Z m.r. nelze odvodit souřazeninu.

↑ přes disj. sjednocení

$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \quad \text{a} \quad \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \end{array} \quad \text{mají stejné marg. rozdělení}$$

↳ jde tedy o r. v. k. d. jsm X, Y nezávislé!

Věta: Pokud je (X, Y) d.m. vektor a $g: \mathbb{R}^2 \rightarrow \mathbb{R}$, potom je $g(X, Y)$ d.m.v. a platí

$$\mu_Z(z) = P[g(X, Y) = z] = \sum_{x, y : g(x, y) = z} \mu_{X, Y}(x, y)$$

Důkaz: Zjistěte je $\{g(x, y) = z\}$ disj. sjednocením fází $\{X=x \& Y=y\}$ pro dvojice, pro které $g(x, y) = z$. Navíc se shání řešení na $x \in \text{Im } X$ a $y \in \text{Im } Y$, takže to sjednocení je speciální \Rightarrow lze použít axiom $P(V) = \sum P$.

Nerávnost d.m.v.

Def: X, Y jsou nerávnost d.m.v., $X \perp Y \equiv \forall x, y \in \mathbb{R} : \{X=x\} \perp \{Y=y\}$

$$\diamond X \perp Y \Leftrightarrow \forall x, y : \mu_{X, Y}(x, y) = \mu_X(x) \mu_Y(y)$$

Def: X_1, \dots, X_m jsou nerávnost \equiv jestli $\{X_1=x_1\}, \dots, \{X_m=x_m\}$ jsou nerávnost.

Multinomické rozdělení

Hájíme m -bráť kostkou ... fírovou $\Rightarrow \mu_1 = \mu_2 = \dots = \mu_6 = \frac{1}{6}$

$$\begin{aligned} X_1 &:= \# \text{ jednice} \sim \text{Bin}(m, \mu_1) \\ X_2 &:= \# \text{ dvouček} \sim \text{Bin}(m, \mu_2) \end{aligned} \quad \left. \begin{array}{l} (X_1, \dots, X_6) \sim \text{Mult}(m, \mu_1, \dots, \mu_6) \end{array} \right.$$

$$\rightarrow \mu_{X_1, \dots, X_6}(\varepsilon_1, \dots, \varepsilon_6) = \binom{m}{\varepsilon_1, \dots, \varepsilon_6} \cdot \mu_1^{\varepsilon_1} \cdot \mu_2^{\varepsilon_2} \cdots \cdot \mu_6^{\varepsilon_6}$$

Def: Multinomický koeficient $(\varepsilon_1, \dots, \varepsilon_m) := \frac{m!}{\varepsilon_1! \cdots \varepsilon_m!}, \quad \sum_i \varepsilon_i = m$

\hookrightarrow intuice: $\#$ různobitých rozdělit m průběhu do m mřímků $\varepsilon_1, \dots, \varepsilon_m$ pravidly

Věta (konvoluční vzorec): Nechť X, Y jsou d.m.v. a $Z := X + Y$. Potom platí

$$\mu_Z(z) = P[X+Y=z] = \sum_{x \in \text{Im } X} P[X=x \& Y=z-x]$$

\diamond pokud $X \perp Y$, potom $\mu_Z(z) = \sum_x \mu_X(x) \mu_Y(z-x)$.

Důkaz: Použijeme předchozí větu.

Věta (PNS pro nekory): Nechť (X, Y) je d.m.r. nekory a $g: \mathbb{R}^2 \rightarrow \mathbb{R}$. Potom platí

$$\mathbb{E}[g(X, Y)] = \sum_{x,y} g(x, y) \cdot p_{X,Y}(x, y)$$

Důkaz: Označme $Z := g(X, Y)$ → podle něj je předchozí stránky

$$\begin{aligned}\mathbb{E} Z &= \sum_z z \cdot \mathbb{P}_Z(z) = \sum_z z \cdot \sum_{x,y: g(x,y)=z} p_{X,Y}(x, y) = \sum_z \sum_{x,y: g(x,y)=z} z \cdot p_{X,Y}(x, y) \\ &= \sum_{x,y} \sum_{z: g(x,y)=z} z \cdot p_{X,Y}(x, y) = \sum_{x,y} g(x, y) \cdot p_{X,Y}(x, y).\end{aligned}$$



→ důkaz jen diskretní

Věta: Nechť jsou X, Y d.m.r. až libovolném f.f., $a, b \in \mathbb{R}$. Potom

$$\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y.$$

Důkaz: Použijeme PNS a $g(x, y) := ax + by$.

$$\begin{aligned}\mathbb{E}[aX + bY] &= \sum_{x,y} (ax + by) \cdot p_{X,Y}(x, y) = a \sum_{x,y} x \cdot p_{X,Y}(x, y) + b \sum_{x,y} y \cdot p_{X,Y}(x, y) \\ &\rightsquigarrow \sum_{x,y} x \cdot p_{X,Y}(x, y) = \sum_x x \cdot \sum_y p_{X,Y}(x, y) = \sum_x x \cdot p_X(x) = \mathbb{E}X\end{aligned}$$



Věta: $X \perp Y \Rightarrow \mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y$.

Důkaz: PNS pro funkci $g(x, y) = xy$:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{x,y} xy \cdot p_{X,Y}(x, y) = \sum_{x,y} x \cdot y \cdot p_X(x) \cdot p_Y(y) = \sum_x x \sum_y y \cdot p_X(x) \cdot p_Y(y) \\ &= \sum_x x \cdot p_X(x) \cdot \sum_y y \cdot p_Y(y) = \mathbb{E}X \cdot \mathbb{E}Y\end{aligned}$$

• Kovariance

Def: Pro m.r. X, Y definujeme jejich kovarianci jako

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Tvrdění: $\text{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X \cdot \mathbb{E}Y$... linearita sibiřské hodnoty

$$\Leftrightarrow X \perp Y \Rightarrow \text{cov}(X, Y) = 0$$

Tvrdění: Vlastnosti kovariance:

$$1) \text{var}(X) = \text{cov}(X, X)$$

$$2) \text{cov}(X, aY + bZ + C) = a \cdot \text{cov}(X, Y) + b \cdot \text{cov}(X, Z) \dots \text{linearita}$$

Věta: Pro d.m.v. X, Y platí

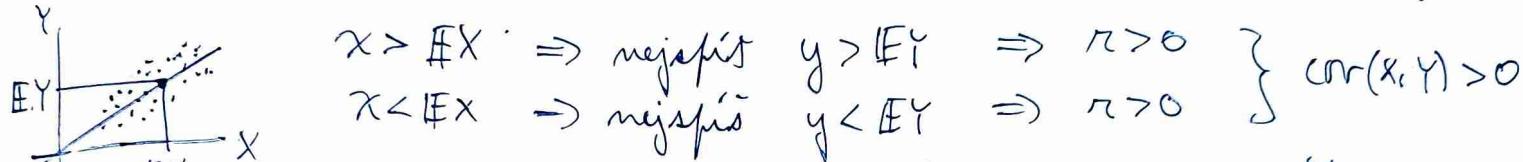
$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2 \cdot \text{cov}(X, Y)$$

Důkaz:

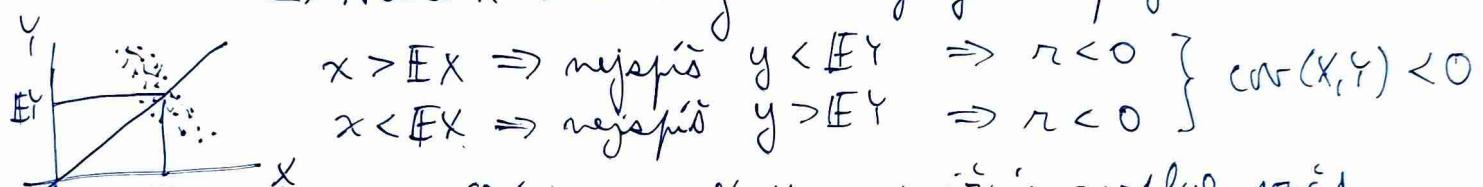
$$\begin{aligned} \text{var}(X+Y) &= E[(X+Y - E(X+Y))^2] = E[(X+Y - EX - EY)^2] = \\ &= E[X^2 - 2XEX + (EX)^2] + E[Y^2 - 2YEY + (EY)^2] + E[2(XY - XEY - YEY + EXEY)] \\ &= E[(X-EX)^2] + E[(Y-EY)^2] + 2 \cdot E[(X-EX)(Y-EY)] \end{aligned}$$

Důsledek: $X \perp Y \Rightarrow \text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$.

Instinkt: $\text{cov}(X, Y) = E[(X-EX)(Y-EY)]$... označme $r(x, y) := (x-EX)(y-EY)$



\Rightarrow větší $x \Rightarrow$ větší y ... rovník rozptyl součtu



\Rightarrow větší $x \Rightarrow$ malý y ... sníží rozptyl součtu

• Korelace vs Kauzalita

Def: Korelace m.v. X, Y je $S(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}}$

$-1 \leq S(X, Y) \leq 1$

Důkaz: Použijeme Cauchyho nerovnost, která říká, že $|EXY| \leq \sqrt{E[X^2]} \cdot \sqrt{E[Y^2]}$, kde dosadíme $X' := X - EX$ a $Y' := Y - EY$, z čehož $|\text{cov}(X, Y)| \leq \sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}$

! Korelace $\not\Rightarrow$ Kauzalita ... rádne říká, že $S(X, Y) = S(Y, X)$

\hookrightarrow malá korelace je často způsobena několika společnou přičinou

Spojité náhodné veličiny

Def: Náhodná veličina (současná) na f.f. (Ω, \mathcal{F}, P) je funkce $X: \Omega \rightarrow \mathbb{R}$ t.j. $\forall x \in \mathbb{R}: \{X \leq x\} \in \mathcal{F}$.

Def: Pro n.v. X definujeme distribuční funkci (cumulative dist. f. CDF) jako $F_X: \mathbb{R} \rightarrow [0, 1]$, $x \mapsto P[X \leq x]$

Věta: Nechť X je n.v. Pak

① F_X je nelesající

② $\lim_{x \rightarrow -\infty} F_X(x) = 0$

③ $\lim_{x \rightarrow \infty} F_X(x) = 1$

④ F_X je zprava spojita

Def
① $a < b \Rightarrow \{X \leq a\} \subseteq \{X \leq b\} \Rightarrow P[X \leq a] \leq P[X \leq b]$

② $A_m := \{X \leq m\} \dots A_1 \subseteq A_2 \subseteq \dots$

$$\hookrightarrow \lim_{m \rightarrow \infty} P(A_m) = P\left(\bigcup_{m \in \mathbb{N}} A_m\right) = P(\Omega) = 1$$

③ $B_m := \{X \leq -m\} \dots B_1 \supseteq B_2 \supseteq \dots$

$$\hookrightarrow \lim_{m \rightarrow \infty} P(B_m) = P\left(\bigcap_{m \in \mathbb{N}} B_m\right) = P(\emptyset) = 0$$

④ Nechť $x \in \mathbb{R}$. Cháme $\lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = F_X(x)$

$$C_n := \left\{X \leq x + \frac{1}{n}\right\} \Rightarrow \lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = \lim_{n \rightarrow \infty} P(C_n) = P(\bigcap C_n) = F_X(x)$$

Def: N.v. X je spojita $\equiv \exists f_X: \mathbb{R} \rightarrow \mathbb{R}$ t.j.

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \dots f_X \text{ je hustota } X$$

⊗ $\frac{d}{dx} F_X(x) = f_X(x) \Rightarrow F_X$ je spojita.

⊗ $f_X \geq 0 \dots \because F_X$ je nelesající

⊗ $\int_{-\infty}^{\infty} f_X(t) dt = 1 \quad \because \lim_{n \rightarrow \infty} F_X(n) = 1$

↗ pdf
probability density function

Intuice:

$$F_x(x) \approx \frac{\# \text{čísel} \leq x}{\# \text{počtu}} \dots \text{empirické CDF} = ECDF$$

$\Rightarrow F_x$ je něco jako limita ECDF

$\rightarrow f_x$ je něco jako limita histogramu sich počtu

Věta: Pro s.m.v. X platí

$$\textcircled{1} \quad \forall x \in \mathbb{R}: P(X=x) = 0$$

$$\textcircled{2} \quad P(a \leq X \leq b) = \int_a^b f_x(t) dt \quad \text{Lagrange}$$

$$\text{Dl: } \textcircled{1} \quad P(X=x) \leq P\left(x - \frac{1}{m} < X \leq x + \frac{1}{m}\right) = \int_{x-\frac{1}{m}}^{x+\frac{1}{m}} f(t) dt \xrightarrow{n \rightarrow \infty} \frac{2}{m} f(x) \rightarrow 0$$

\hookrightarrow je vidět pro merání f , pro neamerané detaily arguejme

$$\textcircled{2} \quad \int_a^b f_x(t) dt = F_x(b) - F_x(a) = P[a < X \leq b] = P[a \leq X \leq b] - P[X=a] \rightarrow 0$$

$$\text{Def: } \underline{\text{Sřední hodnota}} \text{ s.m.v. } X \text{ je } \mathbb{E}X := \int_{-\infty}^{\infty} x f_x(x) dx \quad (\text{ještě je obecně def.})$$

$$\text{Věta (PNS): } \mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

$$\text{Def: } \underline{\text{Var}}(X) := \mathbb{E}[(X-\mu)^2], \quad \mu := \mathbb{E}X$$

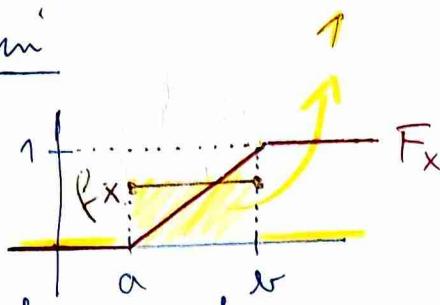
$$\hookrightarrow \text{výpočet: } \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x) dx$$

$$\text{Tvrz: } \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 \dots \text{důkaz linearitou, kterou zde jsem nechal}$$

Vášky spoj. rov dle m'

(1) Uniformní

$X \sim \text{Unif}(a, b)$



$$F_x(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

$$f_x(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{jinak} \end{cases}$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x dx + \int_b^{\infty} x dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{2} \left[\frac{x^2}{b-a} \right]_a^b = \underline{\underline{\frac{a+b}{2}}}$$

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\frac{a+b}{2}\right)^2 = \dots = \underline{\underline{\frac{(b-a)^2}{12}}}$$

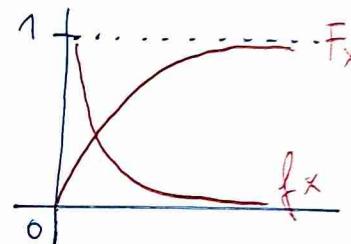
(2) Exponenciální ... obdoba geometrického

$$X \sim \text{Exp}(\lambda) \quad \hookrightarrow \lambda > 0 \quad F_x(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$\Rightarrow f'_x(x) = F' = \lambda e^{-\lambda x}$$

$$\mathbb{E}X = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda}$$

$$\text{Var } X = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} = \dots = \frac{1}{\lambda^2}$$



⊗ Něma 'paměť': $P[X > A+1 | X \geq A] = P[X > 1]$

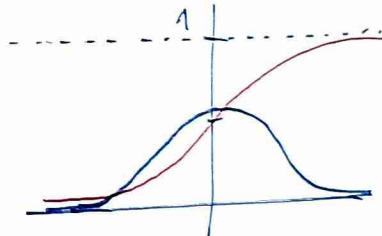
(3) Standardní normální

$$\text{ hustota: } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\text{ distribuce: } \Phi(x) = \int_{-\infty}^x \varphi(u) du$$

$$\mathbb{E}X = 0 \quad X \sim N(0, 1)$$

$$\hookrightarrow \text{Var } X = 1$$



↪ & $\varphi(x)$ se dá dýlit a přibližovat pomocí Stirlingova

(4) Obecné normální

$$Z \sim N(0, 1)$$

$$X \sim N(\mu, \sigma^2) \quad \equiv \quad X = \mu + \sigma Z$$

$$\hookrightarrow \text{shift} = \text{střed}$$

$$\hookrightarrow \text{rozložení}$$

$$f_x(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$$

$$F_x(x) = \int_{-\infty}^x \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(u) du = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\mathbb{E}X = \mu$$

$$\text{Var } X = \sigma^2$$

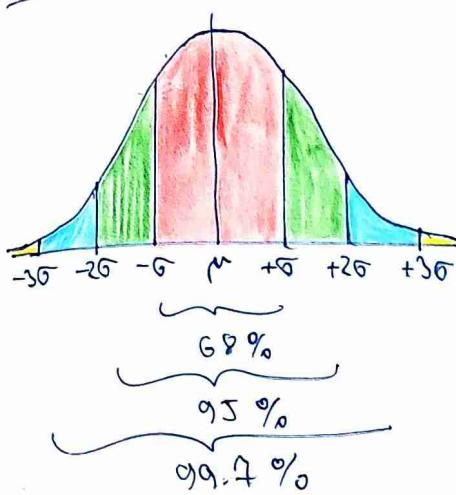
Věta (odolnost noci součtu): X_1, \dots, X_m jsou nez. n. r. $X_i \sim N(\mu_i, \sigma_i^2)$

$$\Rightarrow X_1 + \dots + X_m \sim N(\mu, \sigma^2), \quad \mu = \mu_1 + \dots + \mu_m$$

$$\sigma^2 = \sigma_1^2 + \dots + \sigma_m^2$$

Bruvídlo 36

68-95-99.7 rule



$$P(\mu - \sigma < X < \mu + \sigma) = \Phi(1) - \Phi(-1) \approx 0.68$$

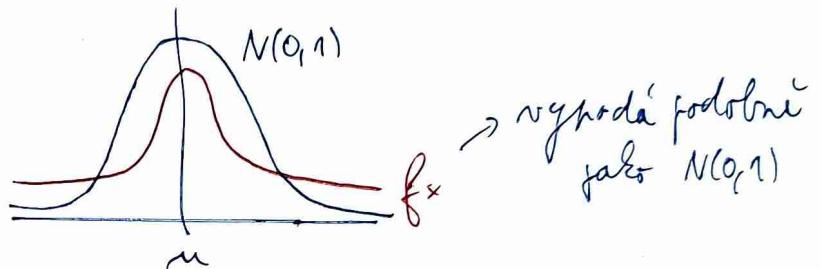
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx 0.95$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = \Phi(3) - \Phi(-3) \approx 0.997$$

⑤ Cauchyho rozdělení

$$f_x(x) = \frac{1}{\pi(1+x^2)}$$

$$F_x(x) = \int_{-\infty}^x f_x = \frac{1}{\pi} \arctg x + \frac{1}{2}$$



! střední hodnota neexistuje

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_x(x) dx = \int_{-\infty}^0 x f_x + \int_0^{\infty} x f_x = \dots = \infty - \infty$$

$$\hookrightarrow \int_0^{\infty} x f_x = \frac{1}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \left| \begin{matrix} u = 1+x^2 \\ du = 2x dx \end{matrix} \right| = \frac{1}{2\pi} \int_1^{\infty} \frac{du}{u} = \frac{1}{2\pi} \ln|u| \Big|_1^{\infty} = \infty$$

→ řečená je dobré?



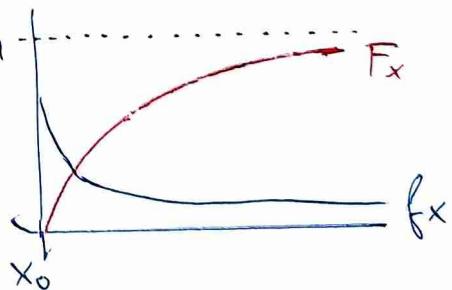
↑ Cauchyho rozdělení

⑥ Paretoho rozdělení

$$F_x(x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^{\alpha}, & x \geq x_0 \\ 0, & x < x_0 \end{cases}$$

$$\alpha, x_0 > 0$$

$$f_x(x) = \begin{cases} \frac{\alpha}{x} \cdot \left(\frac{x_0}{x}\right)^{\alpha}, & x \geq x_0 \\ 0, & x < x_0 \end{cases}$$



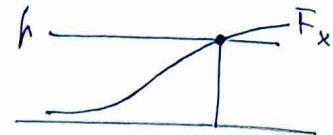
∅ F_x rule & 1 mohou formuleji mít exp \Rightarrow long tail = dlouhý chod

$$\mathbb{E}X = x_0 \frac{\alpha}{\alpha-1}$$

Kvantilová funkcia

Def: Pre m.r. X definujeme kvantilovú funkciu $Q_X: [0,1] \rightarrow \mathbb{R}$ jazec

$$Q_X(p) := \min \{x \in \mathbb{R} \mid p \leq F_X(x)\}$$



Pre súčasťou X je $Q_X(p) = F_X^{-1}(p)$

Pre $\forall x \in \mathbb{R}$: $Q_X(p) \leq x \Leftrightarrow p \leq F_X(x) = P[X \leq x]$

Def: Median je hodnota $m := Q_X(\frac{1}{2})$

Pre súčasťou m.r. X je nejmenejšie také číslo, že

$$P[X \leq m] \geq \frac{1}{2} \text{ a } P[X > m] \leq \frac{1}{2}$$

→ pre d.m.r. môže byť takýchto hodnôt nie

Def: Prvý kvantil je hodnota $q := Q_X(\frac{1}{4})$... 1. čtvrtina hodnôt $\leq q$

Desiaty percentil je $Q_X(10/100)$... hodnota > nie 10% hodnôt

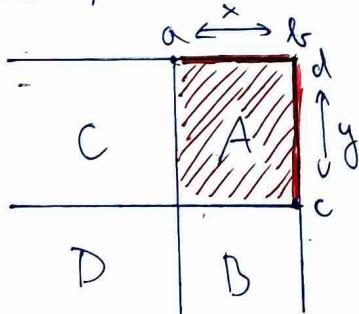
Náhodné vektor

joint cdf

Def: Pre m.r. X, Y ma f.f. (\mathcal{L}, F, P) definujeme súvisiaca distribučná funkcia

$$F_{X,Y}: \mathbb{R}^2 \rightarrow [0,1], \quad F(x,y) = P[X \leq x \text{ a } Y \leq y]$$

Véľkosť (pre obdĺžnik):



$$\begin{aligned} P[(X,Y) \in A] &= P[X \in (a,b) \text{ a } Y \in (c,d)] \\ &= F(b,d) - F(a,d) - F(b,c) + F(a,c) \\ F &:= F_{X,Y} \end{aligned}$$

Pre hustotu:

$$P[(X,Y) \in S] = \int_S f_{X,Y}(x,y) dx dy$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$$

Marginalná hustota

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$

Věta (PNS): Pro sítědlní hodnoty funkce dvoj. m.v. platí

$$\mathbb{E}[g(x, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{x,y}(x, y) dx dy$$

Důsledek (linearity of E) $\mathbb{E}[aX + bY + c] = a\mathbb{E}X + b\mathbb{E}Y + c$

Nerávnost

Def: Náhodné vel. X, Y jsou nerávise, $X \perp Y \Leftrightarrow$

$$\forall x, y \in \mathbb{R} : \{X \leq x\} \perp \{Y \leq y\}$$

$$\Leftrightarrow F_{x,y}(x, y) = F_x(x) \cdot F_y(y).$$

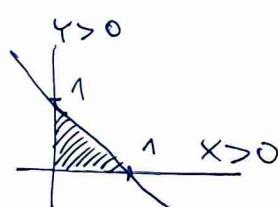
$$\text{Věta: } X \perp Y \Leftrightarrow f_{x,y}(x, y) = f_x(x) \cdot f_y(y).$$

Věta (konvoluce): Nechť X, Y jsou nerávise spojité m.v. Pak $Z := X + Y$ je také s.m.v. a platí

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx.$$

Příklady:

$$\textcircled{1} \quad f_{x,y}(x, y) = \begin{cases} e^{-x-y}, & x, y > 0 \\ 0, & \text{jinak} \end{cases} \quad \rightarrow P(X+Y \leq 1) = ?$$



$$P(X+Y \leq 1) = \underbrace{\iint_A f_{x,y}(x, y) dx dy}_{A} = \int_0^1 \int_0^{1-x} e^{-x-y} dy dx = \dots = \frac{e-2}{e}$$

$$\textcircled{2} \quad X, Y \sim N(0, 1) \Rightarrow Z := X + Y \sim ?$$

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2 - \frac{z^2}{2} + xz} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x-\frac{z}{2})^2} e^{-\frac{z^2}{4}} dx = \frac{e^{-\frac{z^2}{4}}}{2\pi} \int_{-\infty}^{\infty} e^{-(x-\frac{z}{2})^2} dx = \\ &= \frac{e^{-\frac{z^2}{4}}}{2\pi} \int_{-\infty}^{\infty} e^{-u^2} du = \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}} \Rightarrow \sigma = 2 \quad \Rightarrow Z \sim N(0, 2) \end{aligned}$$

$$X \sim N(0, \sigma^2) \dots f_x = \frac{1}{\sigma} \varphi\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x}{\sigma}\right)^2}$$

Kvádrozmírné std. norm. rozdělení

$$z_1, \dots, z_m \rightarrow z_i \sim N(0, 1) \Rightarrow f_{z_i}(t) = \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

$$\hookrightarrow z := (z_1, \dots, z_m) \quad \text{meravise} \Rightarrow f_{x,y} = f_x \cdot f_y$$

$$f_z(z_1, \dots, z_m) = \prod_{i=1}^m f_{z_i}(t_i) = (\sqrt{2\pi})^m e^{-\frac{t_1^2 + \dots + t_m^2}{2}} = \underbrace{(\sqrt{2\pi})^m}_{\text{faktor růžerí pouze na } \|z\|} \cdot e^{-\frac{\|z\|^2}{2}}$$

$\Rightarrow f_z$ rážerí pouze na $\|z\|$, tedy rozdílnosti od počátku
 \hookrightarrow je to radiálně symetrická funkce

\rightarrow pokud $n=2$ (2 dimenze), tak f_z je Bell-curve

\Rightarrow díky této symetrii dává $X := \frac{z}{\|z\|}$
 unif. náhodný bod na jednotkové sféře $\mathbb{S}^n \subset \mathbb{R}^n$

Podmínkování

Def: X je m.v. na (Ω, \mathcal{F}, P) , $B \in \mathcal{F}$ a $P(B) > 0$. Rešíme

$$F_{x|B}(x) := P[X \leq x | B]$$

\rightarrow k tomu patří $f_{x|B}$.

Véta (o rozdělení hustoty): X je s.m.v. a B_1, B_2, \dots je rozdělen \mathcal{F} . Platí

$$F_x(x) = \sum_i P(B_i) \cdot F_{x|B_i}(x)$$

$$f_x(x) = \sum_i P(B_i) \cdot f_{x|B_i}(x)$$

Druhá véta: F_x ... véta o výplné hustoty pro $A = \{X \leq x\}$
 f_x ... rozdělení F_x

$$\text{Def: } \mathbb{E}[X|B] := \int_{-\infty}^{\infty} x f_{x|B}(x) dx$$

Véta: Pokud je B_1, B_2, \dots rozdělen \mathcal{F} , tak

$$\mathbb{E} X = \sum_i P(B_i) \mathbb{E}[X|B_i]$$

Nerovnosti

Příklad: Může být 99 % lidí starších než průměr?

↪ Ano ... 1 člověk ... 1 rok
99 lidí ... 20 let } průměr < 20

Může být 51 % lidí starších než 2x průměr? → NE!

Věta (Markovova ner.): Nechť n.r. X splňuje $X \geq 0$, $a \in \mathbb{R}$. Pak

$$P[X \geq a] \leq \frac{\mathbb{E} X}{a}$$

Důkaz: $P[X \geq b \cdot \mathbb{E} X] \leq \frac{1}{b}$... náleží $a := b \cdot \mathbb{E} X$

$$\text{Dk: } \mathbb{E} X = P[X \geq a] \cdot \mathbb{E}[X | X \geq a] + P[X < a] \cdot \mathbb{E}[X | X < a]$$

$$\geq P[X \geq a] \cdot \mathbb{E}[X | X \geq a] \geq P[X \geq a] \cdot a$$

■

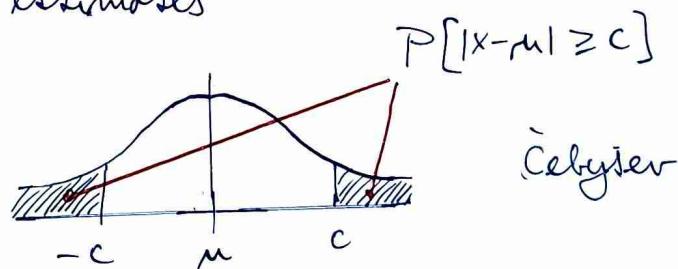
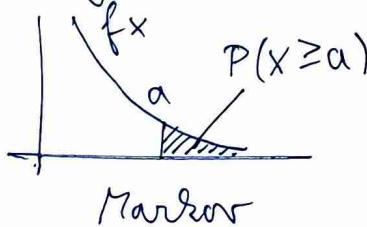
Věta (Čebyshevova ner.): Nechť X má $\mu := \mathbb{E} X$ a $\sigma^2 := \text{Var}(X)$, $\lambda > 0$. Pak

$$P[|X - \mu| \geq \lambda \cdot \sigma] \leq \frac{1}{\lambda^2}$$

$$\text{Dk: } |X - \mu| \geq (\lambda \cdot \sigma) \Leftrightarrow (X - \mu)^2 \geq \lambda^2 \sigma^2 \Rightarrow Y := (X - \mu)^2 \Rightarrow \mathbb{E} Y = \sigma^2$$

$$\Rightarrow P[\dots] = P[Y \geq \lambda^2 \sigma^2] = P[Y \geq \lambda^2 \mathbb{E} Y] \leq \frac{1}{\lambda^2} \quad \dots \text{Markov.} \quad \blacksquare$$

Odhad červotku = tail estimates



Věta (Chernoffova ner.): Nechť $X_i = \begin{cases} +1, & \text{pravd. } 50\% \\ -1, & \text{pravd. } 50\% \end{cases}$. $X := \sum_{i=1}^m X_i$, $\lambda > 0$

$$P[X \leq -\lambda \cdot \sigma_x] = P[X \geq \lambda \cdot \sigma_x] \leq e^{-\frac{\lambda^2}{2}}, \quad \sigma_x^2 = m \Rightarrow \sigma_x = \sqrt{m}$$

↪ Čebyshev je kvadratický odhad, Kohle je exponenciální

Základy velkých čísel

Def: Výberový průměr (sample mean) m.v. X_1, \dots, X_m je $\bar{X}_m := \frac{1}{m} \sum_i X_i$.

Věta (silný ZVC): Nechť X_1, X_2, \dots jsou stejně rozdělené meravisci m.v., se $\mathbb{E}X_i = \mu$ a $\text{Var}X_i = \sigma^2$. Pak platí

$$P\left[\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right] = 1 \quad \dots \quad \lim_{n \rightarrow \infty} \bar{X}_n = \mu \quad \text{stevo jisté}$$

Aplikace: Monte-Carlo integrací

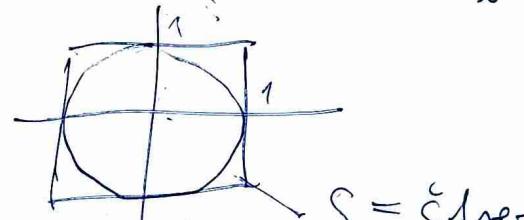
↳ chci sčítat $I := \int_S g(\bar{x}) d\bar{x}$, $S \subseteq \mathbb{R}^d$

↳ objem S ... $V := \text{vol}(S) = \int_S d\bar{x}$

$\Rightarrow \bar{X}_1, \dots, \bar{X}_m$ jsou samplify $\in S$ $\Rightarrow I \approx Q_m = V \cdot \frac{1}{m} \sum_i f(\bar{x}_i)$

Př: Odhad obsahu kruhu

$$g(x, y) := \begin{cases} 1, & x^2 + y^2 \leq 1 \\ 0, & \text{jinak} \end{cases}$$



$S = \text{čtverec}$

$$\Rightarrow V = \pi, \quad \frac{1}{m} \sum_i f(\bar{x}_i) = \frac{\# \text{ bodů v kruhu}}{\# \text{ bodů v kruhu}}$$

Věta (slabý ZVC): Nechť X_1, X_2, \dots jsou stejně rozdělené meravisci m.v.

se $\mathbb{E}X_i = \mu$ a $\text{Var}X_i = \sigma^2$. Pak pro $\forall \varepsilon > 0$ platí

$$\lim_{n \rightarrow \infty} P[|\bar{X}_n - \mu| > \varepsilon] = 0, \quad \text{pisemne } \bar{X}_n \xrightarrow{P} \mu.$$

\Rightarrow Říkáme, že početnost \bar{X}_n konverguje k μ v pravděpodobnosti.

Dk:

$$\odot X_1, \dots, X_m \text{ stejné r.} \Rightarrow \mathbb{E}\bar{X}_m = \frac{1}{m}(\mathbb{E}X_1 + \dots + \mathbb{E}X_m) = \mu$$

$$\hookrightarrow \text{pro Var: } \text{Var}(\bar{X}_m) = \frac{1}{m^2} \text{Var}(\sum_i X_i) \xrightarrow{\text{mer.}} \frac{1}{m^2} (\text{Var}X_1 + \dots + \text{Var}X_m) = \frac{\sigma^2}{m}$$

$$\Rightarrow \text{chceme použít čebysjevova n. } P[|\bar{X}_m - \mu| > 1\sigma] \leq \frac{1}{n^2}$$

$$\hookrightarrow \sqrt{\text{Var}(\bar{X}_m)} = \frac{\sigma}{\sqrt{m}} \Rightarrow A := \frac{\varepsilon \sqrt{m}}{\sigma}$$

$$\Rightarrow P[|\bar{X}_m - \mu| > 1 \cdot \frac{\sigma}{\sqrt{m}}] = P[|\bar{X}_m - \mu| > \varepsilon] \leq \frac{\sigma^2}{m \varepsilon^2} \rightarrow 0. \quad \blacksquare$$

Centrální limitní věta

Věta: Nechť X_1, X_2, \dots jsou stejně rozdílené měřiví m.r.

se $\mathbb{E}X_i = \mu$ a $\text{Var } X_i = \sigma^2$. Označme

$$Y_m := \frac{\bar{X}_m - \mu}{\sigma/\sqrt{m}} = \frac{X_1 + \dots + X_m - m\mu}{\sqrt{m}\sigma}, \quad F_m := \text{distr. fce } Y_m$$

Potom $\forall x \in \mathbb{R}$: $\lim_{m \rightarrow \infty} F_m(x) = \Phi(x)$, píšeme $Y_m \xrightarrow{d} N(0,1)$ a říkáme, že početnost Y_m konverguje k $N(0,1)$ v distribuci.

Intuice:

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}Y_m &= m \cdot \frac{\mathbb{E}\bar{X}_m - \mu}{\sqrt{m}\sigma} = m \cdot \frac{0}{\sigma} = 0 \\ \text{Var } Y_m &= \left(\frac{\sigma}{\sqrt{m}}\right)^2 \cdot \text{Var } \bar{X}_m = \frac{m}{m} \cdot \frac{\sigma^2}{m} = 1 \end{aligned} \quad \left. \begin{array}{l} \text{ } \\ \text{ } \end{array} \right\} Y_m, \text{"má řádiči"} \text{ bývá } N(0,1)$$

\textcircled{2} pokud $X_i \sim N(\mu, \sigma^2)$, tak (odlhost něj součtu) bude Y_m být normální.

\hookrightarrow díky \textcircled{1} bude $Y_m \sim N(0,1)$

\Rightarrow pokud (LV platí pro nejdříve rozdílení), tak jedině $N(0,1)$.

Příklad: Výsah má max. možnost 1000 kg.

\hookrightarrow město má 13 osob, $\mathbb{E}X_m = 75 = \mu$, $\sigma_x = 10$
 \hookrightarrow hmotnosti X_1, \dots, X_{13}

$$\Rightarrow P(X_1 + \dots + X_{13} \geq 1000) = ? \quad \Rightarrow S_{13} := X_1 + \dots + X_{13}$$

$$\hookrightarrow Y_{13} := \frac{S_{13} - 13 \cdot 75}{10\sqrt{13}} \dots S_B > 1000 \Leftrightarrow Y_m > \frac{1000 - 13 \cdot 75}{10\sqrt{13}} \approx 0.69$$

$$(\text{LV: } Y_{13} \approx N(0,1)) \Rightarrow P(S_B > 1000) \approx P(Y_m > 0.7) = 1 - P(Y_m \leq 0.7) = 1 - \Phi(0.7)$$

STATISTIKA

1) Explorační analýza ... pohledné rozborování nějakých dat

2) Konfirmací analýza ... matematická disciplína

↳ snažíme se z dat zjistit nějaké sloučenosti a odhalit nepravidly spisované náhodným kolísáním dat

→ ráckod: jak vybírat z danej populace, objekty, co budeme měřit?

a) uniformě náhodně bez vracení

b) uniformě náhodně s vracením → lepe se analyzuje

Def: Posloupnost ner. m.v. X_1, \dots, X_n se stejnou distribuční funkcí F je náhodný výber s rozsahem n . Píšeme $X_1, \dots, X_n \sim F$.

↳ naměřená data x_1, \dots, x_n = realizace, $x_i = X(\omega)$.

↳ na ráckodě realizace se snažíme určit F = model

① Neparametrický model ... nic o F nepředpokládáme
↳ dosl. dobré nejde

② Parametrický model ... F je z nějaké možnosti $\{F_\theta | \theta \in \Theta\}$

$\text{Exp}(\lambda) \dots \theta = \lambda : \quad \Theta = \mathbb{R}^+$

$U(a, b) \dots \theta = (a, b) : \quad \Theta = \mathbb{R}^2$

$U(0, b) \dots \theta = b : \quad \Theta = \mathbb{R}$

$N(\mu, \sigma^2) \dots \theta = (\mu, \sigma) : \quad \Theta = \mathbb{R} \times \mathbb{R}^+$

možné světy

Def: Statistika je libovolná fce náhodného výberu $T = T(x_1, \dots, x_n)$.

↳ kde to je m.v., ale svoji náhodnost čerpá pouze z x_1, \dots, x_n

⇒ výberový průměr, medián, maximum, ...

Příklad: x_1, \dots, x_n doby běhu programu $\Rightarrow \hat{\mu} = P[X > 100 \text{ ms}] = ?$

1) $\hat{\mu} = \frac{\#\{i : x_i > 100 \text{ ms}\}}{n} \dots x_1, \dots, x_n$ = naměřená data

2) $X \sim N(\mu, \sigma^2) \Rightarrow P(X > 100) = 1 - P(X \leq 100) = 1 - \Phi\left(\frac{100 - \mu}{\sigma}\right)$

↳ normální $\mu, \sigma \Rightarrow$ odhad

$\hat{\mu} = \bar{x}_n \quad \hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \dots$ výberový rozptyl

Bodové odhady

$\hat{\theta}$... odhad parametru θ → estimator

↪ náhodná v. závislá na měřených dotech ⇒ stohistika

Def: Bias (rychýlem) odhadu $\hat{\theta}$ je $E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$

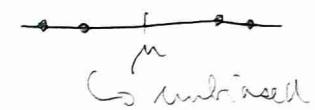
Def: Pro náhodný výber $X_1, \dots, X_m \sim F_\theta$ je bodový odhad $\hat{\theta}_m$

1) nestraný $\equiv E\hat{\theta}_m = \theta$... bias = 0

2) asymptoticky nestraný $\equiv \lim_{m \rightarrow \infty} E\hat{\theta}_m = \theta$... bias → 0

3) konsistentní $\equiv \hat{\theta}_m \xrightarrow{P} \theta$, neboť

$\forall \epsilon > 0 : P[|\hat{\theta}_m - \theta| > \epsilon] \rightarrow 0$. ⇒ malý rozptyl pokud bude:



⊗ ZVČ říká, že \bar{X}_m je konsistentní odhad $E\bar{X}$

Def: Střední kvadratiká odchylka $MSE(\hat{\theta}) := E[(\hat{\theta} - \theta)^2]$.

Věta: $MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2$

Důk: $\text{Var}(\hat{\theta}) = \text{Var}(\hat{\theta} - \theta) = [E[(\hat{\theta} - \theta)^2] - E[(\hat{\theta} - \theta)]^2] = MSE(\hat{\theta}) - \text{bias}(\hat{\theta})^2$

Věta: Mějme náhodný výber X_1, \dots, X_m . Potom

① $\bar{X}_m := \frac{1}{m} \sum_i X_i$... nestraný & konsistentní odhad μ

↪ x. mean() → numpy

② $\bar{S}_m^2 := \frac{1}{m} \sum_i (X_i - \bar{X}_m)^2$... asym. nestr. & kons. odhad σ^2

↪ x. var(ddof=0)

③ $\hat{S}_m^2 := \frac{1}{m-1} \sum_i (X_i - \bar{X}_m)^2$... nestraný & kons. odhad σ^2

↪ x. var(ddof=1) výberový rozptyl

Důk: ① $E\bar{X}_m = \mu$ z linearity, konsistence ze ZVČ

② } konsistence významné, nestranost \bar{S}_m^2 je na další straně

③ } → proč se tam sahá délkou $m-1$? ↗

Dr: Kolik je $E\left[\sum_i (X_i - \bar{X}_m)^2\right]$?

$$\sum_i^m \bar{X}_m$$

$$E[\dots] = E\sum_i (X_i^2 - 2X_i\bar{X}_m + \bar{X}_m^2) = E\left[\sum_i X_i^2 - 2\bar{X}_m \sum_i X_i + m\bar{X}_m^2\right]$$

\rightarrow myšlím, že $X_1, \dots, X_m \sim F_\theta$... stejná dist.

$$= m \cdot E[X_i^2] - m \cdot E[\bar{X}_m]^2$$

$$\rightarrow$$
 myšlím $\text{Var}X = E[X^2] - (EX)^2 \Rightarrow E[X^2] = \text{Var}X + (EX)^2 = \sigma^2 + \mu^2$

$$= m(\sigma^2 + \mu^2) - m\left(\frac{\sigma^2}{m} + \mu^2\right) = (m-1)\sigma^2$$

nerovnost

$$\text{Var}X_m = \frac{1}{m^2} \text{Var}(X_1 + \dots + X_m) = \frac{1}{m^2} m \cdot \sigma^2 = \frac{\sigma^2}{m}$$

$$\Rightarrow E[\hat{S}_m] = \sigma^2 \Rightarrow \hat{S}_m \text{ je nestr.}$$



• Metoda momentů

r -tý moment X ... $m_r := E[X^r]$

r -tý vyčísl. m. X ... $\hat{m}_r := \frac{1}{m} \sum_i X_i^r$ $\underset{||}{E[X]}$ $\underset{||}{\bar{X}_m}$

\Rightarrow pro odhad θ myslíme rovnici $m_r(\theta) = \hat{m}_r(\theta)$

\hookrightarrow případně soustavnou rovnicí $m_r(\theta) = \hat{m}_r(\theta)$ pro $r=1, 2, \dots$

Pří: ① $X_i \sim Ber(p)$ $\rightarrow \theta = p$

$$\begin{aligned} m_1 &= E[X_i] = p \\ \hat{m}_1 &= \bar{X}_m \end{aligned} \quad \left. \begin{array}{l} \hat{p} = \bar{X}_m \\ \dots \text{diky zVČ to je nestr. konz. odhad} \end{array} \right.$$

② $X_1, \dots, X_m \sim N(\mu, \sigma^2) \rightarrow$ dva parametry $\Rightarrow 2$ rovnice

$$m_1 = \mu$$

$$m_2 = E[X^2] = \text{var}(X) + (EX)^2 = \sigma^2 + \mu^2$$

$$\Rightarrow \hat{m}_1 = \bar{X}_m = \mu$$

$$\hat{m}_2 = \frac{1}{m} \sum_i X_i^2 = \mu^2 + \sigma^2 = \bar{X}_m^2 + \sigma^2 \quad \left. \begin{array}{l} \hat{\mu} = \bar{X}_m \\ \hat{\sigma}^2 = \hat{m}_2 - \hat{\mu}^2 \end{array} \right.$$

• Metoda max. verojodnosti - max. likelihood

→ náhodný výber $X_1, \dots, X_m \sim F_\theta$ a realizace x_1, \dots, x_m

↳ předpokládáme $\sim F_\theta \Rightarrow$ to dává distribuci

① diskrétní m.v.: Chceme max. $P[X_1=x_1 \wedge X_2=x_2 \wedge \dots \wedge X_m=x_m]$

↳ X_1, \dots, X_m nerávisek $\Rightarrow P[\forall i : X_i = x_i] = \prod P[X_i = x_i]$

$$\Rightarrow L(\theta, \vec{x}) := \prod_i P[X_i = x_i; \theta]$$

↳ nevětě, ale $X_i \sim F_\theta$

② svojisté m.v.: $L(\theta, \vec{x}) := \prod_i f_{X_i|\theta}(x_i)$

\Rightarrow hledám max $L(\theta, \vec{x})$, → derivací podle θ

$$\text{respektive } l(\theta, \vec{x}) := \log L(\theta, \vec{x}) = \sum_i \log(f_{X_i}(\vec{x}))$$

Příklady

① $X_1, \dots, X_m \sim \text{Ber}(\mu) \Rightarrow \theta = \mu$

$$P(\vec{x}=\vec{x}) = L(\mu, \vec{x}) = P[X_1=x_1, \dots, X_m=x_m] = \mu^{\sum x_i} (1-\mu)^{m-\sum x_i}$$

$$\Rightarrow l(\mu, \vec{x}) = \sum x_i \ln(\mu) + (m-\sum x_i) \ln(1-\mu)$$

$$\Rightarrow \frac{\partial l}{\partial \mu}(\mu, \vec{x}) = \frac{\sum x_i}{\mu} + \frac{m-\sum x_i}{1-\mu} = 0 \underset{\approx \max}{\Downarrow} \Rightarrow \hat{\mu} = \frac{\sum x_i}{m}$$

② $X_1, \dots, X_m \sim U(0, \theta) \Rightarrow f(x) = \begin{cases} \frac{1}{\theta}, & x \in [0, \theta] \\ 0, & \text{jinde} \end{cases}$

$L(\theta, \vec{x}) = \prod_i f(x_i) \dots$ aby to nebyla 0, tak $\theta \geq \max_i x_i$

$$= \prod_i \frac{1}{\theta} = \frac{1}{\theta^m} \dots \max \Rightarrow \theta \text{ co nejméně} \Rightarrow \hat{\theta} := \max_i x_i$$

③ $X_1, \dots, X_m \sim \text{Geom}(\mu) \Rightarrow P[X_i = x_i] = (1-\mu)^{x_i-1} \cdot \mu$

$$L(\mu, \vec{x}) = \prod_i P[X_i = x_i] = \mu^m (1-\mu)^{(\sum x_i) - m}$$

$$\Rightarrow l(\mu, \vec{x}) = m \log(\mu) + (\sum x_i - m) \log(1-\mu)$$

$$\rightarrow \text{přesloužit}: \log(\mu) + (\sum x_i - m) \log(1-\mu)$$

$$\rightarrow \text{derivace}: \frac{1}{\mu} + \frac{1-\sum x_i}{1-\mu} = 0 \Rightarrow \hat{\mu} = \frac{1}{\sum x_i}$$

• Intervalové odhady

→ pries bodový odhad odkadmu θ falo $\hat{\theta}$... ale typicky $\hat{\theta} \neq \theta$
 ⇒ chci si byť dost jistý, že θ nemá moc mino

Def: Nechť $D \subseteq H$ sú sústavy. Ríkame, že určují $[D, H]$
intervalový odhad o spôsobnosťi $(1-\alpha) \equiv P[D \leq \theta \leq H] = 1-\alpha$.

↳ umožňuje chybu $\Delta \in (0, 1)$

↳ anglicky $(1-\alpha)$ -confidence interval $\Rightarrow (1-\alpha)$ -CI

↳ niekedy dáva smysl uvozovať jednostranné odhady ($D = -\infty$ v $H = \infty$)

Značení: Často budeme chcieť CI typu $X \pm \delta := [X - \delta, X + \delta]$.

Věta: Máme nestranný bodový odhad $\hat{\theta}$ pre parameter θ .

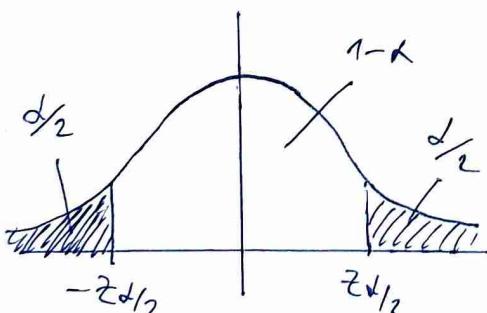
Pokud je $\hat{\theta} \sim N(\mu, \sigma^2)$, tak

$$\hat{\theta} \pm z_{\alpha/2} \sigma(\hat{\theta}) \text{ je } (1-\alpha)\text{-CI}, \quad z_{\alpha/2} := \Phi(1 - \frac{\alpha}{2}).$$

Pokud nemá normálne rozdelenie, tak rečeme, že súčas je (CLV)

Dr: Nejdov povedeme standardizaci, $E\hat{\theta} = \theta$ akoby nestrannosti

$$Z = \text{stand}(\hat{\theta}) = \frac{\hat{\theta} - E\hat{\theta}}{\sigma(\hat{\theta})} = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} \sim N(0, 1)$$



$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$$

$$P(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} \leq z_{\alpha/2}) = 1-\alpha$$

$$P(\underbrace{\hat{\theta} - z_{\alpha/2} \sigma}_{D} \leq \theta \leq \underbrace{\hat{\theta} + z_{\alpha/2} \sigma}_{H}) = 1-\alpha$$



Příklady:

$$\textcircled{1} \quad X_1, \dots, X_n \sim N(\mu, \sigma^2) \quad \rightarrow \hat{\theta} = \bar{X}_n$$

↳ máme nestraný odhad $\hat{\theta} = \bar{X}_n \sim N(\text{?}, \text{?})$

$$\hookrightarrow \mathbb{E} X_n = \mu, \quad \text{Var}(X_n) = \frac{\sigma^2}{n} \Rightarrow \hat{\theta} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow \bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{je } (1-\alpha)\text{-CI.}$$

$$\textcircled{2} \quad X_1, \dots, X_n \text{ libovolné rozdilem'}$$

↳ máme } hledáme $\theta := \mathbb{E} X \Rightarrow$ opět odhad \bar{X}_n pro θ
 n velké }

→ podle CLV je \bar{X}_n přibližně normálně rozdileno ... $\frac{\bar{X}_n - \mu}{\sigma(\bar{X}_n)} \approx N(0,1)$

$$\Rightarrow \bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \text{ je přibližně } (1-\alpha)\text{-CI}$$

$$\hookrightarrow \lim_{n \rightarrow \infty} P(D \leq \theta \leq H) = 1-\alpha \quad \dots \quad \alpha \approx 30 \quad \begin{matrix} \nearrow \\ \text{fazetí} \end{matrix}$$

→ tedy je pro nějaký rozdíl lepší použít nov. Student- t rozdilem'

Pozn: Najdeme CI k rozdiu $X \pm \Delta$.

↳ jaké m nám slouží odhad rozdílu pro dané Δ ?

$$\Delta = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n \geq \left(\frac{z_{\alpha/2} \sigma}{\Delta} \right)^2 = \frac{\text{const}}{\Delta^2}$$

⇒ pokud chci Δ zmenšit 10x, a zachovat spolehlivost α ,
 tak potřebuji 100x víc měření (n)

Dvourůzbový test

$$\left. \begin{array}{l} X_1, \dots, X_n \sim N(\mu_X, \sigma_X^2) \\ Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2) \end{array} \right\} \text{ máme } \sigma_X, \sigma_Y, \text{ chceme odhad pro} \\ \theta := \mu_X - \mu_Y \Rightarrow \hat{\theta} = \bar{X}_n - \bar{Y}_m$$

→ máme nestraný odhad pro $\hat{\theta}$, protože X_i, Y_i jsou norm., takže $\hat{\theta}$ také

→ protože X_i, Y_j jsou všechny nezávislé, tak

$$\text{Var}(\hat{\theta}) = \text{Var}(\bar{X}_n - \bar{Y}_m) = \text{Var}(\bar{X}_n) + \text{Var}(\bar{Y}_m) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$$

$$\Rightarrow \sigma(\hat{\theta}) = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

$$\Rightarrow \bar{X}_n - \bar{Y}_m \pm z_{\alpha/2} \cdot \sigma(\hat{\theta}) \quad \text{je } (1-\alpha)\text{-CI.}$$

• Intervalové odhody, když měřením rozptyl

① Plug-in estimate

$$X_1, \dots, X_m \sim \text{Ber}(\mu) \Rightarrow \theta = \mu, \text{ ale měřením } \bar{x}$$

$$\Rightarrow \sigma(\hat{\mu}) = ? \dots \text{Var } X_i = \mu \cdot (1-\mu) \quad \because X_i \sim \text{Ber}(\mu)$$

$$\hookrightarrow \hat{\mu} = \bar{x}_m \Rightarrow \text{Var } \bar{x}_m = \frac{1}{m} \cdot \mu \cdot (1-\mu)$$

$$\Rightarrow \sigma(\hat{\mu}) = \sqrt{\frac{\mu \cdot (1-\mu)}{m}} \approx \sqrt{\frac{\hat{\mu} \cdot (1-\hat{\mu})}{m}} \dots \text{faktor měření } \mu$$

$$\Rightarrow \bar{x}_m \pm z_{\alpha/2} \sqrt{\frac{\hat{\mu} \cdot (1-\hat{\mu})}{m}} \quad \text{je } (1-\alpha)-CI$$

② Studentův test

$$X_1, \dots, X_m \sim N(\theta, \sigma^2) \dots \text{G měřením}$$

$$\hookrightarrow nestranný odhad \hat{\theta} = \bar{x}_m, \quad \sigma(\hat{\theta}) = \frac{\sigma}{\sqrt{m}}, \text{ ale G měřením}$$

$$\Rightarrow \text{odhad } \hat{\sigma} = \hat{s}_m = \sqrt{\frac{1}{m-1} \sum_i (x_i - \bar{x}_m)^2} \Rightarrow \sigma(\hat{\sigma}) = \frac{\hat{s}_m}{\sqrt{m}}$$

! → důkaz věty dělán normální $\hat{\theta}$ na stand. normální r.

$$\Rightarrow T_m := \frac{\hat{\theta} - \mathbb{E}(\hat{\theta})}{\sigma(\hat{\theta})} = \frac{\bar{x}_m - \mu}{\hat{s}_m / \sqrt{m}} \rightarrow \text{tahle není konst, ale n. veličina}$$

$\Rightarrow T_m$ nemá $N(0, 1)$, ale něco podobného

Def: Students t-distribution s $m-1$ stupni volnosti je písni t_{m-1} rozdilem.

→ distribuční funkci nazíváme $\Psi_{m-1}(x)$.

↪ python: scipy.stats.t.cdf(x, m-1) ↗

⇒ použijeme odhad

$$\hat{\theta} \pm t_{\alpha/2} \cdot \frac{\hat{s}_m}{\sqrt{m}}, \text{ kde } t_{\alpha/2} := \Psi_{m-1}^{-1}(1-\alpha/2)$$

→ analogie pro $\hat{\theta}^{-1}$

! $t_{\alpha/2} > z_{\alpha/2} \Rightarrow$ interval je větší ↗ analogie pro $z_{\alpha/2}$
↪ nezávislé měření σ

Věta: Nechť $X_1, \dots, X_m \sim N(\mu, \sigma^2)$. Která máme μ ani σ , chceme určit μ .
Odešleme $\alpha \in (0, 1)$. Nechť $\Psi_{m-1}(1_{\alpha/2}) = 1 - \alpha/2$. Pak

$$\bar{X}_m \pm 1_{\alpha/2} \frac{\hat{S}_m}{\sqrt{m}} \quad \text{je } (1-\alpha)\text{-CI pro } \mu.$$

• Testování hypotéz

1) otázka ... mohouží správnou mísou piv? \rightarrow

H_0 ... mulová hypotéza (conservativní model) ... $E(X) = 500 \text{ ml}$

H_1 ... alternativní hypotéza (významná odchyly) ... $E(X) < 470 \text{ ml}$

2) design experimentu - rozvrhnout před začátkem měření

- vybereme vhodný statistický model ... $X \sim N(500, \sigma^2)$

- chyba 1. druhu: Zamítne H_0 , i když platí ... „Trefas“

\Rightarrow zvolíme hladinu významnosti α ... typicky $\alpha = 0.05$

\hookrightarrow test konstruujeme tak, aby $P[\text{ch. 1. d.}] = \alpha$

- chyba 2. druhu: Nezamítne H_0 , ale ona neplatí ... „Promarněna možnost“

\Rightarrow zvolíme parametr β \Rightarrow pak $P[\text{ch. 2. d.}] = \beta$

\hookrightarrow hodnota $1 - \beta$ se nazývá síla testu

- wrátíme statistický test $T = h(X_1, \dots, X_m)$... $T = \bar{X}_m$

- wrátíme kritický obor (rejection region) - množina W podle α, β

3) naměříme data $X_1, \dots, X_m \rightarrow$ realizace X_1, \dots, X_m

4) výhodnocení: pokud $T = h(X_1, \dots, X_m) \in W \Rightarrow$ zamítne H_0

$\Rightarrow \alpha = P[T \in W | H_0] \quad \dots$ i renamerá: neplatí, kde platí H_0

$\Rightarrow \beta = P[T \notin W | H_1] \quad \dots$ neplatí, kde platí H_1

5) výpočet dalsích rávání \rightarrow fórum $\alpha = 0.05 \Rightarrow$ neplatí, kde platí H_0
 $\alpha = 0.11 \Rightarrow$ zamítne

t-value := min. α , pro kterou bychom H_0 zamítli
 \hookrightarrow pro menší α ne

Příklady:

① Jednovýberový test

Náhled mi správou měru fira?

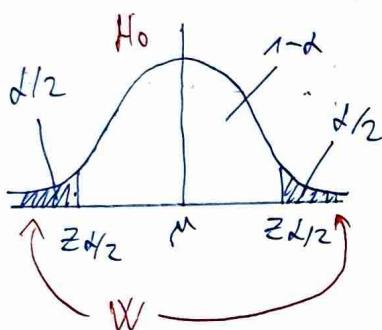
$X_1, \dots, X_m \sim N(\theta, \sigma^2)$, řešenéme, že σ známe režisérnost

$$H_0: \theta = 500 \text{ ml} \rightarrow \mu := 500 \text{ ml}$$

$$H_1: \theta \neq 500 \text{ ml} \rightarrow \text{pro jednoduchost}$$

→ chci statistiku něco jako \bar{X}_m , ale musím ho udělat W

$$\Rightarrow Z = \frac{\bar{X}_m - \mathbb{E}\bar{X}_m}{\sigma/\sqrt{m}} = \frac{\bar{X}_m - \mu}{\sigma/\sqrt{m}} \rightarrow \text{fodle CLV: } Z \sim N(0,1)$$



$$\Rightarrow \text{chci } P[Z \in W; H_0] = \alpha \Rightarrow P[Z \notin W; H_0] = 1-\alpha$$

$$\Rightarrow W = \{x \in \mathbb{R} \mid |x| > z_{\alpha/2}\}$$

↳ $(1-\alpha)$ -CI použití

→ B. rávise na 1om, jak blízko je $\theta \approx 500 \text{ ml}$.

② Dvouvýberový test

Náhled mi stejně jako kamarádovi?

$$\begin{aligned} X_1, \dots, X_m &\sim N(\theta_X, \sigma^2) & H_0: \theta_X = \theta_Y \\ Y_1, \dots, Y_m &\sim N(\theta_Y, \sigma^2) & H_1: \theta_X \neq \theta_Y \end{aligned}, \text{ předpokládáme, že známe } \sigma$$

→ rade chci statistiku něco jako $\bar{X}_m - \bar{Y}_m =: S$

↳ použud platí H_0 , takže $\mathbb{E}S = 0$

$$\hookrightarrow \text{Var } S = \text{Var } \bar{X}_m - \text{Var } \bar{Y}_m = \frac{\sigma^2}{m} + \frac{\sigma^2}{m}$$

$$\Rightarrow Z = \frac{S - \mathbb{E}S}{\sigma(S)} = \frac{\bar{X}_m - \bar{Y}_m}{\sigma \sqrt{\frac{1}{m} + \frac{1}{m}}} \rightarrow \text{fodle CLV: } Z \sim N(0,1)$$

• Z-test: jde o ①: $W = \{x \in \mathbb{R} \mid |x| < z_{\alpha/2}\}$

③ Co daly bychom σ neznati?

↳ v ① a ② jsme využili Z -test. Tedy ①, aby σ neznáme

$$\Rightarrow \text{T-test}: \sigma \approx \hat{S}_m = \frac{1}{m-1} \sum_i (X_i - \bar{X}_m)^2 \Rightarrow \sigma(\bar{X}_m) \approx \frac{\hat{S}_m}{\sqrt{m}}$$

$$\Rightarrow T = \frac{\bar{X}_m - \mu}{\hat{S}_m / \sqrt{m}} \quad \text{a} \quad W = \{x \in \mathbb{R} \mid |x| > t_{\alpha/2}\}, t_{\alpha/2} = \chi_{m-1}^{-1}(1-\alpha/2)$$

Hypothesis pro kategorická data - testy dobré' shody

- definice numerická data: kolik mali, #dětí, doba řešení programu
- hříčí řešení ... je spravedlivá? → G kategorie

	1	2	3	4	5	6
Observed: O_1, \dots, O_6	92	120	88	98	95	107
Expected: E_1, \dots, E_6	100	100	100	100	100	100

→ H_0 = řešení je spravedlivá

→ dán. se odvodit nějaký rozumný výrocík pro statistický test

$$G\text{-test: } G = 2 \cdot \sum_i O_i \cdot \log \frac{O_i}{E_i} \quad \forall i: E_i = O_i \Rightarrow G = \chi^2 = \emptyset$$

$$\chi^2\text{-test: } \chi^2 = \sum_i \frac{(E_i - O_i)^2}{E_i} \quad \text{Taylor. f.}$$

↳ měří vzdálenost od ideálního stavu $\chi^2 \doteq 6.86$

→ jak znít kritický obor pro dané α ?

⇒ chceme $P[T > \gamma | H_0] = \alpha$... pro $T = G$ nebo $T = \chi^2$

↳ jak znít γ ? ... co už je moc velké?

① exaktní test - pro malé $n = \# \text{ samplů}$

→ projde všechn $(\# \text{kategorií})^n$ možných výsledků

↳ respektive $\binom{n+k-1}{n}$... musí se to sečít na n

→ pro \forall výsledek vyhodnocím T a výšim γ , aby $T > \gamma$
pouze v 100 α procentech případů

pro málo pravdě
 $\gamma \doteq 11.1 \Rightarrow H_0$

② velké $n \Rightarrow$ užívám ① s nějakou směsí mnoha místních výzkumů

③ velké n & $T = \chi^2 \rightarrow$ aproximace formou Asy. χ^2 rozdělení s $k-1$ stupni volnosti

χ^2_{k-1} je rozdělen m.n. $Q = Z_1^2 + \dots + Z_{k-1}^2$, $Z_i \sim N(0,1)$ → poslední kategorie dostatečně

↳ dá se ukázat, že $\chi^2 \approx Q$. (používám H_0)

⇒ potom $\gamma = F_Q^{-1}(1-\alpha) \Rightarrow P[Q > \gamma] = P[\chi^2 > \gamma] = \alpha$.

Druhy testů

- 1- nýberový test $X_1, \dots, X_n \sim N(\mu, \sigma^2) \rightarrow H_0: \mu = 5$

stats. ttest_1samp(x, np.mean=5)

- 2-nýberový test $X_1, \dots, X_m \sim N(\mu_x, \sigma^2)$ $Y_1, \dots, Y_n \sim N(\mu_y, \sigma^2)$ $\left. \right\} H_0: \mu_x = \mu_y$

stats. ttest_ind(x, y, equal_var=True)

\hookrightarrow pokud $\sigma_x \neq \sigma_y \Rightarrow \text{equal-var=False}$

- fárový test $X_1, \dots, X_m \sim N(\mu_x, \sigma^2)$ $Y_1, \dots, Y_n \sim N(\mu_y, \sigma^2)$ $\left. \right\} H_0: \mu_x = \mu_y$

$\rightarrow X_i, Y_i$ sponěj nějak souvisí ... běh algoritmu jezd / optimální

\Rightarrow mohlo bych udělat 2-nýber. 1-test, ale jde se o hřív

\Rightarrow udělím $D_i := X_i - Y_i$ a dělám

1-nýberový 1-test $D_1, \dots, D_m \rightarrow H_0: \underbrace{\mu_D}_{\mu_X - \mu_Y} = 0$

stats. ttest_rel(x, y)

Lineární regrese

data: (x_i, y_i) pro $i=1, \dots, n$

↳ cíl: $y = ax + b$... x = nezávislá 'proměna' - predictor
 y = závislá 'proměna' - response

↳ cíl: zjednodušit model pomocí MSE

$$\sum_{i=1}^n (y_i - (ax_i + b))^2 \Rightarrow \text{chceme} \rightarrow \text{minimalizovat}$$

řešení:

$$b = \bar{x}_m / \bar{y}_m$$

$$a = \frac{\sum_i (x_i - \bar{x}_m)(y_i - \bar{y}_m)}{\sum_i (x_i - \bar{x}_m)^2} = \frac{\text{cov}(\bar{x}_m, \bar{y}_m)}{\text{var}(\bar{x}_m)}$$

spis závislosti corr a var