23.9.

I. POČET PRAVDĚPODOBNOSTI

1. ÚVOD

Požadavky: průměrná hmotnost 250g

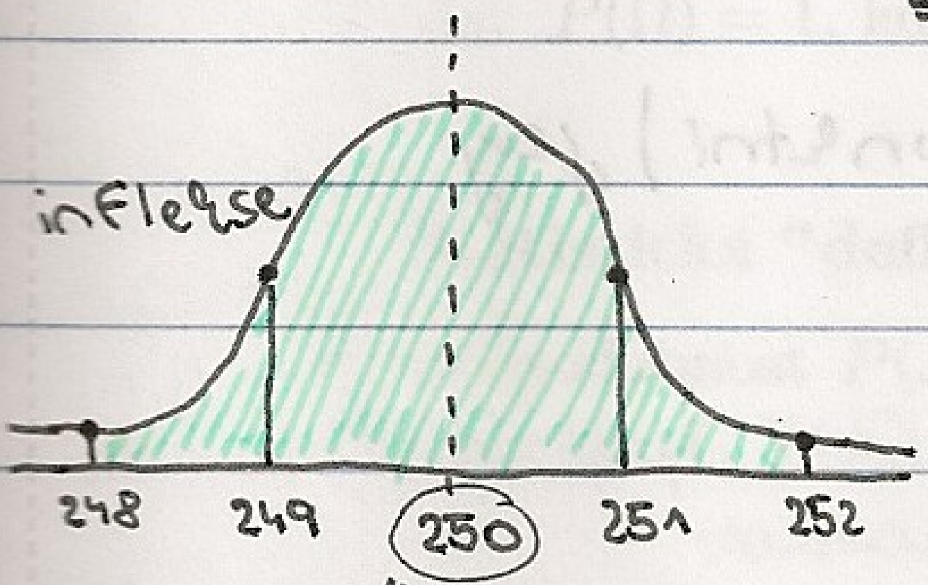
$\underbrace{19 \text{ z } 20 \text{ výrobků}}_{95\%} \in (248, 252\,g)$ interval gramáže

• malé vzdálenosti (odchylky) od 250g jsou více časté než velké

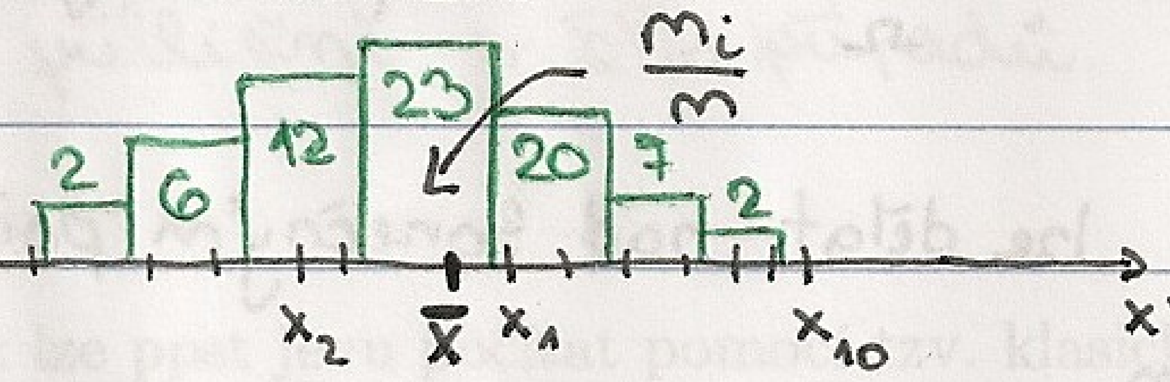
=> Gaussovo (normální) rozdělení pravděp.

• Gaussova křivka je symetrická - odchylky na obou stranách nastanou se stejnou pravděp.

• kontrolní měření: m = 100 (100 měření)



≈ 95% plochy



histogram relativních četností

$\bar{x} \approx \mu$

$\parallel$

249,3

• tento model neodpovídá Gaussovu rozdělení úplně - jen s minimálními odchylkami

2. NÁHODNÝ JEV

Náhodný pokus - aspoň dva různé výsledky

Náhodné jevy - podmnožiny množiny Ω všech možných výsledků nějakého pokusu

Označení: $A \subset \Omega, B \subset \Omega, \ldots$

$\emptyset$ ... nemožný jev

Ω ... jistý jev

Operace s jevy:

- sjednocení dvou jevů: $A \cup B$ (A nebo B)
- průnik dvou jevů: $A \cap B$ (A a B)
- negace jevu A: $\overline{A}$ (jev opačný neboli doplňkový k A)

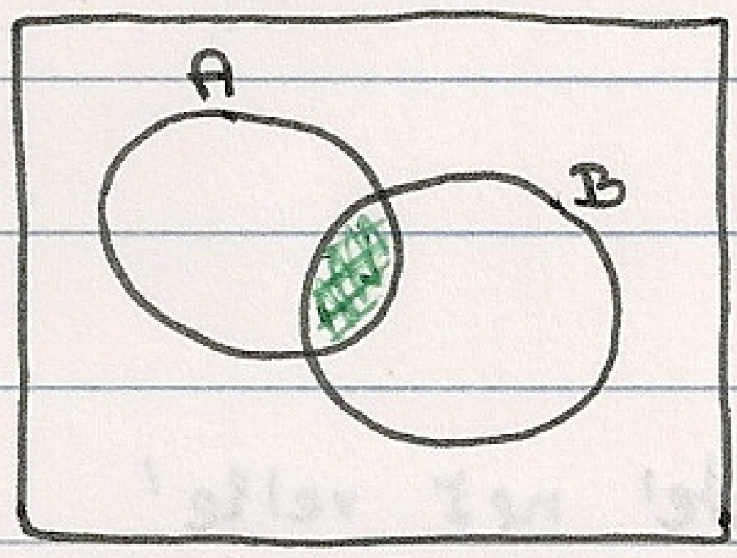
Neslučitelné (disjunktní) jevy:

Jevy A, B se nazývají neslučitelné, je-li $A \cap B = \emptyset$.

Náhodný pokus - např. hody kostkou nebo mincí

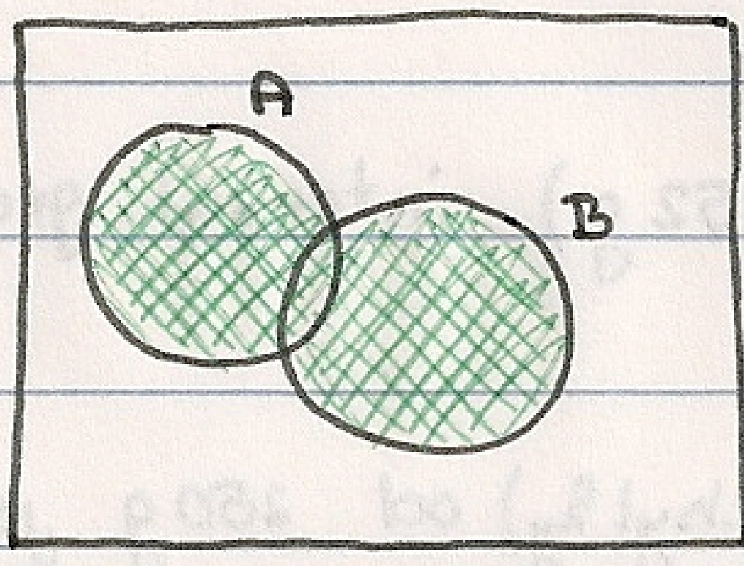
Nemožný jev - na kostce padne 7

<u>Znázornění jevů</u>: s jevy můžeme dělat operace jako s množinami



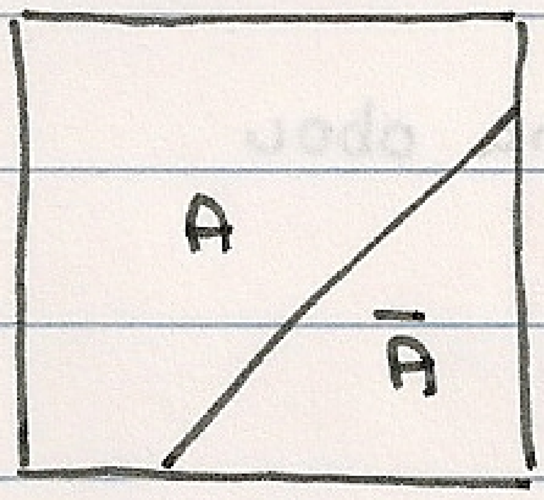
Ω

$A \cap B$



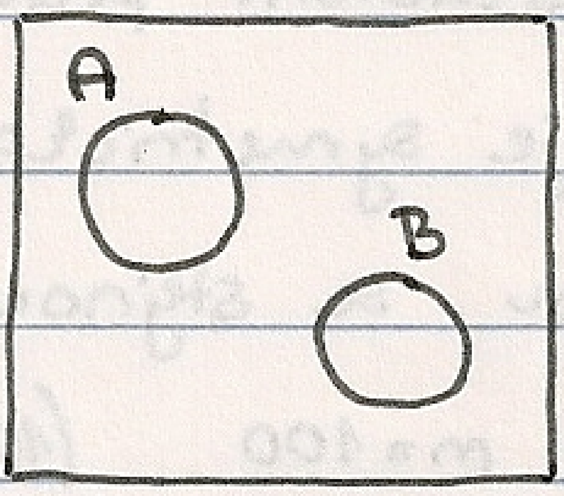
Ω

$A \cup B$



Ω

negace



Ω

neslučitelné (disjunktní) jevy

$A \cap B = \emptyset$

(př: negace)

- operace nad množinami lze dělat nad konečným počtem ~~jev~~ pokusů

$\bigcup_{i=1}^{m} A_i$ $\bigcap_{i=1}^{m} A_i$

př:

Náh. pokus	Ω	Příklad jev	Množinový zápis
Hod hrací kostkou	$\Omega = 1,2,3,4,5,6$	l ... padne liché číslo m ... padne číslo < 3	$L = \{1,3,5\}$ $M = \{1,2\}$
Počet zmetků z 20-ti kontrolovaných	$\Omega = \{0,1,2,\dots,20\}$	d ... nejvýše 2 zmetky	$D = \{0,1,2\}$
Doba fungování součástky [h]	$\Omega = \langle 0, +\infty)$ teoreticky	c ... součástka funguje < 100[h]	$C = \langle 0, 100)$

pak např: $L \cup M = \{1,2,3,5\}$ $\overline{D} = \{3,4, \dots 20\}$ $L \cap M = \emptyset$

$L \cap M = \{1\}$ $\overline{L} = \{2,4,6\}$ $C = \langle 100, +\infty)$

3. PRAVDĚPODOBNOST JEVU

$$A \longrightarrow P(A)$$

A...jev, P(A)...**pravděpodobnost (ppst) jevu A**

Axiomy ppsti:

A_1: $0 \leq P(A) \leq 1$;

A_2: pro neslučitelné jevy $A_1, A_2, \ldots$ platí: $P(\bigcup_i A_i) = \sum_i P(A_i)$;

A_3: $P(\Omega) = 1$, kde Ω je jev jistý.

Statistická "definice" ppsti

Pravděpodobnost $P(A)$ jevu A je limita relativní četnosti jevu A, zvětšujeme-li počet pokusů $n \to \infty$. Např. $P(A) = 0{,}05$ znamená, že při velkém množství pokusů nastane jev A přibližně v 5% případů.

Klasická definice ppsti

Za určitých specifických podmínek lze ppst jevu počítat pomocí tzv. klasické definice ppsti - více **viz cvičení PSA**

$$P(A) = \frac{N_A}{N} \quad .$$

Některé další možné definice ppsti (tzv. **geometrická, axiomatická (Kolmogorovova)** budou v závěru ZS.

př. Relativní četnost: A ... jev

$$h(A) = \frac{m_A}{m} \xrightarrow{\text{"}m \to \infty\text{"}} P(A)$$

↑ relativní četnost jev A v m pokusech

N_A ... počet výsledků příznivých jev A

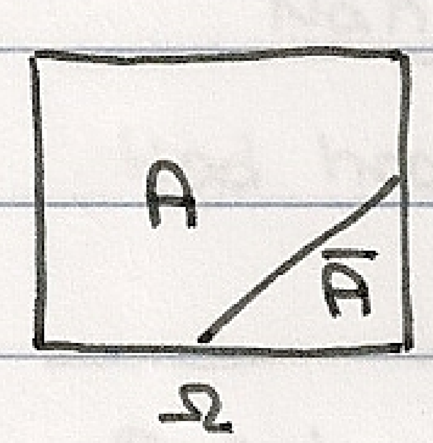
N ... počet výsledků možných, je konečné

každý výsledek má stejnou pravděp.

Základní pravidla pro pravděpodobnost

Základní pravidla pro ppst (lze je odvodit z axiomů A_1, A_2, A_3):

- $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(\overline{A}) = 1 - P(A)$ **(ppst opačného jevu)**
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  (ppst sjednocení jevů)



axiom 3 (nejspíš)

$$1 \overset{A_3}{=} P(\Omega) = \underbrace{P(A \cup \bar{A})}_{\text{neslučitelné}} \overset{A_2}{=} P(A) + P(\bar{A})$$

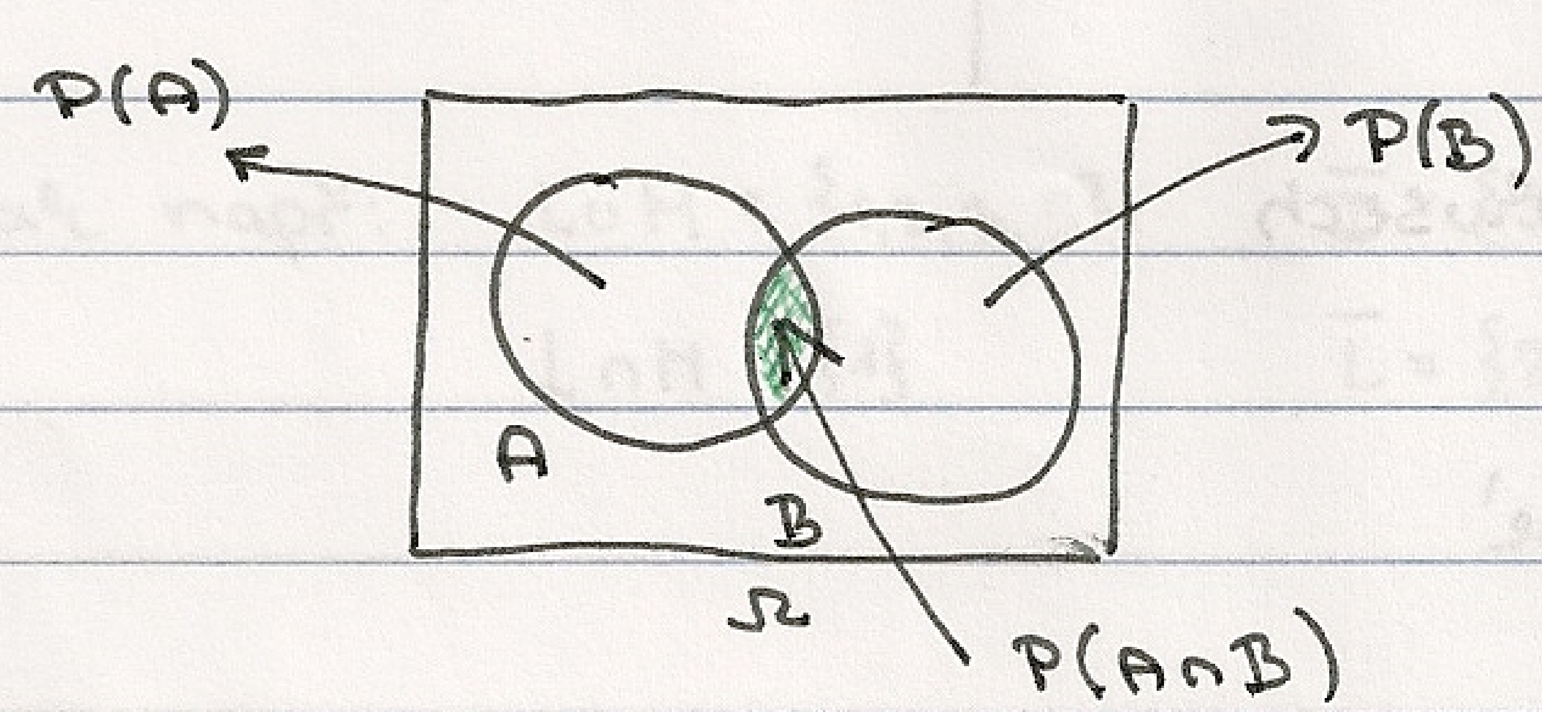
axiom 2

$A \cap \bar{A} = \emptyset$ $\qquad A \cup \bar{A} = \Omega$

$A = \Omega \qquad \bar{\Omega} = \emptyset$

$$1 \overset{A_3}{=} P(\Omega) = P(\Omega \cup \emptyset) \overset{A_2}{=} \underbrace{P(\Omega)}_{1} + P(\emptyset)$$

$\Rightarrow P(\emptyset) = 0$



A, B neslučitelné $\Leftrightarrow A \cap B = \emptyset$

Př.: Náhodně vybereme jednu z cifer 0,1,2, ... 9

A ... vybraná cifra bude < 6

B ... vybraná cifra bude lichá

určete $P(A \cup B)$

Řešení: $\Omega = \{0,1, \dots 9\}$

$A = \{0,1,2,3,4,5\}$

$B = \{1,3,5,7,9\}$

$A \cap B = \{1,3,5\}$

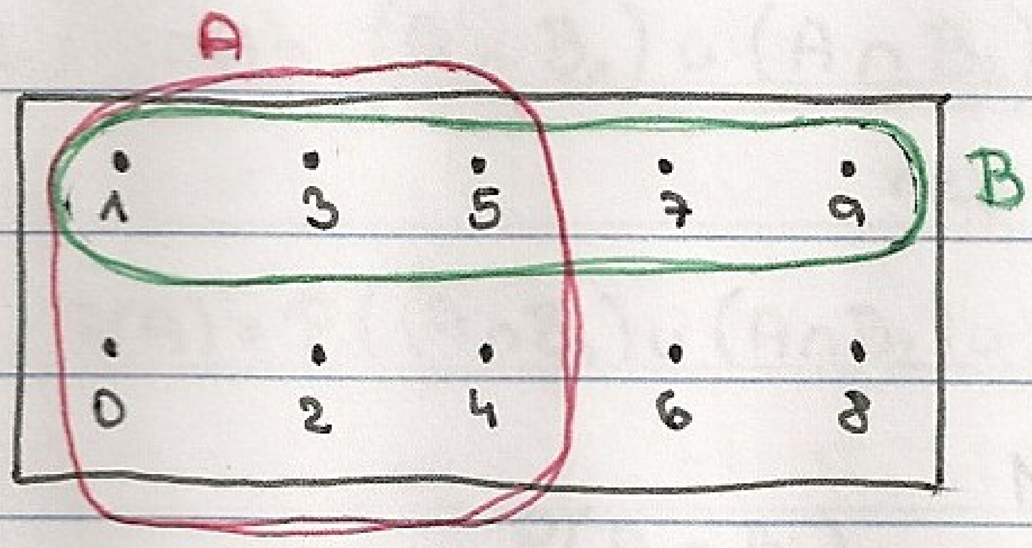
$P(A) = \frac{6}{10}$

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

$P(A \cap B) = \frac{3}{10}$

↑ pravděp. průniku

$P(A \cup B) = \underbrace{\frac{6}{10} + \frac{5}{10}}_{1,1} - \frac{3}{10} = 0,8$



Musíme dávat pozor na závislé nebo nezávislé jevy.

A ... padne "6" B ... padne "6" $P(A \cap B) = \frac{1}{36}$

$P(A) = \frac{1}{6}$ $P(B) = \frac{1}{6}$ ⇒ A,B - nezávislé jevy

A ... 1. vybraný

B ... 2. vybraný

(bez vrácení 1. zpět)

⇒ závislé jevy

$A_1, \dots A_n$ nezávislé ⇒ $P\left(\bigcap_{i=1}^{m} A_i\right) = \prod_{i=1}^{m} P(A_i)$

4. PODMÍNĚNÁ PRAVDĚPODOBNOST

Definice. Nechť A, B jsou jevy, $P(B) \neq 0$.
Pak **pravděpodobnost jevu A podmíněná jevem B** se značí $P(A|B)$ a je definována jako podíl

$$\mathbf{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Analogicky (pro $P(A) \neq 0$):

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

Nezávislost jevů A, B.

Požadavek: Výskyt jevu B nemá ovlivňovat ppst jevu A, resp. naopak, tj. $P(A|B) = P(A)$, resp. $P(B|A) = P(B)$.

Dosazením těchto rovností do levé strany (1), resp. (1') a úpravou dostaneme **pro nezávislé jevy** A, B :

$$\mathbf{P(A \cap B) = P(A) \cdot P(B)} \qquad (2)$$

Rovnost (2) lze použít jako **definici nezávislosti jevů** A, B.

Nejsou-li jevy nezávislé, pak se nazývají **závislé** (a rovnost (2) pro ně neplatí).

Pojem nezávislosti lze rozšířit na n jevů $A_1, A_2, \dots, A_n$ $(n \geq 3)$.

Př: Výrobní linka

A ... jev "výrobek má vyhovující rozměry"

B ... -II- ————— II ————— hmotnost"

Dlouhodobým pozorováním byly zjištěny tyto relativní četnosti (≈ pravděp.): $P(A) = 0{,}96$

$P(B) = 0{,}93$

$P(\overline{A} \cup \overline{B}) = 0{,}09$

a, jsou jevy A, B nezávislé?

b, spočtěte $P(B|A)$

c, spočtěte $P(A \cup B)$

a, $P(A \cap B) \overset{?}{=} P(A) \cdot P(B)$

$$P(A \cap B) = 1 - P(\overline{A \cap B}) = 1 - \overbrace{P(\overline{A} \cup \overline{B})}^{0{,}09} = 0{,}91$$

deMorganovy zákony

$$0{,}91 \neq \underbrace{0{,}96 \cdot 0{,}93}_{0{,}8928} \Rightarrow A, B \text{ jsou } \underline{\underline{\text{závislé}}}$$

b, pravděp. jev B za podmínky, že nastane jev A

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0{,}91}{0{,}96} \doteq \underline{0{,}948} \quad (\neq P(B))$$

c, $P(A \cup B) = P(A) + P(B) = 0{,}96 + 0{,}93 - 0{,}91 = \underline{\underline{0{,}98}}$

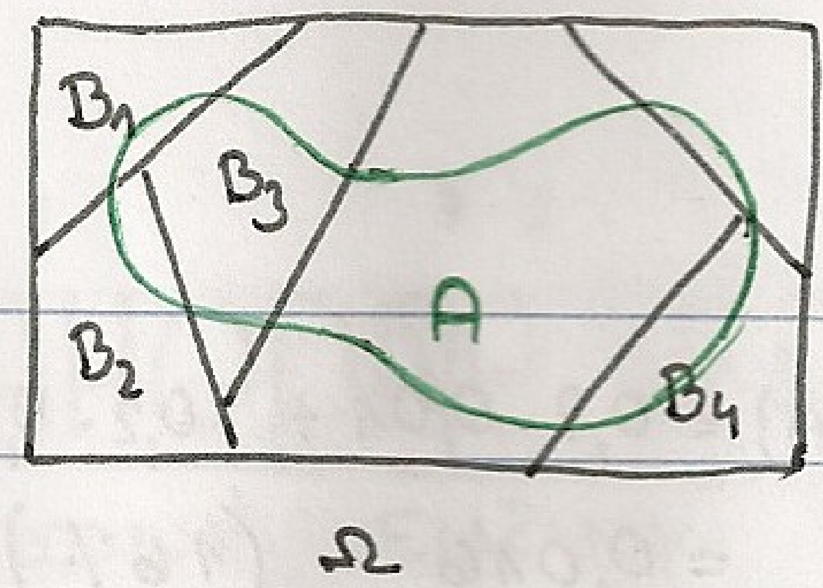
5. VĚTA O ÚPLNÉ PRAVDĚPODOBNOSTI, BAYESOVA VĚTA

Nechť pro jevy $B_1, B_2, \ldots, B_n$ platí: $B_i \cap B_j \quad \forall i \neq j$, $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$, $P(B_i) > 0 \quad \forall i = 1, 2, \ldots, n$, a nechť A je libovolný jev (tj. $A \subset \Omega$).

Pak platí:

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

(tzv. věta o úplné ppsti)

Znázornění: 

$$\bigcup_{i=1}^{m} B_i = \Omega$$

$B_i \cap B_j = \emptyset \quad \forall i \neq j$

$A \subset \Omega \quad P(A) = P(A \cap B_1) \cup (A \cap B_2) \cup \ldots (A \cap B_m)$

neslučitelné

$$\overset{A_2}{=} \sum_{i=1}^{m} P(A \cap B_i) = \sum_{i=1}^{m} P(B_i) \cdot P(A|B_i)$$

Je-li navíc $P(A) > 0$, pak pro $k = 1, 2, \ldots, n$ platí:

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

(tzv. **Bayesova věta**, věta o inverzní ppsti)

Důkaz: $A = \underbrace{(A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_m)}_{\text{neslučitelné}}$

$$P(A) = P((A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_m)) \overset{\text{axiom } A_2}{=} \sum_{i=1}^{m} P(A \cap B_i) = \sum_{i=1}^{m} P(B_i) P(A|B_i)$$

$$P(A|B_i) = \frac{P(A \cap B_i)}{P(B_i)}$$

(Bayes) $\quad P(B_k|A) = \dfrac{P(A \cap B_k)}{P(A)} = \dfrac{P(B_k) \cdot P(A|B_k)}{P(A)}$

$P(A|B_k)$ ... apriorní pravděp.

$P(B_k|A)$ ... aposterio pravděp.

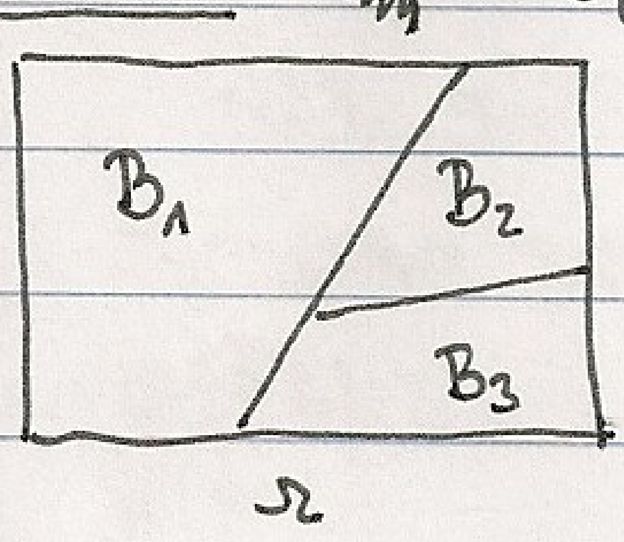
Př: Výrobek je vyráběn na právě jedné ze tří linek

1. linka, 70% produkce, zmetkovitost 1%
2. linka, 20% -//- , -//- 2%
3. linka, 10% -//- , -//- 5%

a, určete pravděp., že náhodně vybraný výrobek bude zmetek (věta o úplné pravd.)

b, je-li náhodně vybraný výrobek zmetek, určete pravděp., že tento výrobek byl vyroben na 1., resp. 2, resp. 3. lince

Řešení: B_i ... výrobek vyráběn na i-té lince $(i = 1, 2, 3)$



A ... jev "náh. vybraný výrobek bude zmetek" $\quad B_1 \cup B_2 \cup B_3 = \Omega$

a) $P(B_1) = 0{,}7$ $\quad P(A|B_1) = 0{,}01$ $\quad P(A) = 0{,}7 \cdot 0{,}01 + 0{,}2 \cdot 0{,}02 + 0{,}1 \cdot 0{,}05 =$

$P(B_2) = 0{,}2$ $\quad P(A|B_2) = 0{,}02$ $\quad = \underline{\underline{0{,}016}}$ (1,6%)

$P(B_3) = 0{,}1$ $\quad P(A|B_3) = 0{,}05$ $\quad$ průměrná zmetkovitost

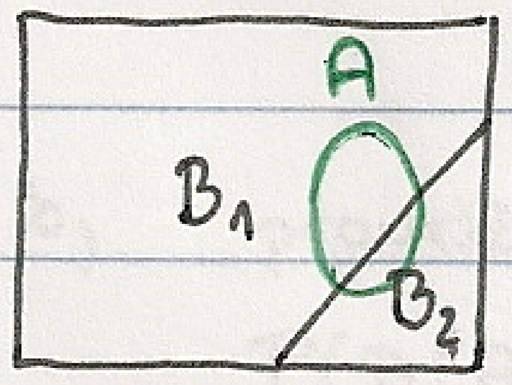
(Bayes)

b) $$P(B_1|A) = \frac{P(B_1) \cdot P(A|B_1)}{P(A)} = \frac{0{,}7 \cdot 0{,}01}{0{,}016} = \frac{7}{16} \; (= 0{,}4375)$$

$$P(B_2|A) = \frac{0{,}2 \cdot 0{,}02}{0{,}016} = \frac{4}{16} \; (=0{,}25) \qquad P(B_3|A) = \frac{0{,}1 \cdot 0{,}05}{0{,}016} = \frac{5}{16} \; (= 0{,}3125$$

<u>Př:</u> V produkci je průměrně 4% vadných výrobků. Je-li výrobek vadný, kontrolní přístroj tuto vadu nalezne s pravděp. 0,95. Není-li výrobek vadný, kontr. přístroj s pravděp. 0,02 signalizuje vadu. Náhodně vybraný výrobek byl přístrojem označen jako vadný. S jakou pravděp. tento výrobek skutečně vadný je?

<u>Řešení:</u> B_1 ... jev "náh. vybraný výrobek bude dobrý"



$B_2 = \overline{B_1}$ jev " —//— vadný "

$P(B_1) = 0{,}96$ $\quad P(A|B_1) = 0{,}02$

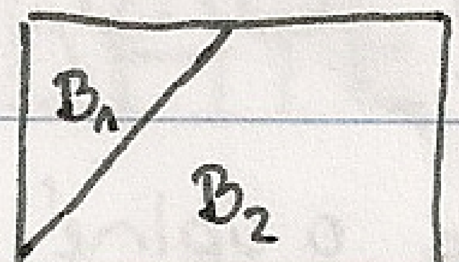
$P(B_2) = 0{,}04$ $\quad P(A|B_2) = 0{,}95$

A ... jev "přístroj signalizuje vadu"

$$P(B_2|A) \overset{\text{Bayes}}{=} \frac{0{,}04 \cdot 0{,}95}{0{,}96 \cdot 0{,}02 + 0{,}04 \cdot 0{,}95} \doteq 0{,}664$$ zhruba třetina je dobrá

"nízká pravděp."

"2. kolo"



$P(B_1) = 0{,}336$

$P(B_2) = 0{,}664$ $\quad$ $$P(B_2|A) = \frac{0{,}664 \cdot 0{,}95}{0{,}336 \cdot 0{,}02 + 0{,}664 \cdot 0{,}95} = 0{,}989$$

Výrobek 2x označený jako vadný s pravděp. 0,989 vadný je.

II. NÁHODNÁ VELIČINA

1. POJEM NÁHODNÉ VELIČINY

Náhodná veličina (náhodná proměnná) je "vhodná" reálná funkce definovaná na množině Ω

$$X : \Omega \to R$$

(zpřesnění bude dále)

Označení:

X, Y, Z ... **náhodné veličiny** (jsou to funkce)

$x,\ y,\ z$ ... jejich **realizace** (jsou to reálná čísla)

Náhodná veličina: a) **diskrétní** - může nabývat (s nenulovou ppstí) jen konečně, popř. spočetně mnoha hodnot ;

b) **spojitá** - může nabývat libovolné hodnoty z nějakého intervalu (intervalů) reálných čísel.

U náh. veličiny nutno určit též její **rozdělení ppsti**.

$P(X \le x)$ .. pravděp. jev "náh. veličina X nabyde hodnoty $\le x$"

$P(X = x)$ .. pravděp. jev "náh. veličina X nabyde hodnoty x", krátce píšeme $P(x)$

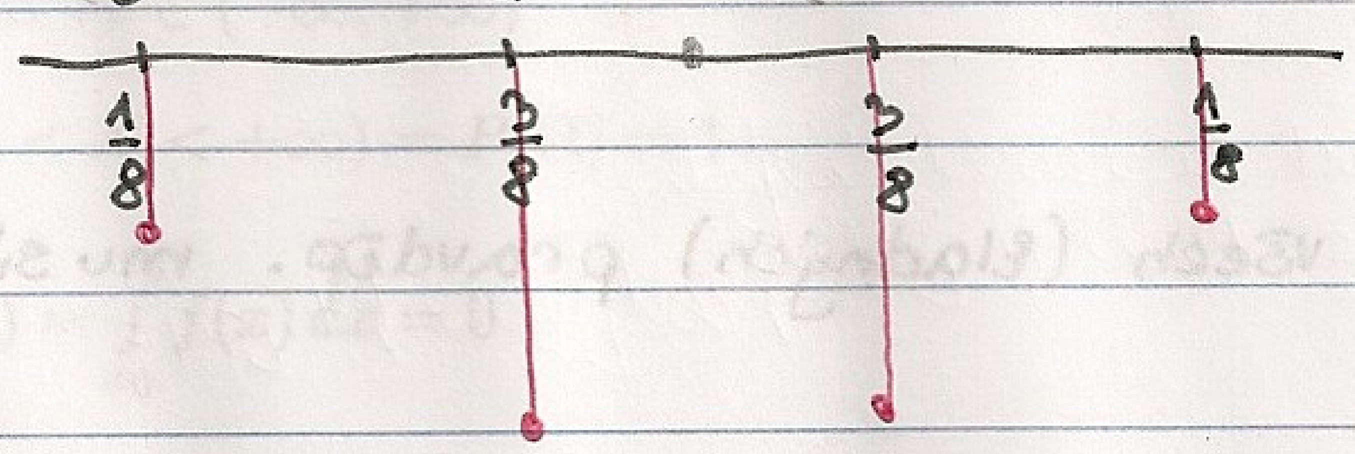
př.: 1, Nechť X je počet líců při hodu třemi mincemi. Pak X je diskr. náh. veličina (s nenulovou pravděp. může nabýt jen hodnot 0,1,2,3). Pravděp. přiřadíme použitím klasické definice:

$\Omega = \{ \circ\circ\circ, \bullet\circ\circ, \circ\bullet\circ, \circ\circ\bullet, \bullet\bullet\circ, \bullet\circ\bullet, \circ\bullet\bullet, \bullet\bullet\bullet \}$

$\circ$ ... ~~líc~~ rub

$\bullet$ ... ~~rub~~ líc

0 1 2 3



$P(0) = \frac{1}{8}$

$P(1) = \frac{3}{8}$

$P(2) = \frac{3}{8}$

$P(3) = \frac{1}{8}$

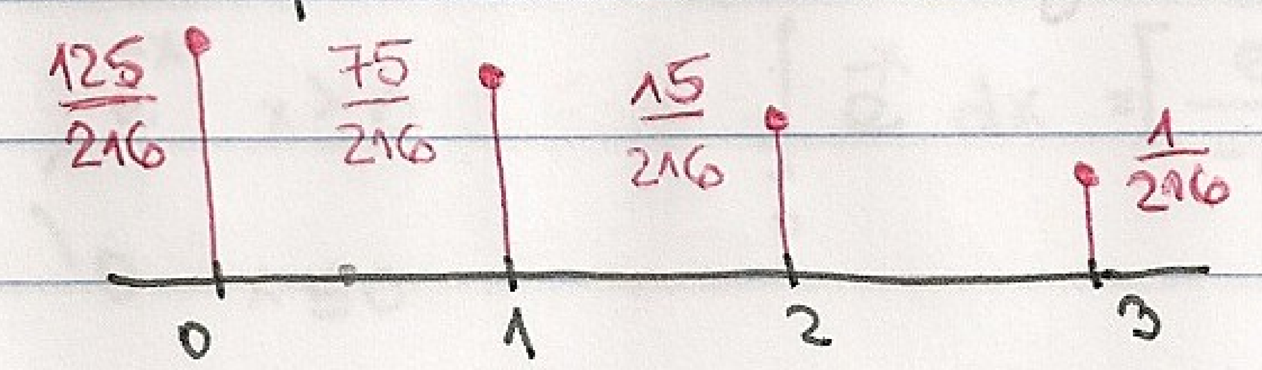
$E(X) = \frac{3}{2}$

př.: 2, Nechť X je počet šestek při hodu třemi kostkami. Pak X je diskr. náh. veličina.

Ω = jako v př. 1

$\circ\ \{\neq 6\}$

$\bullet\ \{= 6\}$



$E(X) = \frac{1}{2}$

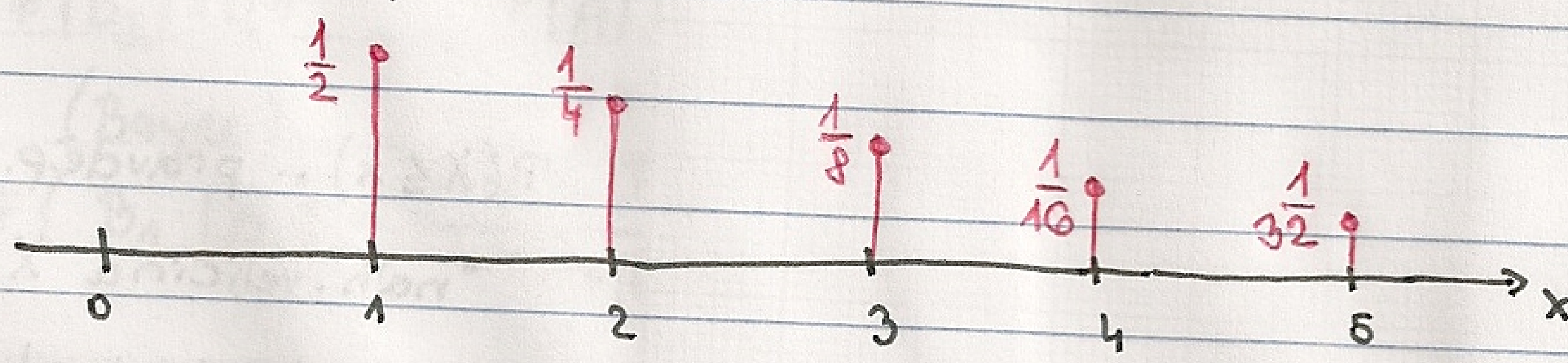
$P(0) = \left(\frac{5}{6}\right)^3$

$P(1) = 3 \cdot \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6}$

$P(2) = 3 \cdot \frac{5}{6} \cdot \frac{1}{6^2}$

$P(3) = \left(\frac{1}{6}\right)^3$

př.: 3, Házíme mincí, dokud nepadne líc. Nechť X je počet potř. hodů. Pak X je diskr. náh. veličina (s nenul. pravděp nabývá hodnot 1,2,3, ...)



$$P(x) = \begin{cases} \frac{1}{2^x} & \dots x \in \mathbb{N} \\ 0 & \dots x \notin \mathbb{N} \end{cases}$$

$\Omega = \{\bullet, \circ\bullet, \circ\circ\bullet, \dots\}$

Definice: Řekneme, že náhodná veličina X má rozdělení diskrétního typu (krátce: **DISKRÉTNÍ** náh. veličina), právě když existuje konečná nebo spočetná množina reálných čísel $x_1, x_2, \dots$ taková, že:

$$P(X = x_j) > 0 \quad pro \quad j = 1, 2, \dots$$

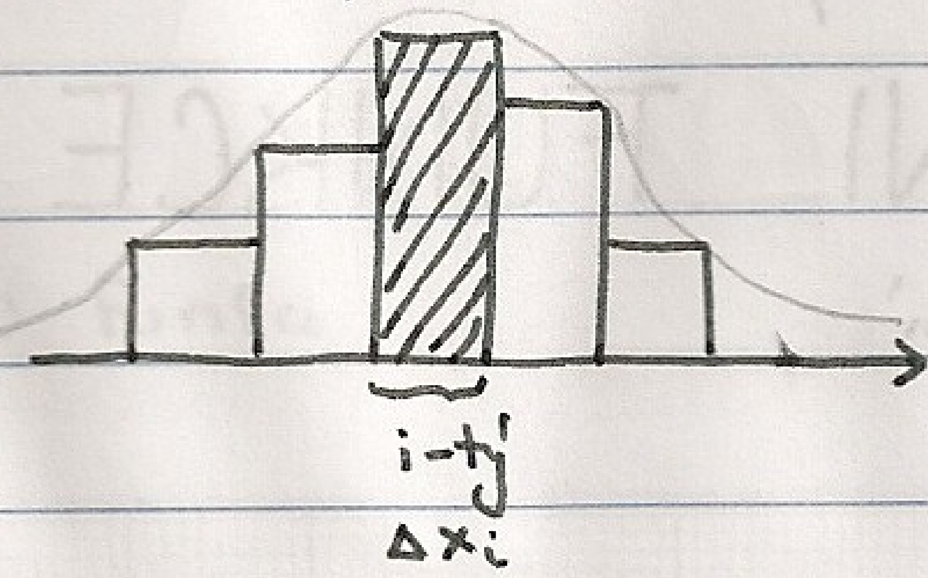
$$a \qquad \sum_{x_j} P(x_j) = 1$$

Funkci, která u **diskrétní** náh. veličiny X přiřadí každému $x \in R$ ppst $P(X = x)$, krátce $P(x)$, nazveme **PRAVDĚPODOBNOSTNÍ FUNKCE**

! Nezapomeňme: Součet všech (kladných) pravděp. musí být 1.

Další příklady diskr. náh. veličin:
- počet vadných výrobků z n kontrolovaných
- počet pojistných událostí (např. za rok)
- počet zákazníků čekajících na obsluhu

obsah ▨ $\approx \frac{N_i}{N}$



"$N \to \infty$"

"$\Delta x_i \to 0$"

př: N ... počet pokusů

N_i ... -||- -||- ,

jejichž výsledek je v

i-tém podintervalu

$\sum \frac{N_i}{N} = 1$

Jiným typem náh.veličin jsou náh.veličiny spojitého typu (krátce: **SPOJITÉ** náh.veličiny).

Pro **spojitou** náh. veličinu se definuje tzv. **HUSTOTA PPSTI** jako funkce $f(x)$, k níž se "blíží" histogram relativních četností při zjemňujícím se dělení viz obr.

Zřejmě hustota ppsti je **nezáporná funkce**
a plocha pod jejím grafem je rovna 1.

Definice: HUSTOTA ppsti náh.veličiny X je funkce $f(x)$ definovaná na intervalu $(-\infty, +\infty)$ taková, že:

$$\int_{x_1}^{x_2} f(x)\,dx = P(x_1 < X < x_2)$$

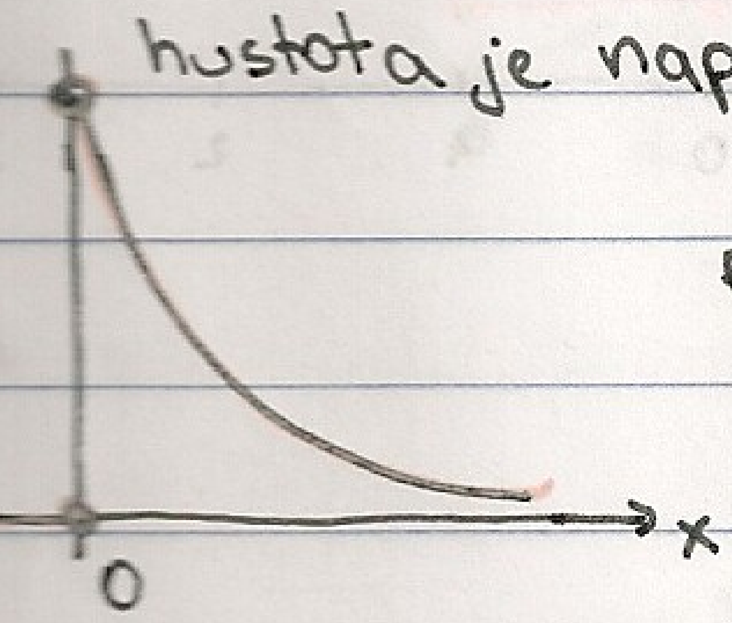
pro libovolná x_1, x_2 taková, že $x_1 < x_2$.

Pozn.: 1) $f(x) \geq 0$ pro $\forall x \in (-\infty, +\infty)$

2) $\int_{-\infty}^{+\infty} f(x)\,dx = P(-\infty < X < +\infty) = P(\Omega) = 1$

3) $P(X = x_0) = P(x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$

př.: 4, Životnost výrobku X lze chápat jako spojitou náh. veličinu. Její hustota je např.



$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$\int_0^{+\infty} e^{-x}\,dx = \left[\frac{e^{-x}}{-1}\right]_0^{+\infty} = 0 - (-1) = 1$$

exp. funkce popisující životnost

2. DISTRIBUČNÍ FUNKCE = nejdůležitější definice PSA !

distribuce = rozdělení

Popsat rozdělení ppsti náh. veličiny X (**diskr.** nebo **spoj.**) lze též pomocí tzv. **DISTRIBUČNÍ FUNKCE** $F(x)$, kterou definujeme:

$$\boxed{F(x) = P(X \leq x) \quad \text{pro} \quad x \in (-\infty, +\infty)}$$

Z definice $F(x)$ ihned plyne:

- $0 \leq F(x) \leq 1$ pro $x \in (-\infty, +\infty)$;
- $F(x)$ je **neklesající** funkce ;
- $\lim_{x \to -\infty} F(x) = 0$ a $\lim_{x \to +\infty} F(x) = 1$.

Distribuční fce **diskrétní** náh. veličiny je **nespojitá**,
(je *"schodovitá"*
- neklesající a po částech konstantní).

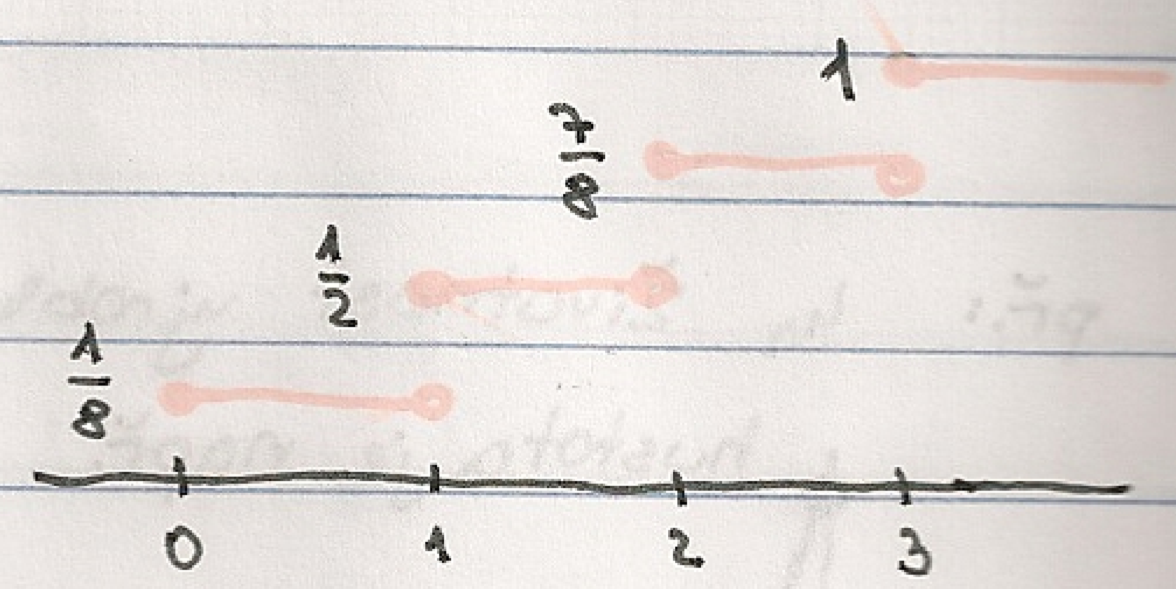
Distribuční fce **spojité** náh. veličiny je **spojitá**.

$x_1 < x_2$

$$\{X \leq x_1\} \subset \{X \leq x_2\} \Rightarrow \underbrace{P(X \leq x_1)}_{F(x_1)} \leq \underbrace{P(X \leq x_2)}_{F(x_2)}$$

z 1. příkladu:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} & x \in \langle 0, 1) \\ \frac{1}{8} + \frac{3}{8} = \frac{1}{2} & x \in \langle 1, 2) \\ \frac{7}{8} & x \in \langle 2, 3) \\ 1 & x \geq 3 \end{cases}$$



Pozn.: X ... diskr. náh. veličina

$P(x)$ ... pravděpodobnostní funkce

$F(x)$ ... distrib. funkce

Je-li dána pravděp. funkce $P(x)$, pak $F(x) = \sum\limits_{\substack{x_j :\, x_j \le x \\ P(x_j) > 0}} P(x_j)$

=> když budu mít danou pravděp. funkci, můžu z ní vypočítat distrib. funkci.

Je-li dána distrib. funkce $F(x)$, pak $P(x) = F(x) - \lim\limits_{\hat{x} \to x^-} F(\hat{x}) =$ zleva

$$= \begin{cases} 0 & \text{je-li } x \text{ bod spojitosti} \\ > 0 & \text{-\textbar\textbar- } x \text{ bod nespojitosti} \end{cases}$$

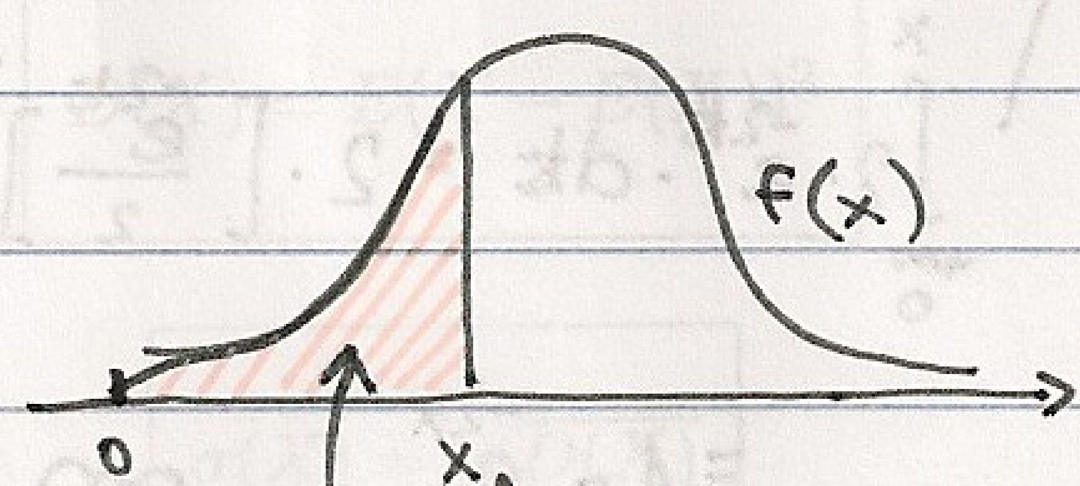
"výška schodu"

např.: $P(2) = F(2) - \underbrace{\lim\limits_{x \to 2^-} F(2)}_{F(1)} = \frac{7}{8} - \frac{1}{2} = \frac{3}{8}$

př.:

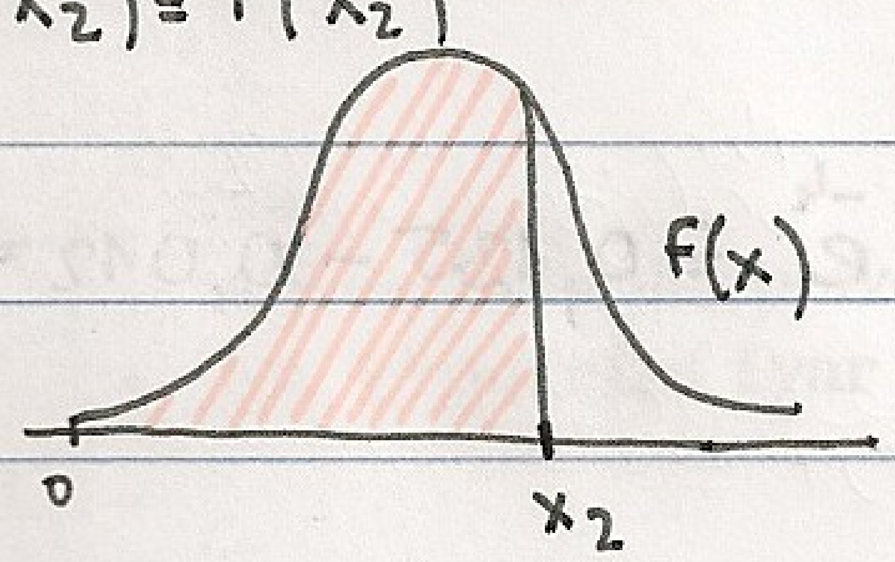
X ... spojitá

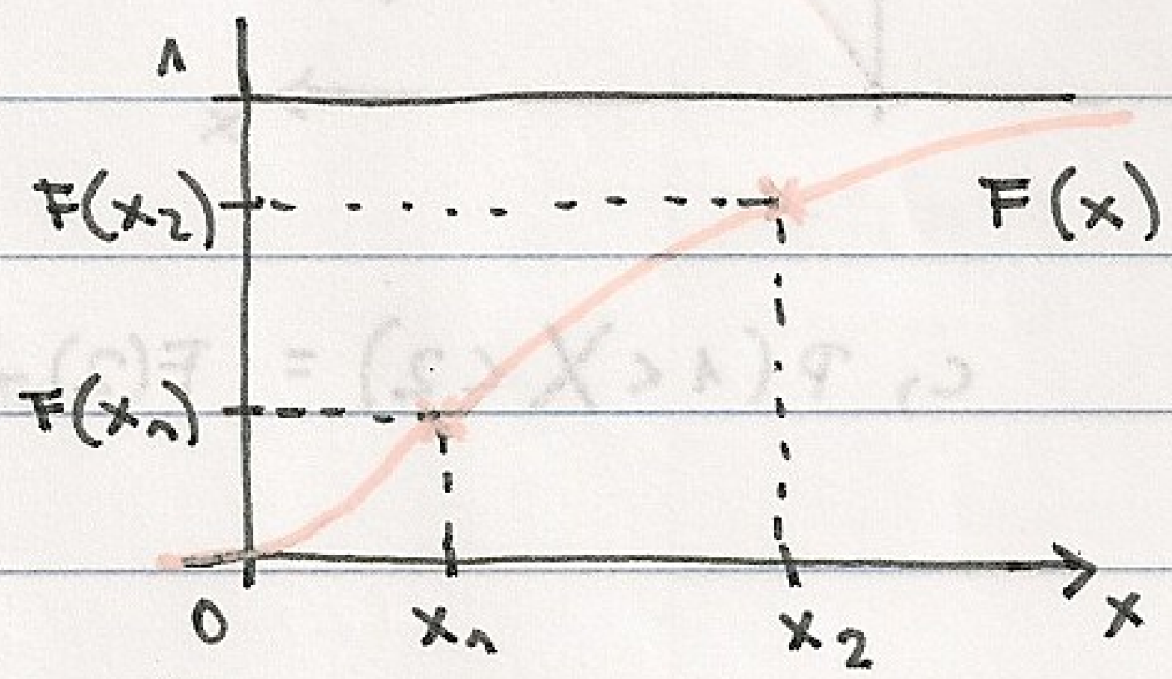
$f(x)$ ... hustota

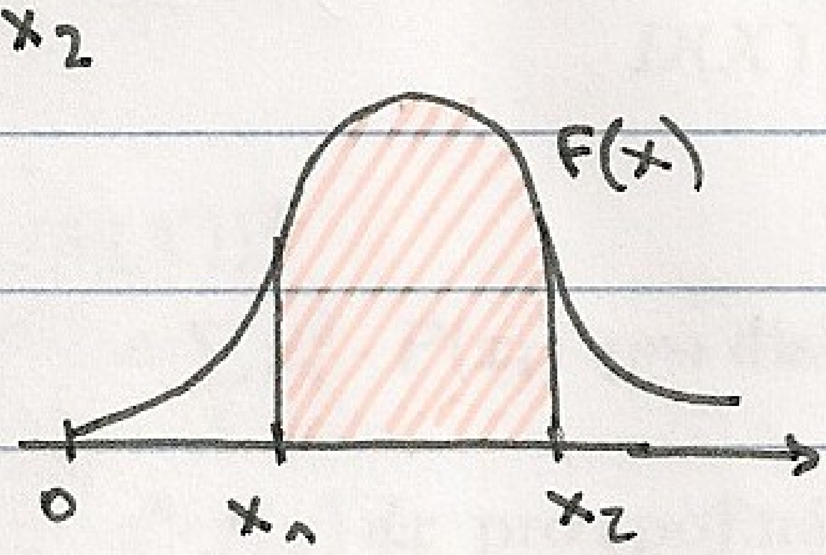


$P(X \le x_1) = F(x_1)$

$P(X \le x_2) = F(x_2)$







$$\boxed{P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = F(x_1) - F(x_2)}$$

Pozn.: 1, je-li dána hustota $f(x)$, pak

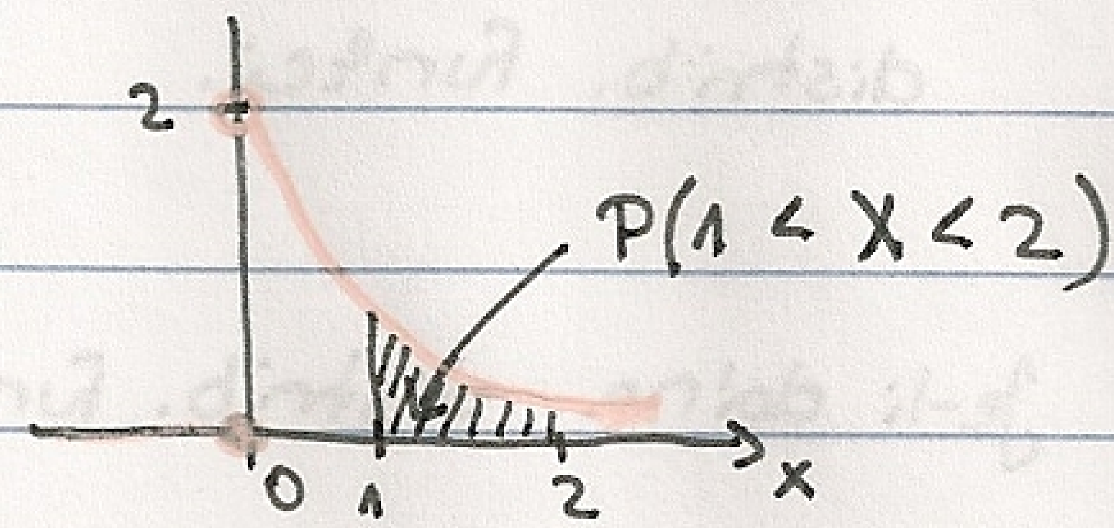
$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

2, je-li dána distrib. funkce $F(x)$, pak

$$f(x) = F'(x) \quad \text{pro "skoro všechna" } x \in \mathbb{R}$$

př.: 5. Náh. veličina X je dána hustotou pravděp.:

$$f(x) = \begin{cases} 0 & \dots\ x \le 0 \\ c \cdot e^{-2x} & \dots\ x > 0 \end{cases}$$



a, určete hustotu $\underline{c}$

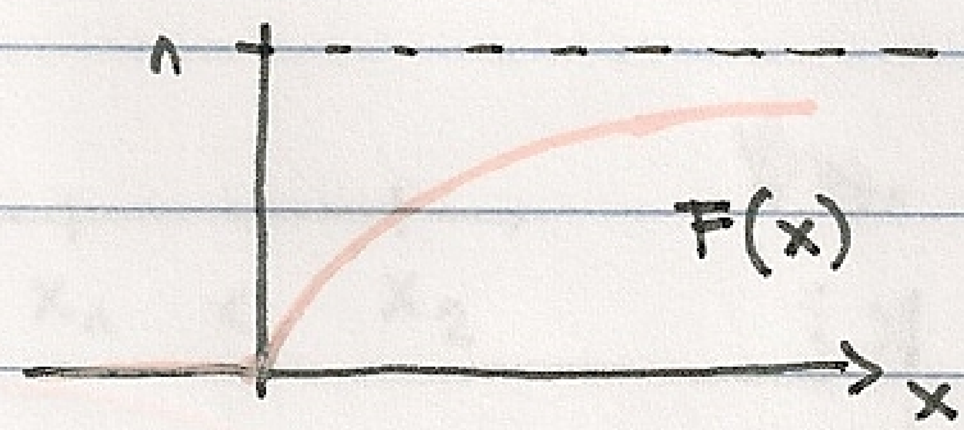
b, určete distrib. funkci $F(x)$ + graf

c, spočtěte $P(1 < X < 2)$

Řešení: a, $1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{\infty} c \cdot e^{-2x}\,dx = 0 + c \cdot \left[\frac{e^{-2x}}{-2}\right]_0^{\infty} =$

$$= c\left(0 - \left(-\frac{1}{2}\right)\right) = \frac{c}{2} \Rightarrow \boxed{c = 2}$$

b, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \int_{-\infty}^{x} 0 \cdot dx & \text{pro } x \le 0 \\ \int_{0}^{x} 2 \cdot e^{-2t} \cdot dt = 2 \cdot \left[\frac{e^{-2t}}{2}\right]_0^{x} = -e^{-2x} - (1) = \\ = 1 - e^{-2x} & \text{pro } x > 0 \end{cases}$



c, $P(1 < X < 2) = F(2) - F(1) = 1 - e^{-2 \cdot 2} - (1 - e^{-2 \cdot 1}) = e^{-2} - e^{-4} \doteq 0{,}135 - 0{,}012 = 0{,}1$

3. STŘEDNÍ HODNOTA, ROZPTYL

Základní charakteristikou **polohy** náh. veličiny X je tzv. **STŘEDNÍ HODNOTA** , píšeme $E(X)$.

Pro **diskrétní** náhodnou veličinu X definujeme:

$$\boxed{E(X) = \sum_i x_i \cdot P(x_i)}$$

kde sčítáme přes všechna i, pro která $P(x_i) > 0$;

Pro **spojitou** náhodnou veličinu X definujeme:

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx}$$

kde $f(x)$ je hustota ppsti.

Zákl. charakteristikami **variability** (proměnlivosti) náhodné veličiny X jsou **ROZPTYL** , píšeme $D(X)$, a **SMĚRODATNÁ ODCHYLKA**, píšeme $\sigma(X)$.

$$\boxed{D(X) = E(X - E(X))^2}$$

$$\boxed{\sigma(X) = \sqrt{D(X)}}$$

Pozn.:

- Roznásobením pravé strany definice $D(X)$ a úpravou dostaneme tzv. **výpočetní tvar rozptylu**:

$$D(X) = E(X^2) - E^2(X)$$

kde $E^2(X) = (E(X))^2$

a kde $E(X^2) = \sum_i x_i^2 \cdot P(x_i)$ pro diskr.náh.veličinu,

$E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx$ pro spoj.náh.veličinu.

- $E(X)$, $D(X)$, $\sigma(X)$ jsou reálná čísla (navíc: $D(X) \geq 0$, $\sigma(X) \geq 0$)

- Jsou-li X, Y náh. veličiny a c je reálná konstanta, pak:

$E(c \cdot X) = c \cdot E(X)$,

$E(X+Y) = E(X) + E(Y)$.

$D(cX) = c^2 D(X)$

$\sigma(cX) = |c|\sigma(X)$

- Jsou-li X, Y *nezávislé*, pak platí:

$$D(X+Y) = D(X) + D(Y)$$

(Pro X, Y *závislé* může nastat: ”<”, -”, ”>”.)

- Jsou-li $X_1, X_2, \ldots, X_n$ *nezávislé*, pak platí:

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i)$$

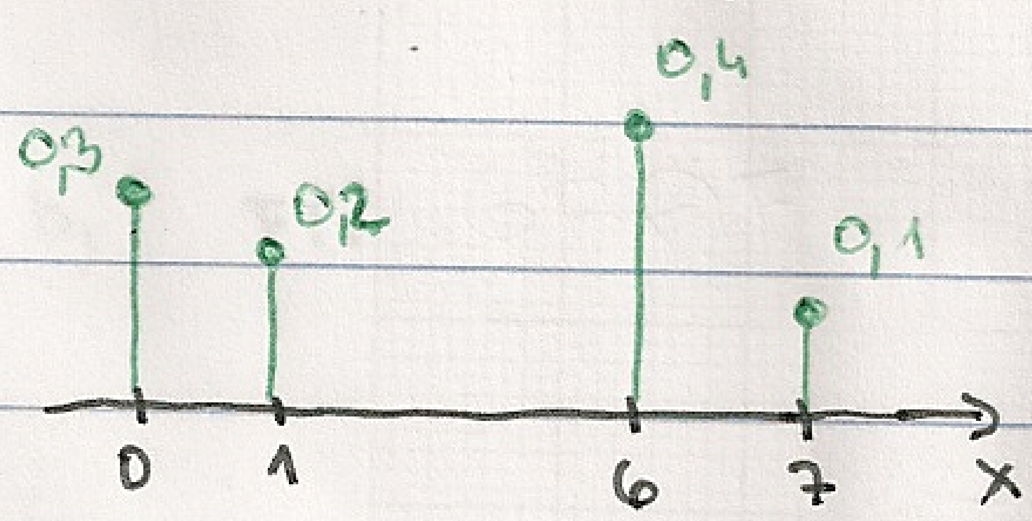
tj. "rozptyl součtu je roven součtu rozptylů"

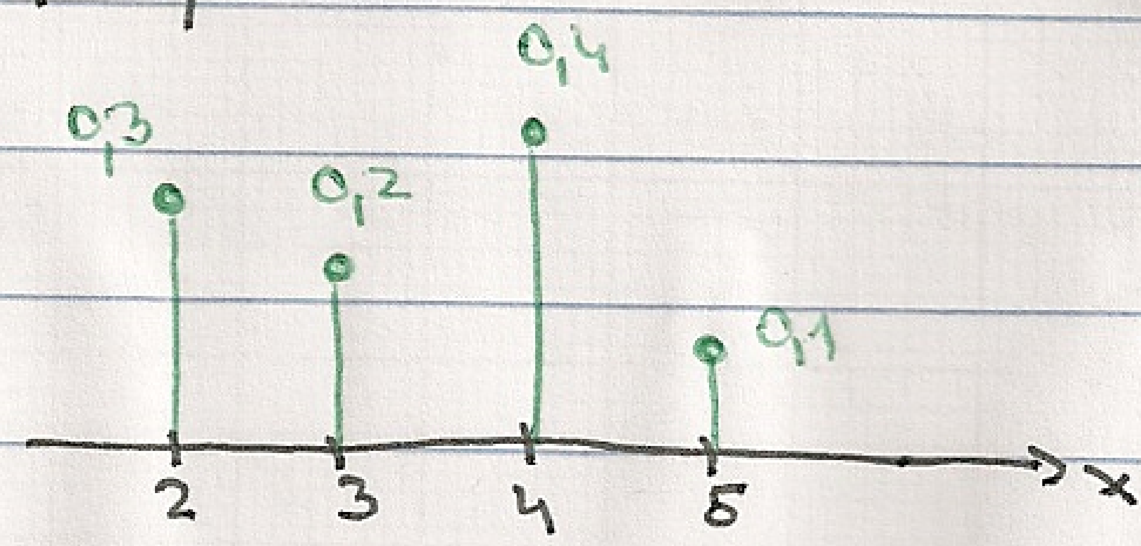
$$D(X) = E(X - E(X))^2 = E(X^2 - 2X\overbrace{E(X)}^{\text{konst.}} + \overbrace{E^2(X)}^{\text{konst.}}) =$$

$$= E(X^2) - 2\underbrace{E(X)E(X)}_{E^2(X)} + E^2(X) = E(X^2) - E^2(X)$$

př: Nechť ~~rozptyl~~ diskrétní náh. veličina X, resp. Y má, pravděp. funkci $P(x)$, resp. $P(y)$, určenou dle obrázku.

Vypočtěte: $E(x)$, $E(Y)$, $D(X)$, $D(Y)$,





Řešení: $E(x) = 0 \cdot 0{,}3 + 1 \cdot 0{,}2 + 6 \cdot 0{,}4 + 7 \cdot 0{,}1 = 3{,}3$

$D(x) = E(x^2) - E^2(x) = 19{,}5 - 3{,}3^2 = 8{,}61$

$0^2 \cdot 0{,}3 + 1^2 \cdot 0{,}2 + 6^2 \cdot 0{,}4 + 7^2 \cdot 0{,}1 = 19{,}5$

$\sigma(x) = \sqrt{8{,}61} \doteq 2{,}93$

$E(Y) = 2 \cdot 0{,}3 + 3 \cdot 0{,}2 + 4 \cdot 0{,}4 + 5 \cdot 0{,}1 = 3{,}3$

$D(Y) = 11{,}9 - 3{,}3^2 = 1{,}01$

$E(Y^2) = 2^2 \cdot 0{,}3 + 3^2 \cdot 0{,}2 + 4^2 \cdot 0{,}4 + 5^2 \cdot 0{,}1 = 11{,}9$

$\sigma(y) = \sqrt{1{,}01} \doteq 1{,}00$

$$D(X) > D(Y)$$

př: Spočtěte $E(x)$, $D(x)$ pro náh. veličinu z př. 5

Řešení:

$$E(x) = \int_0^{+\infty} x \cdot 2 e^{-2x}\,dx = \left[x \cdot 2 \frac{e^{-2x}}{-2}\right]_0^{+\infty} + \int_0^{+\infty} 1 \cdot e^{-2x}\,dx = \left[\frac{e^{-2x}}{-2}\right]_0^{\infty} =$$

$$= 0 - \left(-\frac{1}{2}\right) = \frac{1}{2}$$

$$D(x) = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4} \qquad E(x^2) = \int_0^{\infty} x^2 \cdot 2 \cdot e^{-2x}\,dx = \dots = \frac{1}{2}$$

per partes

K podrobnějšímu popisu náh.veličiny X se zavádí další charakteristiky.
Pro $k = 1, 2, \dots$ definujeme:

$\mu'_k(X) = E(X^k)$ ... k–tý obecný moment

$\mu_k(X) = E(X - E(X))^k$ ... k–tý centrální moment

Charakteristika

$$\alpha_3 = \frac{\mu_3(X)}{\sigma^3(X)}$$

se nazývá **koeficient šikmosti**.

Charakteristika

$$\alpha_4 = \frac{\mu_4(X)}{\sigma^4(X)} - \boxed{3}$$

aby Gaussovo rozdělení mělo koef. špič. 0

se nazývá **koeficient špičatosti**.

$$\boxed{\mu_3} = E(x - E(x))^3 =$$

$$= E\Big(x^3 - \underbrace{3E(x)}_{\text{konst}} \cdot x^2 + \underbrace{3E^2(x)}_{\text{konst}} \cdot x - E^3(x)\Big) =$$

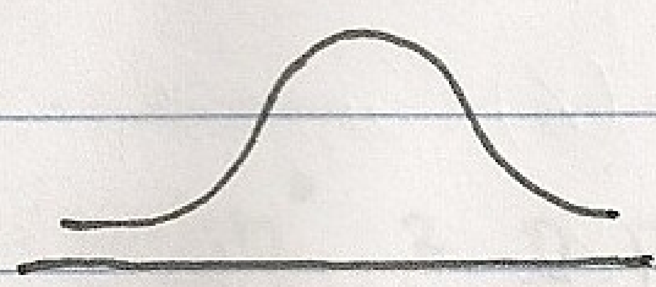
$$= E(x^3) - 3E(x)E(x^2) + \underbrace{3E^2(x)E(x) - E^3(x)}_{E^3(x)} =$$

$$\boxed{= E(x^3) - 3E(x) \cdot E(x^2) + 2E^3(x)}$$

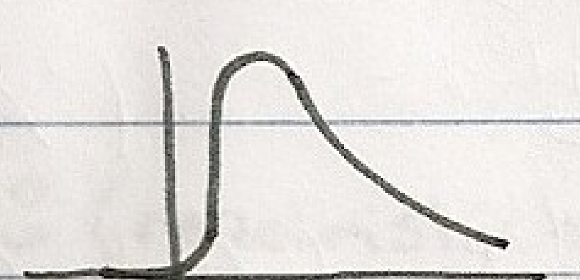
$$\boxed{\mu_4 = E(x^4) - 4E(x) \cdot E(x^3) + 6E^2(x) \cdot E(x^2) - 3E^4(x)}$$

$$F(x) = P(X \le x)$$

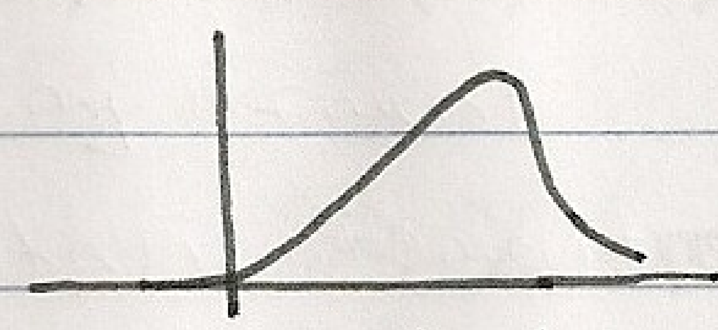
normální (Gaussovo) rozdělení



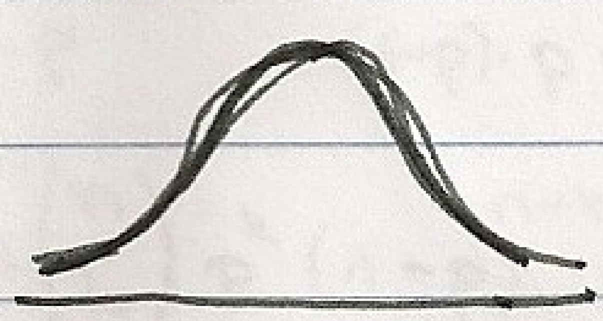
$\alpha_3 = 0$



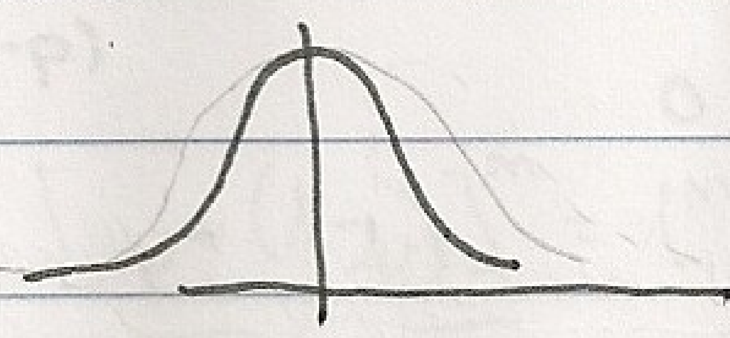
$\alpha_3 > 0$



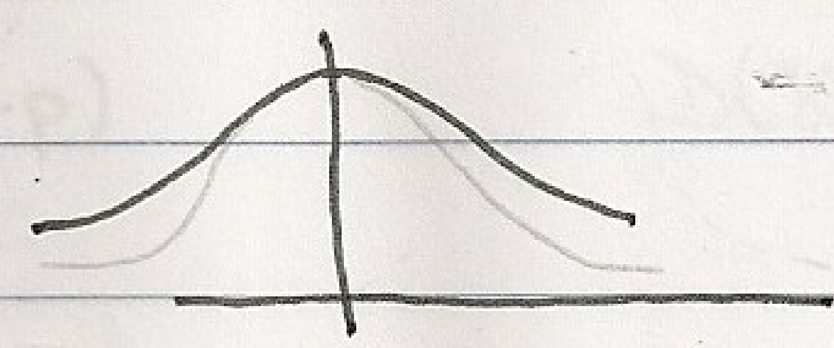
$\alpha_3 < 0$



$\alpha_4 = 0$



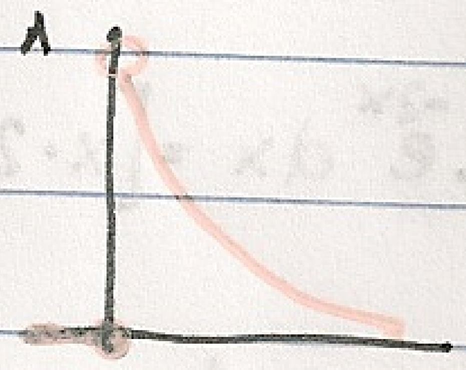
$\alpha_4 > 0$



$\alpha_4 < 0$

př: Pro náh. veličinu X s hustotou pravděp.

$$F(x) = \begin{cases} 0 & x \le 0 \\ e^{-x} & x > 0 \end{cases}$$



a, koef. ~~hustoty~~ šikmosti α_3

b, -||- ~~hustoty~~ špičatosti α_4

• koeficient šikmosti a špičatosti jsou sladné

Řešení: 7.10.

$$E(x) = \int_0^\infty x \cdot e^{-x}\,dx \overset{p.p.}{=} \underbrace{\left[-x e^{-x}\right]_0^\infty} + \int_0^\infty 1 \cdot e^{-x}\,dx =$$

$$= 0 + \left[\frac{e^{-x}}{-1}\right]_0^\infty = \lim_{x \to \infty} \frac{-x}{e^x} \overset{L.H.}{=} \lim \frac{-1}{e^x} = 0$$

$$= 0 - (-1) = 1$$

$$D(x) = E(x^2) - E^2(x) = 2! - 1^2 = 1$$

$$\Rightarrow \sigma(x) = 1$$

vypadne

$$E(x^m) = \int_0^\infty x^m \cdot e^{-x}\,dx = \cancel{\left[x^m \cdot \frac{e^{-x}}{-1}\right]_0^\infty} + m \underbrace{\int_0^\infty x^{m-1} \cdot e^{-x}\,dx}_{E(x^{m-1})}$$

dostáváme rekurzi ⇐

$$\boxed{E(x^m) = m!}$$

---

a, $\alpha_3 = \dfrac{\mu_3(x)}{\sigma^3(x)} = \dfrac{3! - 3\cdot 1\cdot 2! + 2\cdot 1^3}{1^3} = \boxed{2} > 0$ udává míru sešikmení

b, $\alpha_4 = \dfrac{\mu_4(x)}{\sigma^4(x)} - 3 = \dfrac{4! - 4\cdot 1\cdot 3! + 6\cdot 1^2\cdot 2! - 3\cdot 1^4}{1^4} - 3 = 9 - 3 = \boxed{6}$

- **HYPERGEOMETRICKÉ rozdělení** s param. $N, K, n \in \mathbf{N}, n \le N, K \le N$:

$$\boxed{P(k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}}$$

pro všechna $k \in \mathbf{N} \cup \{0\}$, taková, že $k \le K$, $0 \le n-k \le N-K$

Píšeme: $X \sim HG(N, K, n)$.

př.: Náh. výběr rozsahu $\underline{m}$ ze série $\underline{N}$ výrobků, z nichž $\underline{k}$ je vadných (tj. $p = \frac{k}{N}$ je podíl vadných), vybíráme postupně, zkontrolovaný výrobek nevracíme zpět (tzv. výběr bez vracení)

$\Rightarrow$ závislost $\quad \underline{p \neq \text{konst}}$

- binomické rozdělení není přesné, když je velký rozsah výběru

X ... počet vadných

$X \sim HG(N, k, m)$ $\qquad$ doporučení: $\boxed{\frac{m}{N} < 0{,}1}$

$$\left.\begin{array}{l} m \ll N \\ m-k \ll N-k \end{array}\right\} \Rightarrow p = \frac{k}{N} \doteq \text{konst.}$$

$\boxed{HG(N, k, m) \approx Bi(m \cdot p)}$, kde $p = \frac{k}{N}$ aproximace binom. rozdělením

př.: Mezi 10-ti výrobky jsou 3 vadné (výrobky s nižší životností). Náhodně vybereme 2 výrobky, označme X počet vadných (ze dvou vybraných). Určete pravděp. funkci $P(X)$, jestliže 1. vybraný prvek mezi zbylé
- a, vrátíme zpět (výběr s vracením povede k binom. rozděl.)
- b, nevracíme zpět (povede k hypergeom. rozděl.)

○○○●○
○●○●○

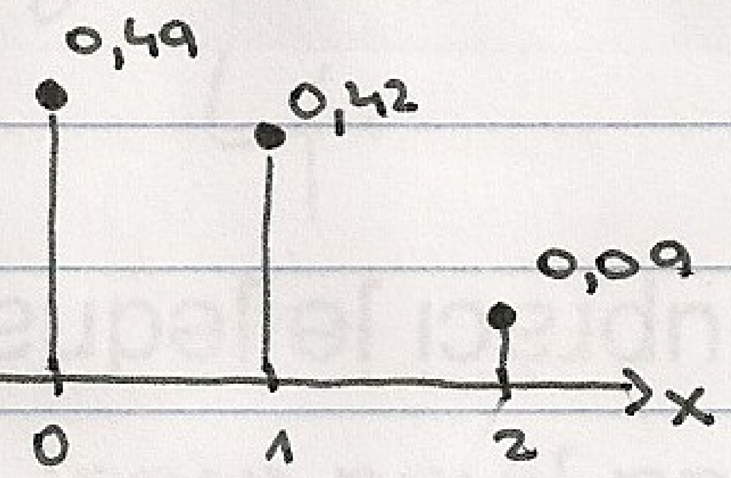
a, $X \sim Bi(2; 0{,}3)$

$0{,}3$ - pravděp., že najdu zmetek

$P(0) = 0{,}7^2 = 0{,}49$

$P(1) = 2 \cdot 0{,}7 \cdot 0{,}3 = 0{,}42$

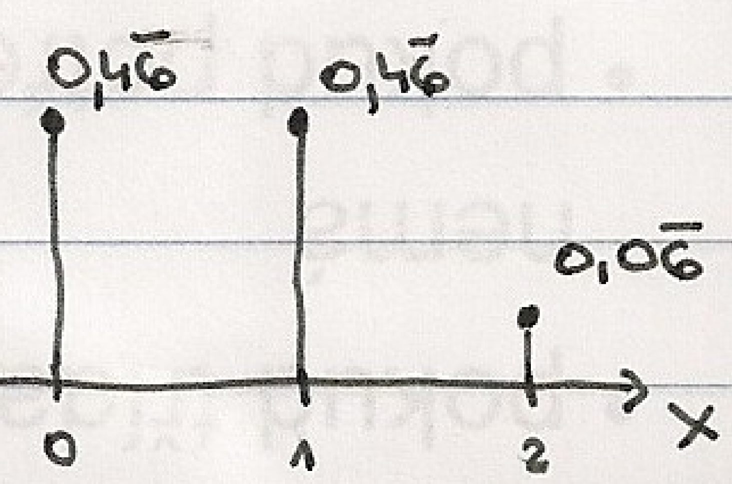
$P(2) = 0{,}3^2 = 0{,}09$



b, $X \sim HG(\overset{N}{10}, \overset{k}{3}, \overset{m}{2})$

$P(0) = \frac{7}{10} \cdot \frac{6}{9} = \frac{42}{90} = \frac{7}{15}$ $\qquad$ $P(2) = \frac{3}{10} \cdot \frac{2}{9} = \frac{6}{90} = \frac{1}{15}$

$P(1) = \frac{7}{10} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{15}$



v tomto případě nemůžeme aproximovat !

4. NĚKTERÁ DISKRÉTNÍ ROZDĚLENÍ

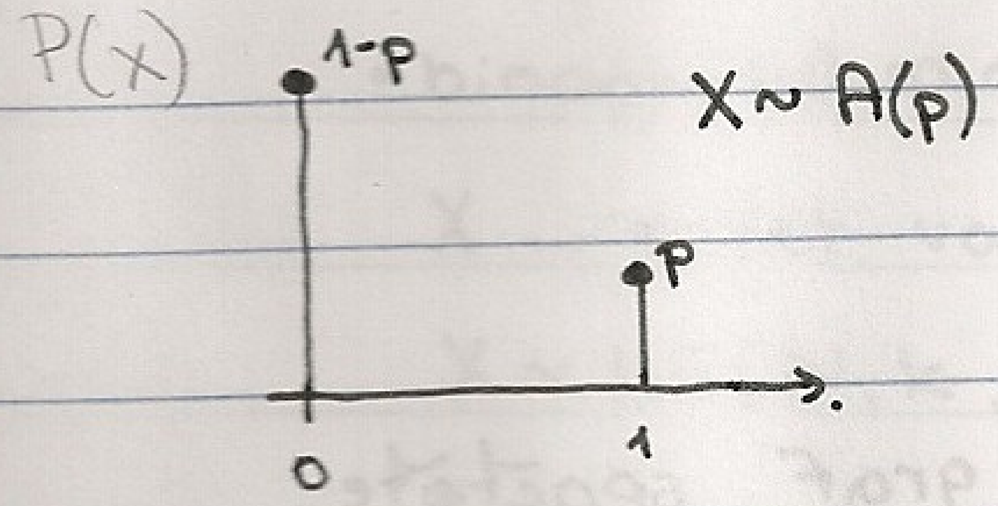
X ... **diskrétní náhodná veličina**

- **ALTERNATIVNÍ rozdělení** s parametrem $p \in (0,1)$:

X nabývá jen hodnot 0 nebo 1, přičemž $\boxed{P(0) = 1-p,\ P(1) = p}$

Píšeme: $X \sim A(p)$.

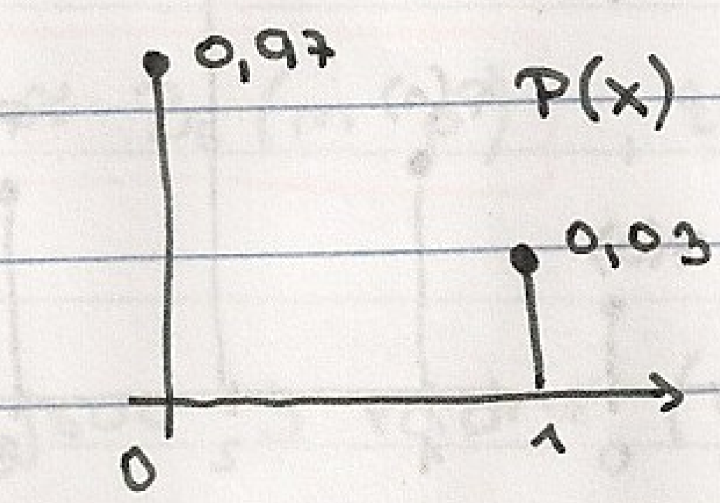
Výpočtem: $E(X) = p, \quad D(X) = p\,(1-p)$.



př: kontrola 1 výrobku ze série s 3% zmetkovitostí.

X ... počet zmetků

$X \sim A(0{,}03)$



$X \sim A(p)$

- hod mincí
  X ... počet líců $\left(p = \frac{1}{2}\right)$
- hod kostkou
  X ... počet 6 $\left(p = \frac{1}{6}\right)$
- výběr 1 výrobku z produkce s průměr. zmetkovitostí 2,5%
  X ... počet zmetků
  $p = 0{,}025$

$E(x) = 0\,(1-p) + 1 \cdot p = p \qquad E(x^2) = 0^2(1-p) + 1^2 \cdot p = p$

$D(x) = E(x^2) - E^2(x) = p - p^2 = p\,(1-p)$

- **BINOMICKÉ rozdělení** s parametry $n \in \mathrm{N},\ p \in (0,1)$:

X nabývá jen $0, 1, \ldots, n$ a platí: $\boxed{P(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pro } k = 0, 1, \ldots, n}$

Píšeme: $X \sim Bi(n,p)$

$Bi(n,p)$ je součtem n nezáv. veličin s rozd. $A(p)$, takže: $E(X) = n\,p, D(X) = n\,p\,(1-p)$

k - "úspěšných" z n pokusů (nezávislých)

$\circ \ \circ \ \bullet \ \ldots \ \circ \ \bullet \ \circ \qquad \bullet \ p$

$(1-p)\cdot(1-p)\cdot p \qquad (1-p)\cdot p\cdot(1-p) \qquad \circ \ (1-p)$

$\sum_{k=0}^{n} P(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = [\underbrace{p + (1-p)}_{1}]^n = 1^n = 1$

$A_i \ (i = 1, \ldots, m)$ nezávislé

$P\left(\bigcap_{i=1}^{m} A_i\right) = \prod_{i=1}^{m} P(A_i)$

Př. X ... počet líců ze 3 hodů mincí

$X \sim Bi\left(3, \frac{1}{2}\right) \quad E(X) = 3 \cdot \frac{1}{2} = 1{,}5$

$D(X) = 3 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = 0{,}75$

ze 7-mi max 1 vadný

př: Při statistické kontrole kvality je domluven přijímací plán (7,1), tj. je-li ze 7 kontrolovaných výrobků – bez vracení zpět – nejvýše 1 vadný, je zásilka odběratelem přijata.

Spočtěte pravděp. přijetí zásilky, budou-li mezi 60-ti výrobky

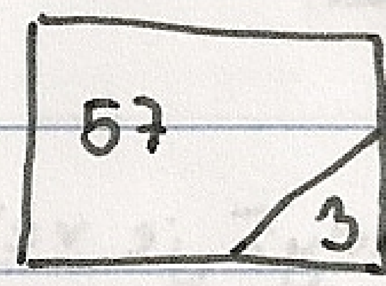
a, 3 zmetky b, 12 zmetků

-> povede na hypergeom. rozdělení (málo prvků)

X ... počet zmetků (ze 7 kontrolovaných)

a, $P(X \leq 1) =$

$X \sim HG(60, 3, 7)$



$= P(0) + P(1) =$

$$= \frac{\binom{57}{7} \cdot \binom{3}{0}}{\binom{60}{7}} + \frac{\binom{57}{6} \cdot \binom{3}{1}}{\binom{60}{7}} \doteq 0{,}966$$

b, $X \sim HG(60, 12, 7)$

$$P(X \leq 1) = \frac{\binom{48}{7} + \binom{48}{6} \cdot \binom{12}{1}}{\binom{60}{7}} \doteq \underline{0{,}572}$$

vysoká pravděp. přijetí nekvalitní zásilky

$\frac{12}{60} = 0{,}2$ (20% zmetkovitost)

⊳ Plán (7,1) je nevýhodný pro odběratele.

Pozn: $\frac{m}{N} = \frac{7}{60} > 0{,}1 \Rightarrow$ aproximace binom. rozdělením by byla nepřesná

• **POISSONOVO rozdělení** s parametrem $\lambda > 0$:
X může nabývat jen hodnot $k = 0, 1, 2, \ldots$ a platí

$$\boxed{P(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{pro} \quad k = 0, 1, 2, \ldots .}$$

Píšeme: $X \sim Po(\lambda)$

Výpočtem: $E(X) = \lambda,\ D(X) = \lambda$

Jsou-li $X_i \sim Po(\lambda_i),\ i = 1, 2, \ldots, n$ **nezávislé** náh. veličiny, $X = \sum_{i=1}^{n} X_i$,

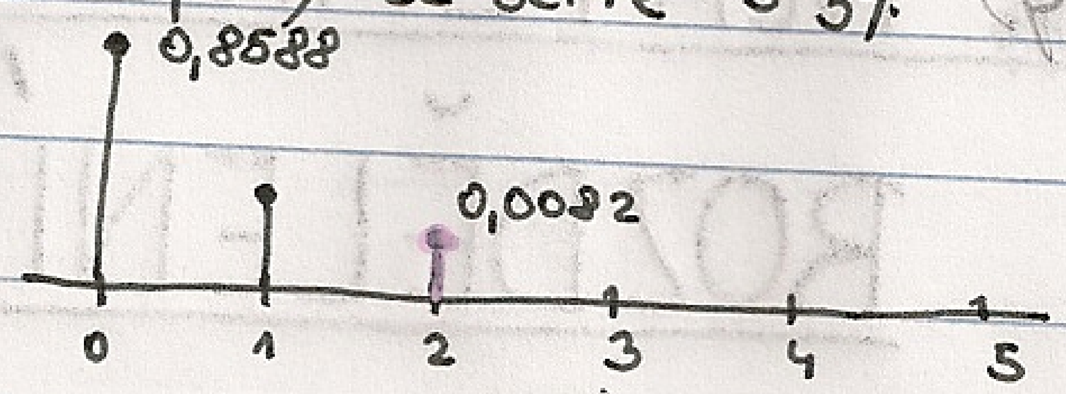
pak $X \sim Po(\lambda)$, kde $\lambda = \sum_{i=1}^{n} \lambda_i$.

př: kontrola 5 náh. vybraných výrobků (s vracením zpět) ze série s 3% zmetkovitostí. ($p = 0{,}03$)

X ... počet zmetků (v pětici)

$X \sim Bi(5;\ 0{,}03)$

např.: $P(2) = \binom{5}{2} \cdot 0{,}03^2 \cdot 0{,}97^2 \doteq 0{,}0082$



Pozn.: $X_i \sim A(p)$ $(i = 1 \dots m)$ nezávislé $\Rightarrow X = \sum_{i=1}^{m} X_i \sim Bi(m, p)$

$E(X) = m \cdot p$

$D(X) = m \cdot p \cdot (1-p)$

př: Nechť $X \sim Bi(\overset{m}{3};\ \overset{p}{0{,}6})$. Určete pravděp. funkci $P(x)$ + graf, spočtěte střední hodnotu $E(X)$

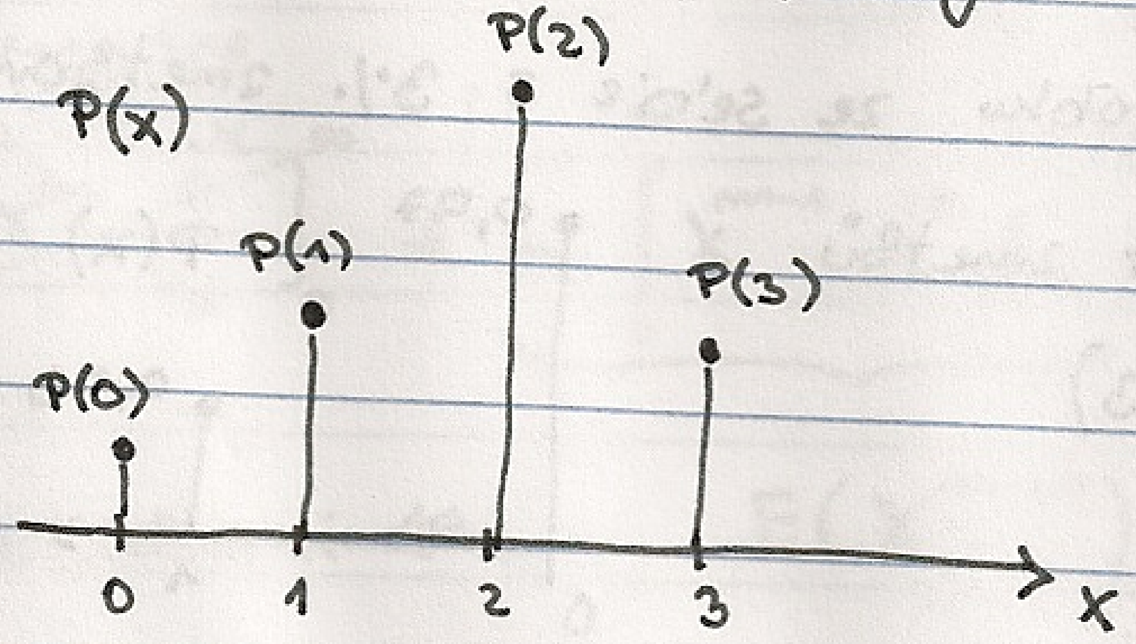
Řešení:

$P(0) = 0{,}4^3 = 0{,}064$

$P(1) = \binom{3}{1} \cdot 0{,}6 \cdot 0{,}4^2 = 0{,}288$

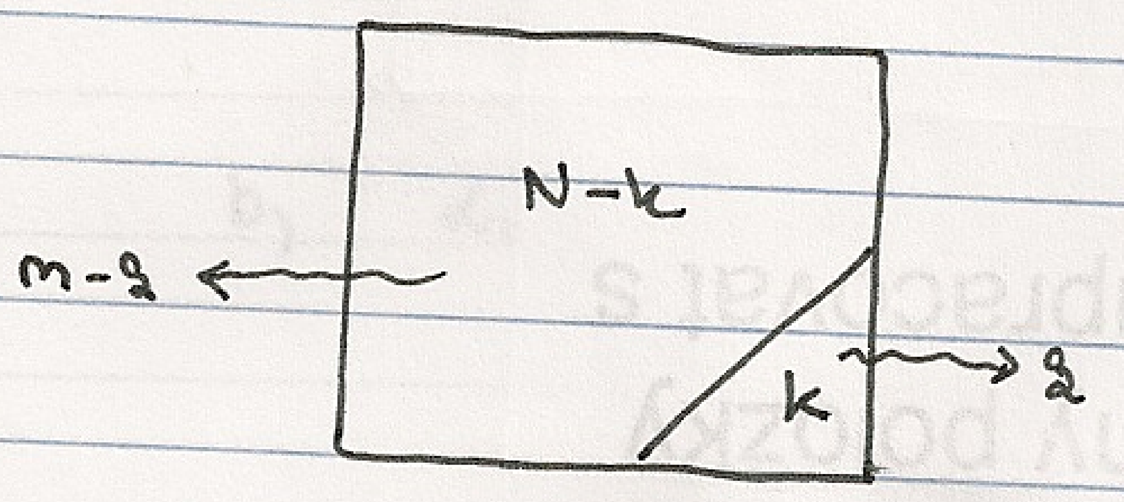
$P(2) = \binom{3}{2} \cdot 0{,}6^2 \cdot 0{,}4 = 0{,}432$

$P(3) = 0{,}6^3 = 0{,}216$



$E(x) = 3 \cdot 0{,}6 = 1{,}8$

OBECNĚ:



N - prvků

k má nějakou vlastnost, zbytek ne

$0 \le k \le m$

dle klasické def. pravděp.

$$\frac{\binom{k}{q} \cdot \binom{N-k}{m-q}}{\binom{N}{m}}$$

počet příznivých jevů

Binomická věta: $1 = 1^n = [p + (1-p)]^n = \sum$

Pozn.: 1, $\sum_{k=0}^{n} P(k) = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$

Taylorova řada

2, $E(X) = \sum_{k=0}^{\infty} k \cdot P(k) = \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$

$= \lambda \cdot e^{-\lambda} \cdot \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \boxed{\lambda}$

$\| \; e^{\lambda}$ (z Taylora)

Poisson. rozdělení lze použít k aproximaci binomického rozdělení: pro $\boxed{n \geq 30}$ a $\boxed{p \leq 0,1}$

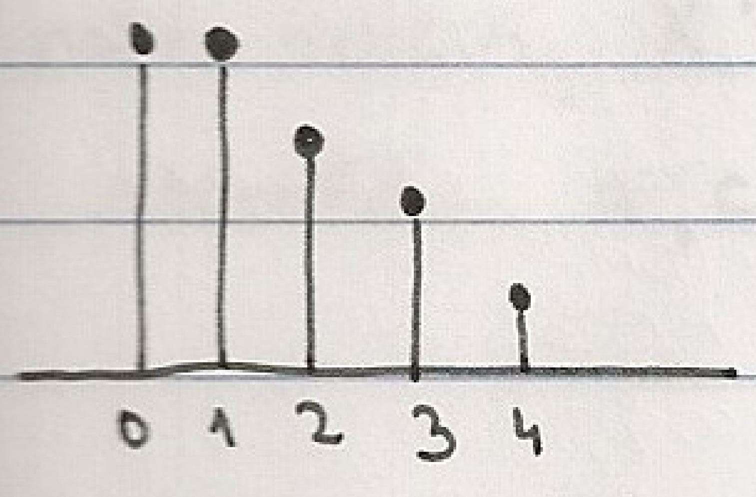
$$\boxed{Bi(n,p) \approx Po(\lambda), \quad \text{kde } \lambda = n \cdot p}$$

Funkční hodnoty distribuční funkce $F(x)$ Poissonova rozdělení bývají **tabelovány** pro některá $\lambda \leq 10$.

Pro $\lambda \geq 9$ používáme aproximaci normálním rozdělením - bude později.

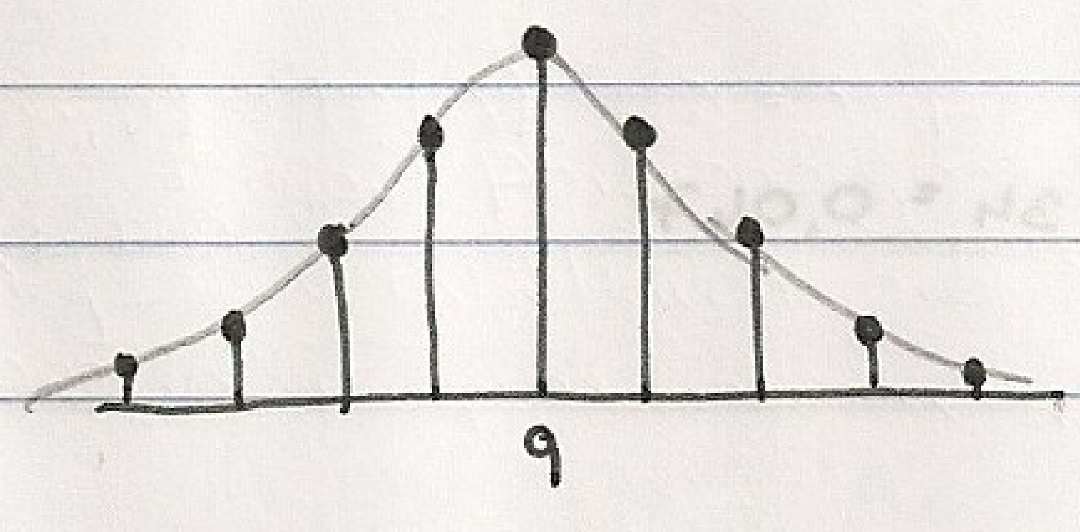
Po(1)

$\lambda = 1$



Po(3)

$\lambda = 3$



př: průměrná zmetkovitost produkce (3 tisíce výrobků => téměř konst. zmetkovitost i když vracíme i když ne). Vybereme a zkontrolujeme 75 výrobků. Spočtěte (s přesností na 3 des. místa)

a) $P(0)$ (ve vzorku není zmetek)

b) $P(X \leq 1)$ (nejvýše 1 zmetek)

d, $P(4)$ (právě 4 vadné)

<u>Řešení:</u>

a, $P(0) = 0{,}98^{75} \doteq 0{,}220$

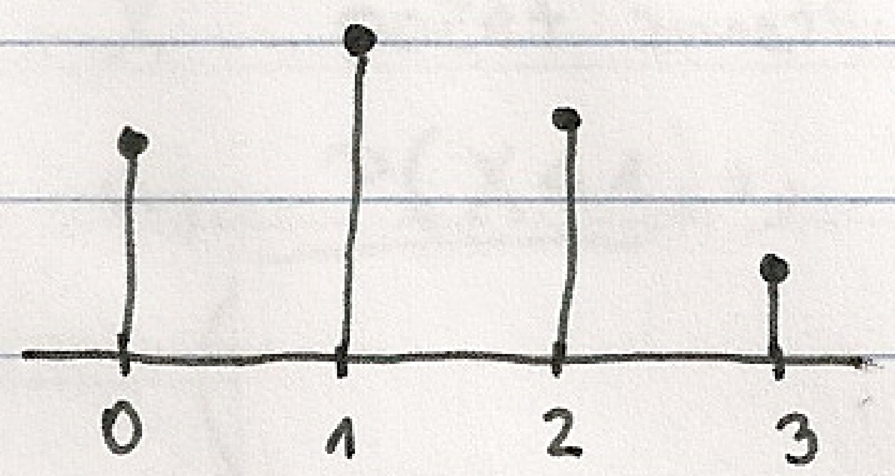
$P(2) = \binom{75}{2} 0{,}02^2 \cdot 0{,}98^{73} = 0{,}254$

$P(3) = \binom{75}{3} 0{,}02^3 \cdot 0{,}98^{72} \doteq 0{,}126$

b, $P(X \leq 1) = P(0) + \underbrace{P(1)}_{\binom{75}{1} \cdot 0{,}02^1 \cdot 0{,}98^{74} \doteq 0{,}536} \doteq 0{,}556$

c, $P(X \geq 4) = 1 - P(X \leq 3) = 1 - (P(0) + P(1) + P(2) + P(3)) =$

$= 1 - (0{,}220 + 0{,}556 + 0{,}254 + 0{,}126) = 1 - 0{,}936 \doteq \boxed{0{,}063}$



d, $P(4) = \binom{75}{4} \cdot 0{,}02^4 \cdot 0{,}98^{71} =$

př: Spočtěte předchozí příklad použitím aproximace Poissonovým rozdělením

$n = 75 > 30$

$p = 0{,}02 < 0{,}1$

$\Bigg\} \quad X \approx Po(1{,}5)$

$\lambda = 75 \cdot 0{,}02 = 1{,}5$

a, $P(0) = F(0) = 0{,}223$ b, $P(X \leq 1) = F(1) = 0{,}558$

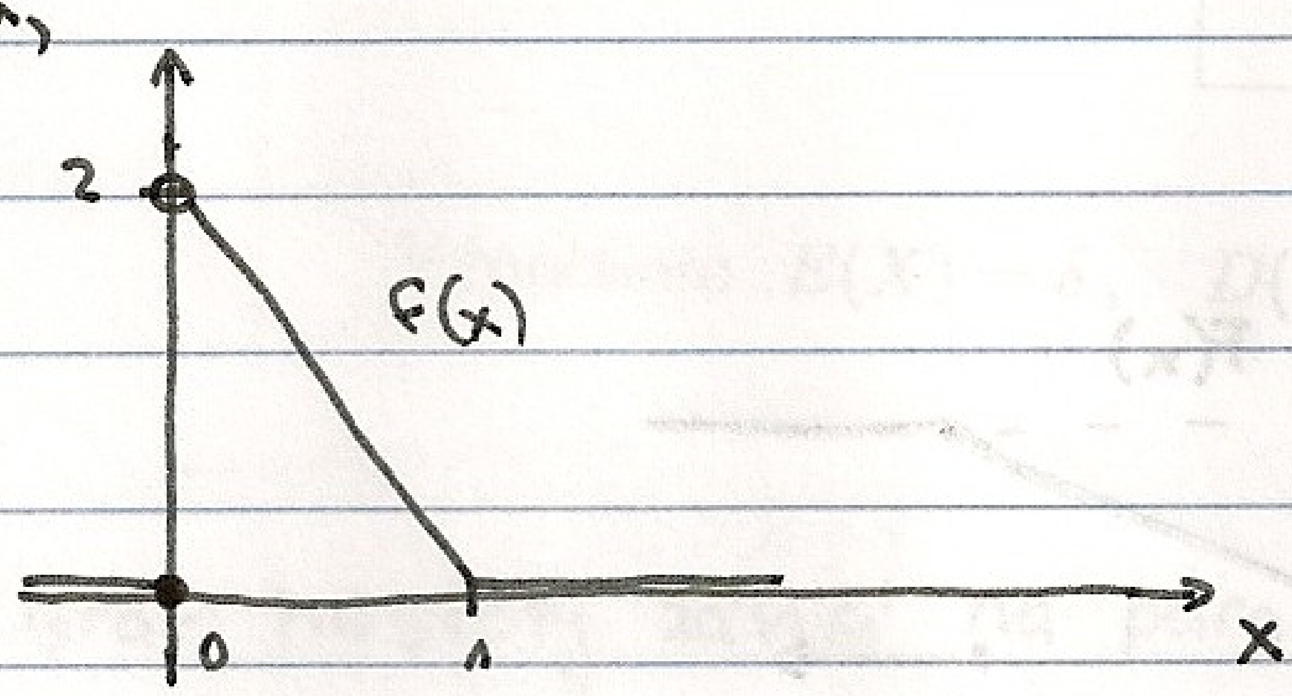
c, $P(X \geq 4) = 1 - \underbrace{P(X \leq 3)}_{F(3)} = 1 - 0{,}934 = 0{,}066$

d, $P(4) = F(4) - F(3) = 0{,}981 - 0{,}934 = 0{,}047$

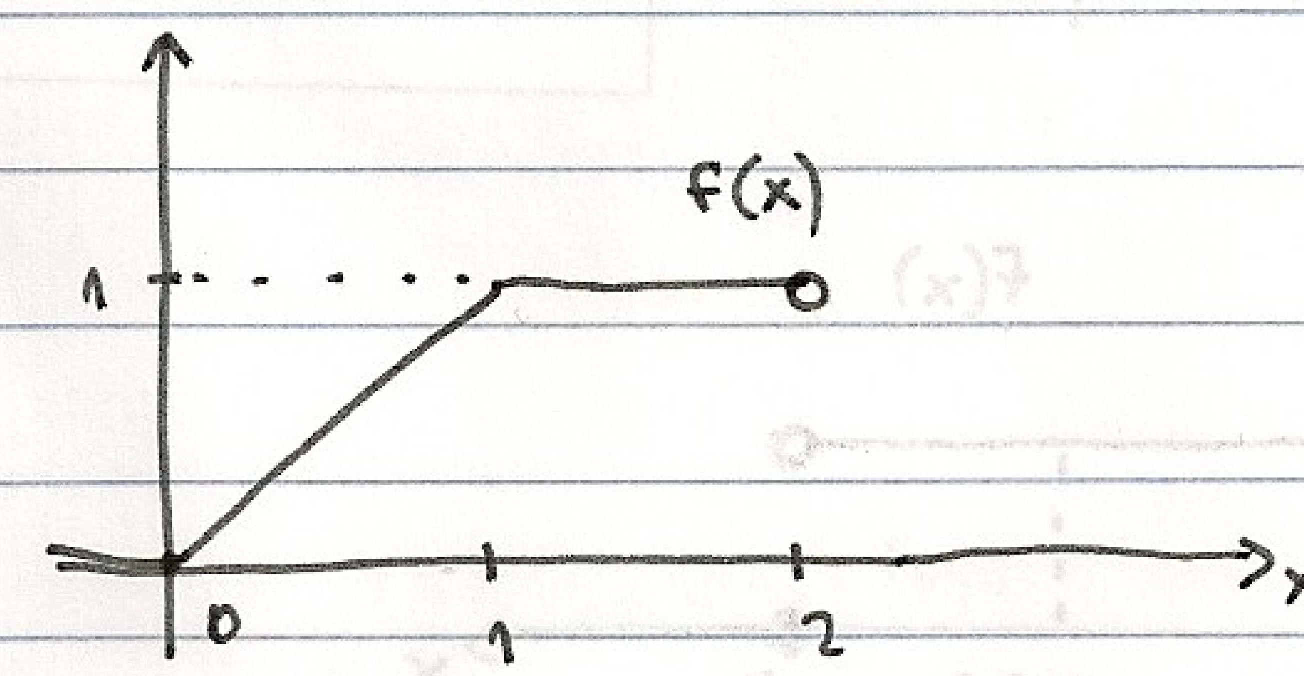
5. NĚKTERÁ SPOJITÁ ROZDĚLENÍ

př:

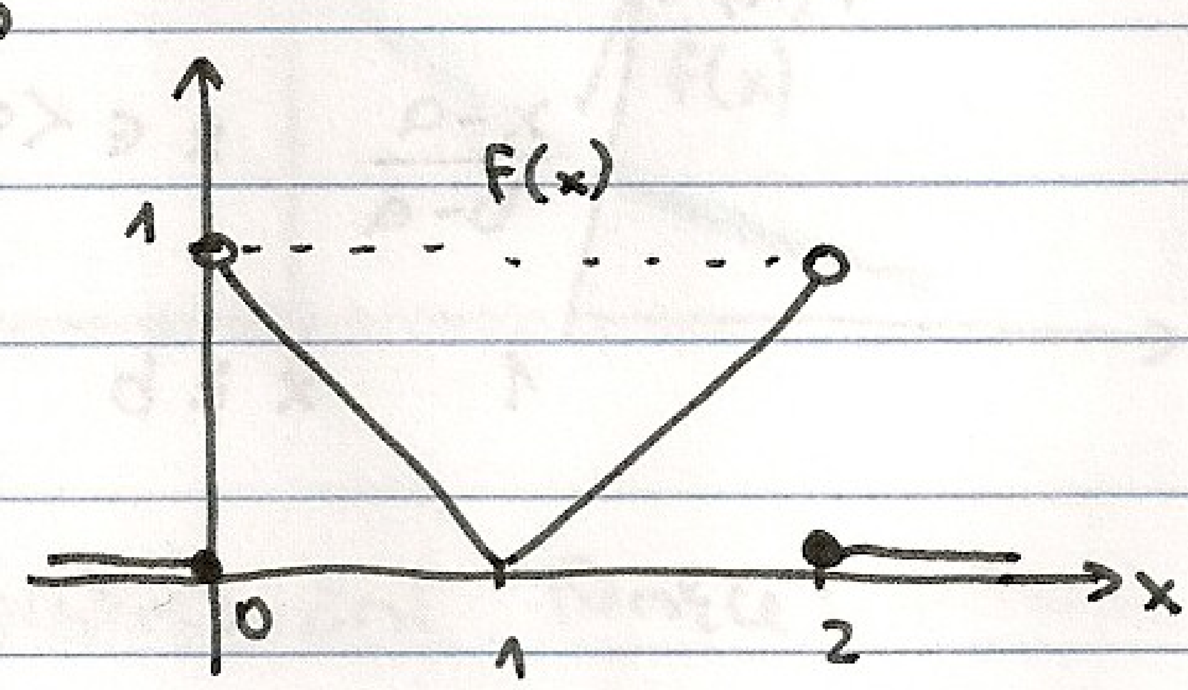
a,



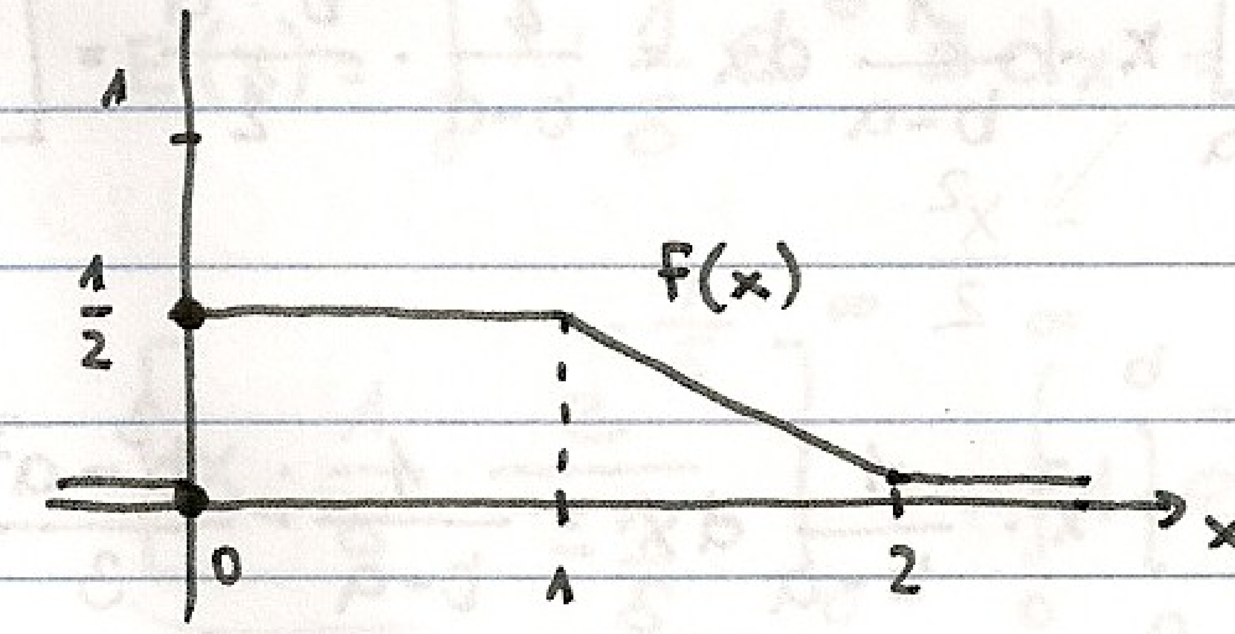
b,



c,



d,

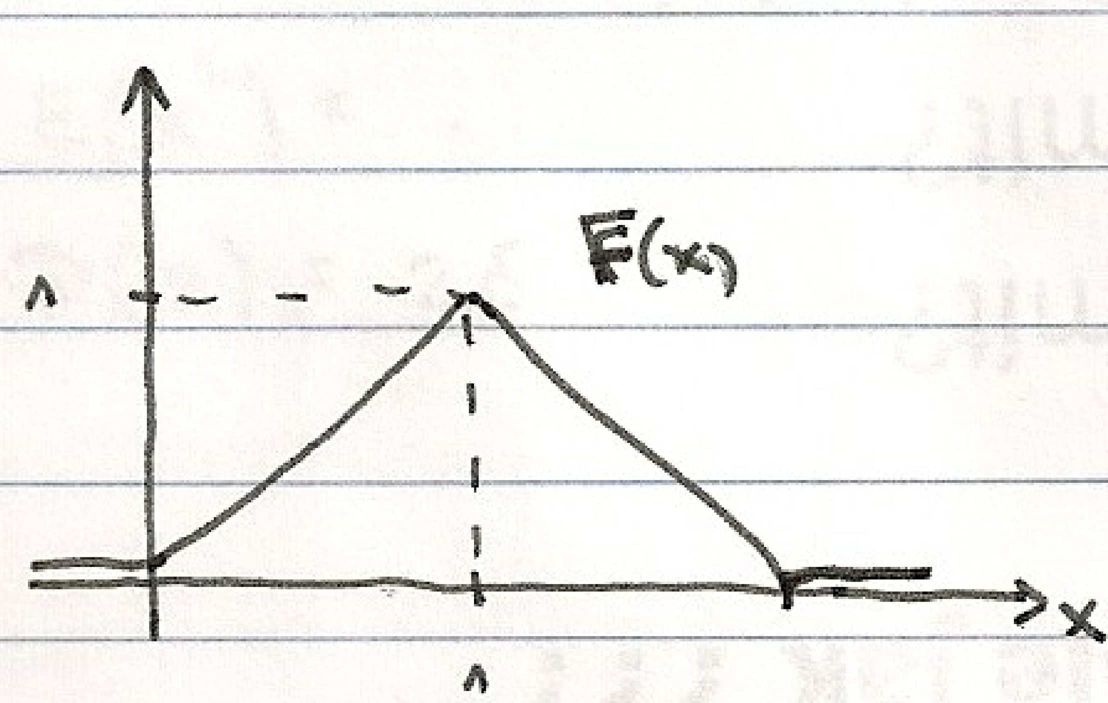


Která z těchto funkcí, jejíž graf je na obr., může být hustotou pravděp. nah. veličiny X?

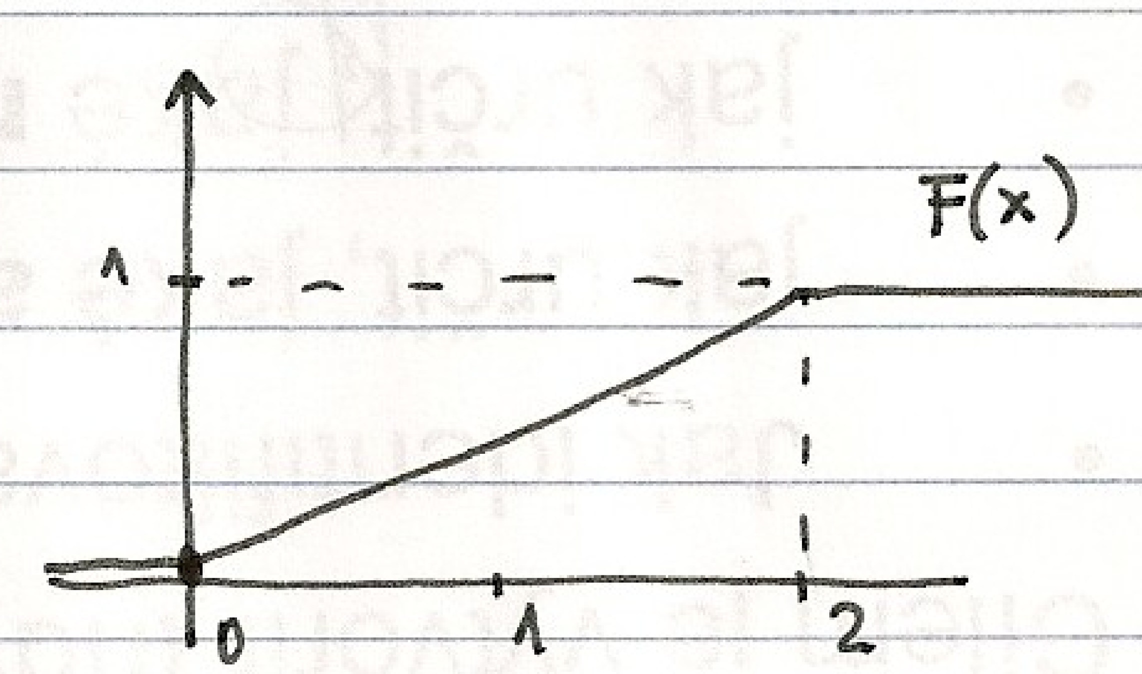
odpověď: a, c

př:

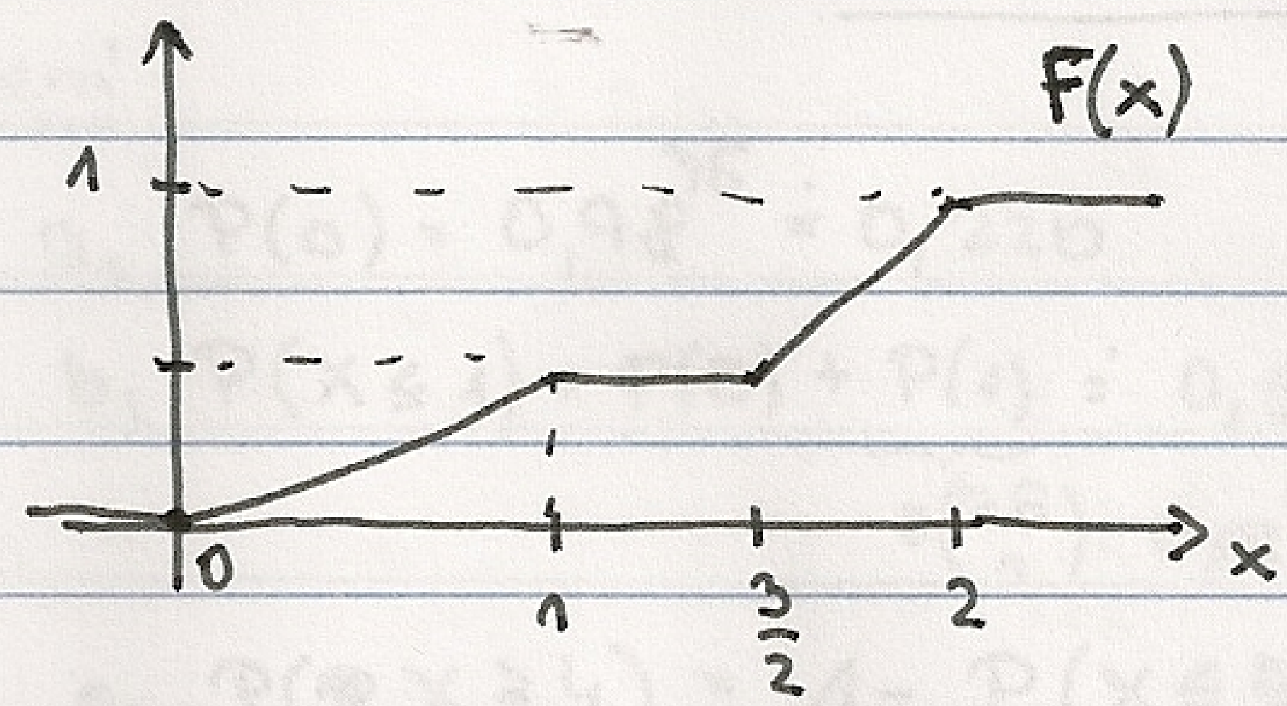
a,



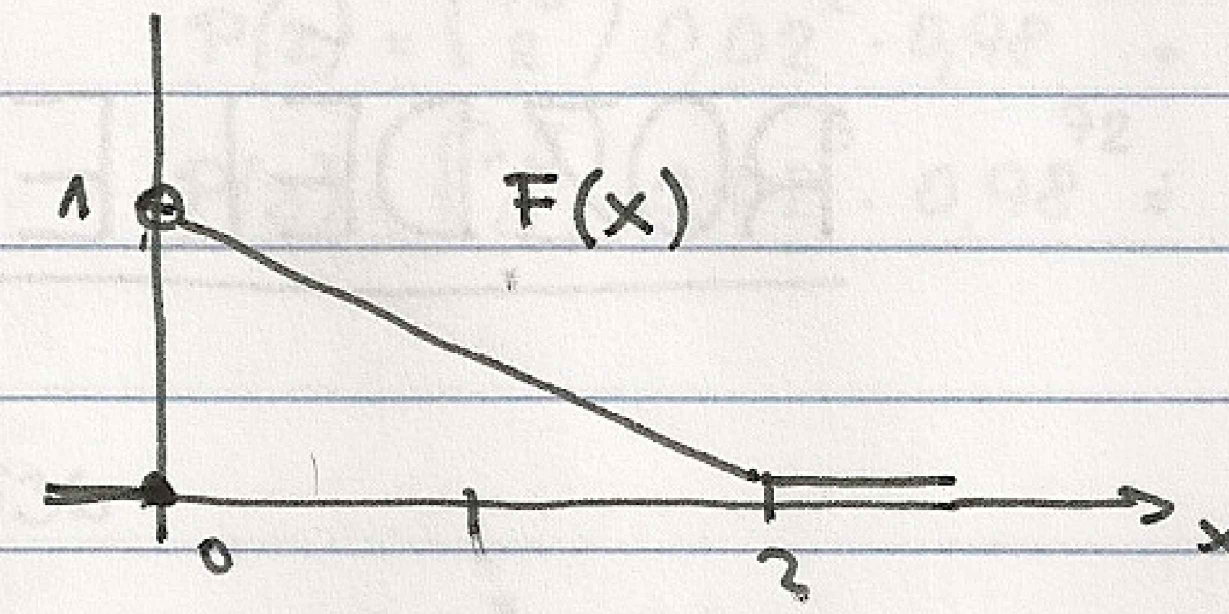
b,



c,



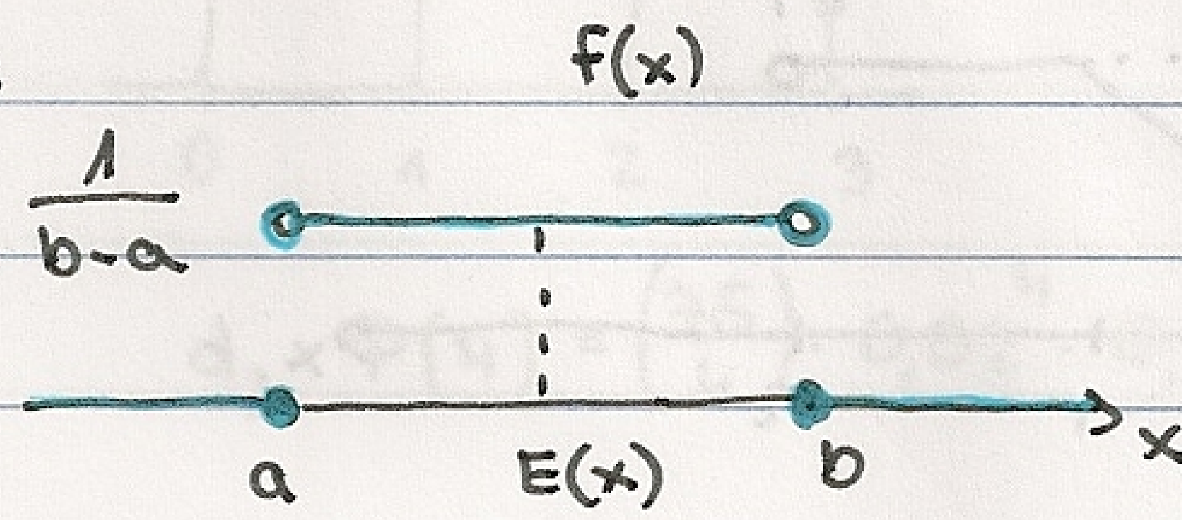
d,

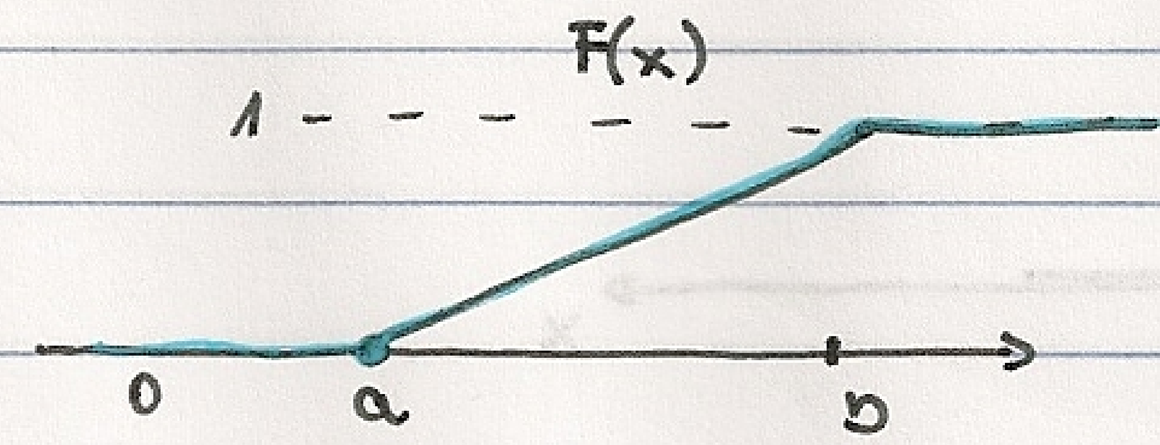


která z těchto funkcí může být distrib. funkcí náh. veličiny X ?

odpověď: b, c

př:





$$E(X) = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \boxed{\frac{a+b}{2}}$$

$$\frac{x^2}{2}$$

$$F(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & x \in \langle a, b \rangle \\ 1 & x \geq b \end{cases}$$

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} = \frac{a^2+ab+b^2}{3}$$

$$D(X) = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

X ... **spojitá náhodná veličina**

ROVNOMĚRNÉ rozdělení na intervalu (a, b) :
X má hustotu ppsti

$$\boxed{f(x) = \begin{cases} \dfrac{1}{b-a} & \text{pro } x \in (a, b) \\ 0 & \text{jinde.} \end{cases}}$$

Píšeme: $X \sim R(a, b)$

Výpočtem: $E(X) = \frac{a+b}{2}$, $D(X) = \frac{(b-a)^2}{12}$.

EXPONENCIÁLNÍ rozdělení s parametrem $\delta > 0$:
X má hustotu

$$f(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0] \\ \frac{1}{\delta} e^{-\frac{x}{\delta}} & \text{pro } x \in (0, +\infty) \end{cases}$$

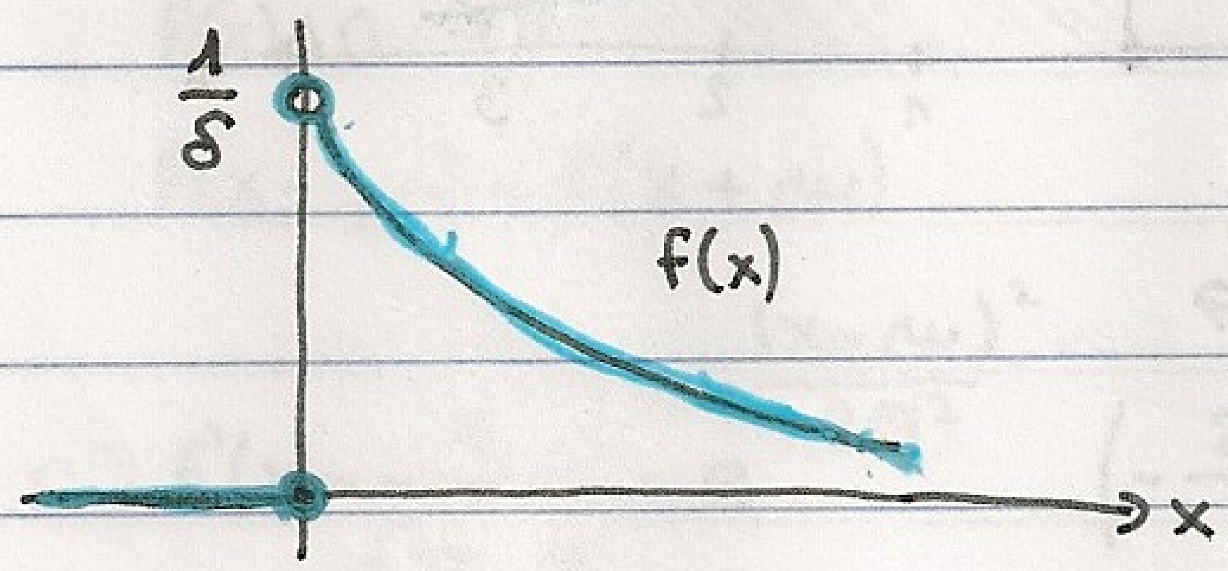
Píšeme: $X \sim Exp(\delta)$.

Distribuční funkce :

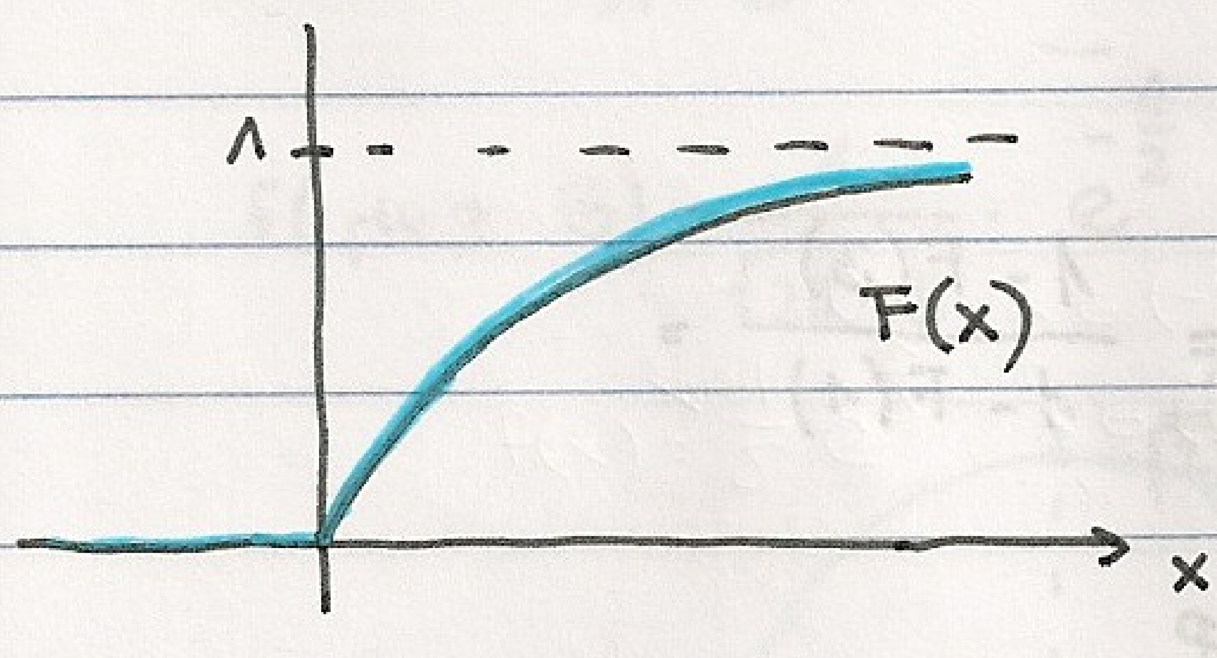
$$F(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0] \\ 1 - e^{-\frac{x}{\delta}} & \text{pro } x \in (0, +\infty) \end{cases}$$

Výpočtem: $E(X) = \delta$, $D(X) = \delta^2$.

- Graf hustoty závisí na parametru δ.



- Distribuční funkce:



$$E(x) = \int_0^\infty x \cdot \frac{1}{\delta} e^{-\frac{x}{\delta}} \, dx =$$

$$= \underbrace{\left[x \cdot \frac{1}{\delta} \cdot \frac{e^{-\frac{x}{\delta}}}{-\frac{1}{\delta}} \right]_0^\infty}_{\to 0} + \int_0^\infty 1 \cdot e^{-\frac{x}{\delta}} \, dx =$$

$$\left(\lim_{x \to \infty} \frac{-x}{e^{\frac{x}{\delta}}} = \lim_{x \to \infty} \frac{-1}{\frac{1}{\delta} \cdot e^{\frac{x}{\delta}}} = \underline{\underline{0}} \right)$$

$$= 0 - 0 + \left[\frac{e^{-\frac{x}{\delta}}}{-\frac{1}{\delta}} \right]_0^\infty = 0 - \left(\frac{1}{-\frac{1}{\delta}} \right) = \underline{\underline{\delta}}$$

$$E(X^2) = \dots = \underline{\underline{2\delta^2}}$$

$$D(X) = 2\delta^2 - \delta^2 = \underline{\underline{\delta^2}}$$

př.: Nechť $X \sim Exp(2)$

(např.: X .. životnost výrobků s průměrnou životností 2 roky)

Spočtěte: a) $P(X<3)$

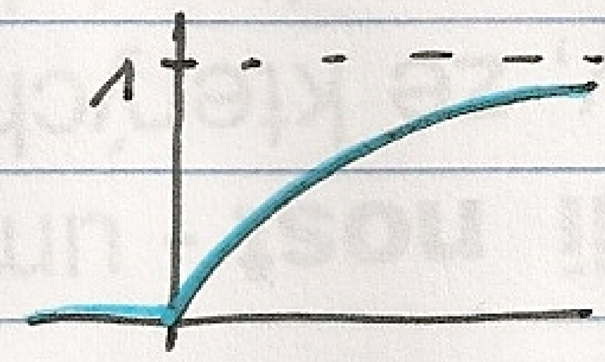
b) $P(X>3)$

c) $P(1<X<2)$

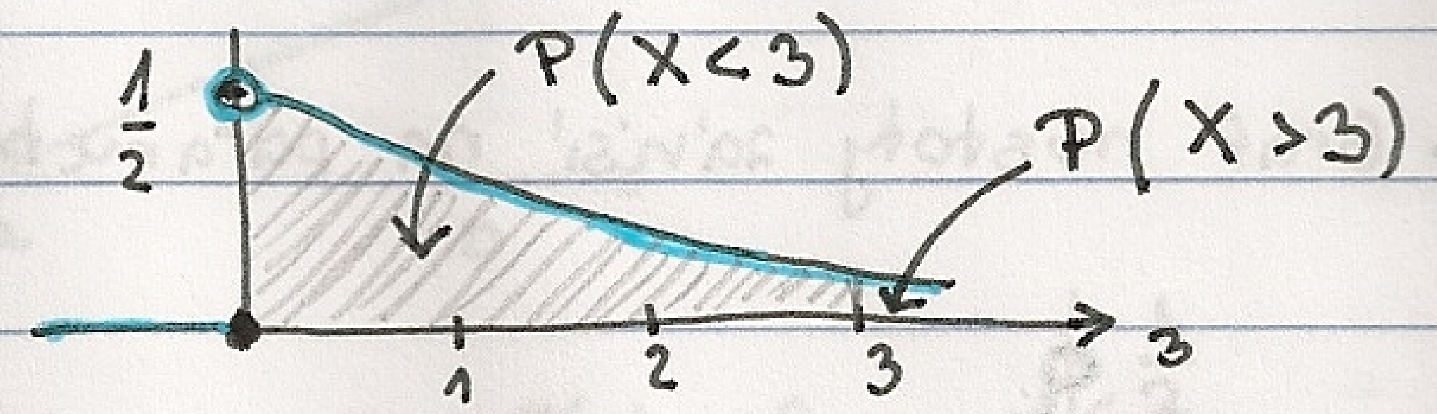
d) $P(X>2)$

e) $P(X>3 \mid X>1)$

Řešení: $F(x) = \begin{cases} 0 & x \leq 2 \\ 1-e^{-\frac{2}{x}} & x>0 \end{cases}$



$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2}e^{-\frac{x}{2}} & x>0 \end{cases}$



a) $P(X<3) = P(X \leq 3) = F(3) = 1 - e^{-\frac{3}{2}} \doteq 0{,}777$

b) $P(X>3) = 1 - \underbrace{P(X \leq 3)}_{F(3)} \doteq 0{,}223$

c) $P(1<X<2) = F(2) - F(1) = \ldots = e^{-\frac{1}{2}} - e^{-1} \doteq 0{,}239$

d) $P(X>2) = 1 - F(2) = e^{-1} \doteq 0{,}368$

podmíněná pravděp.

e) $P(X>3 \mid X>1) = \frac{P(A \cap B)}{P(B)} = \frac{P(X>3)}{P(X>1)} = \frac{1-F(3)}{1-F(1)} =$

$= \frac{1-(1-e^{-\frac{3}{2}})}{1-(1-e^{-\frac{1}{2}})} = \frac{e^{-\frac{3}{2}}}{e^{-\frac{1}{2}}} = e^{-1} \doteq 0{,}368$

př.: Nechť $X \sim Exp(\delta)$, $\delta > 0$, nechť $x_0 > 0$

$|\Delta > 0|$

$P(X > x_0 + \Delta \mid X > x_0) = \frac{P(X > x_0 + \Delta)}{P(X > x_0)} = \frac{1 - F(x_0+\Delta)}{1 - F(x_0)} =$

$$= \frac{1-\left(1-e^{-\frac{x_0+\Delta}{\delta}}\right)}{1-\left(1-e^{-\frac{x_0}{\delta}}\right)} = \frac{\cancel{e^{-\frac{x_0}{\delta}}}\cdot e^{-\frac{\Delta}{\delta}}}{\cancel{e^{-\frac{x_0}{\delta}}}} = 1-\underbrace{\left(1-e^{-\frac{\Delta}{\delta}}\right)}_{F\left(-\frac{\Delta}{\delta}\right)} = P(X>\Delta)$$

NORMÁLNÍ rozdělení s parametry $\mu \in \mathrm{R}$, $\sigma > 0$:
X má pro $x \in (-\infty, +\infty)$ hustotu

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

(její graf je symetrický kolem přímky $x = \mu$)

Píšeme: $X \sim N(\mu, \sigma^2)$

Výpočtem lze ověřit: $E(X) = \mu$, $D(X) = \sigma^2$.

Používá se též název **GAUSSOVO rozdělení**.

$f(x) > 0$

$f(x-\mu) = f(x+\mu)$

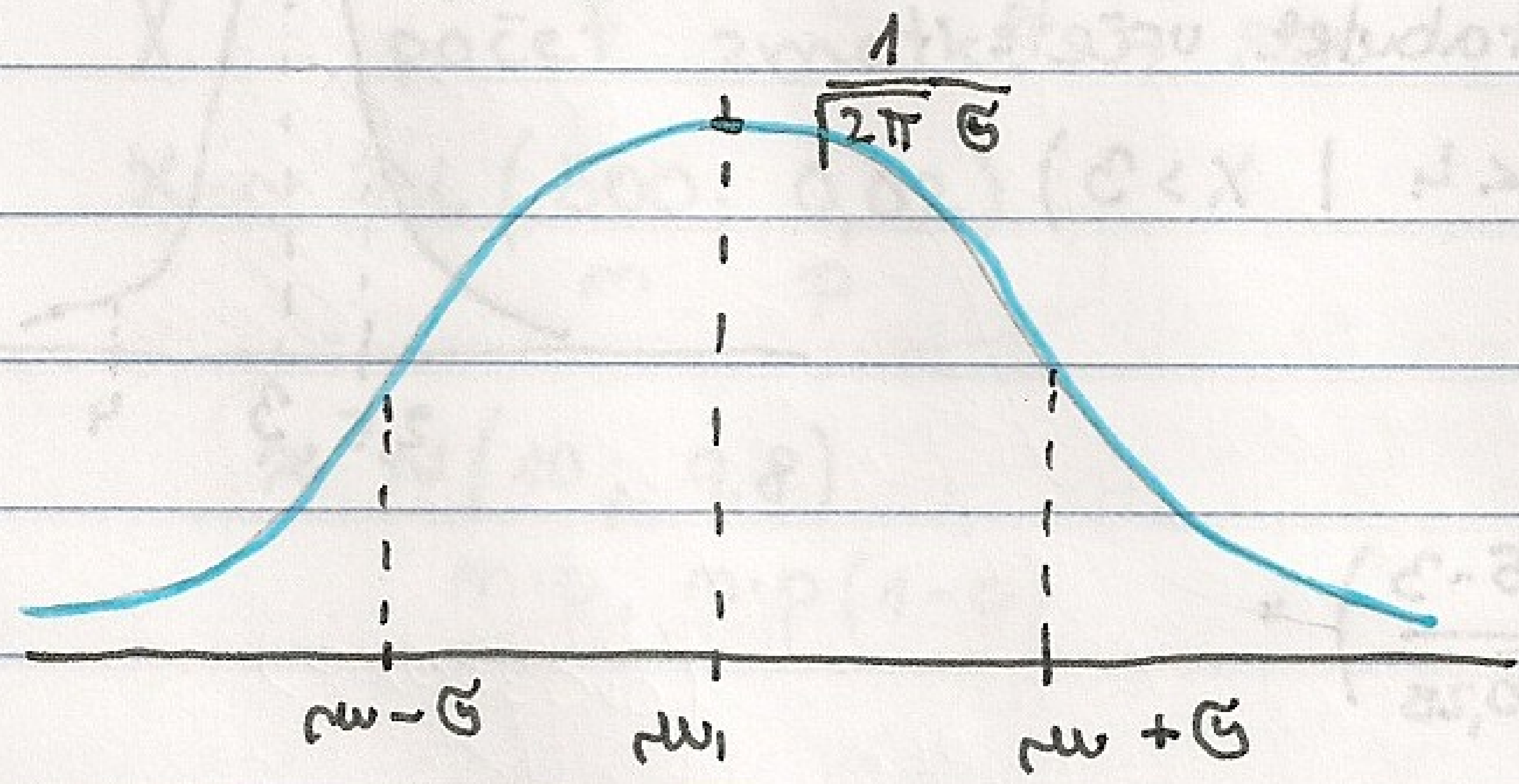
$$f'(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\frac{2(x-\mu)}{2\sigma^2}\right)$$

$f'(x) = 0 \iff x = \mu$

$f''(x) = 0 \iff x = \mu \pm \sigma$ inflexe

$$f(\mu) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \quad \dots \text{ max}$$

$$f(\mu \pm \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\cdot e^{-\frac{1}{2}} = 0{,}61\cdot f(\mu)$$

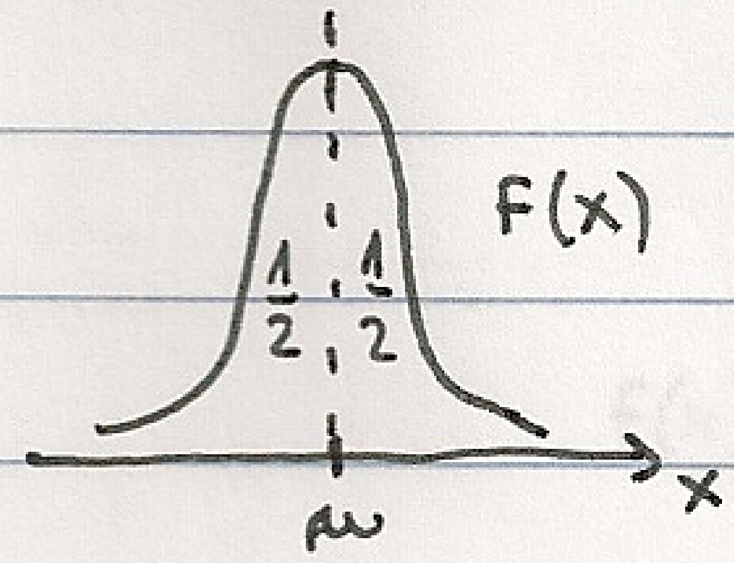


- integrál plochy pod Gaussovou křivkou je rovný 1

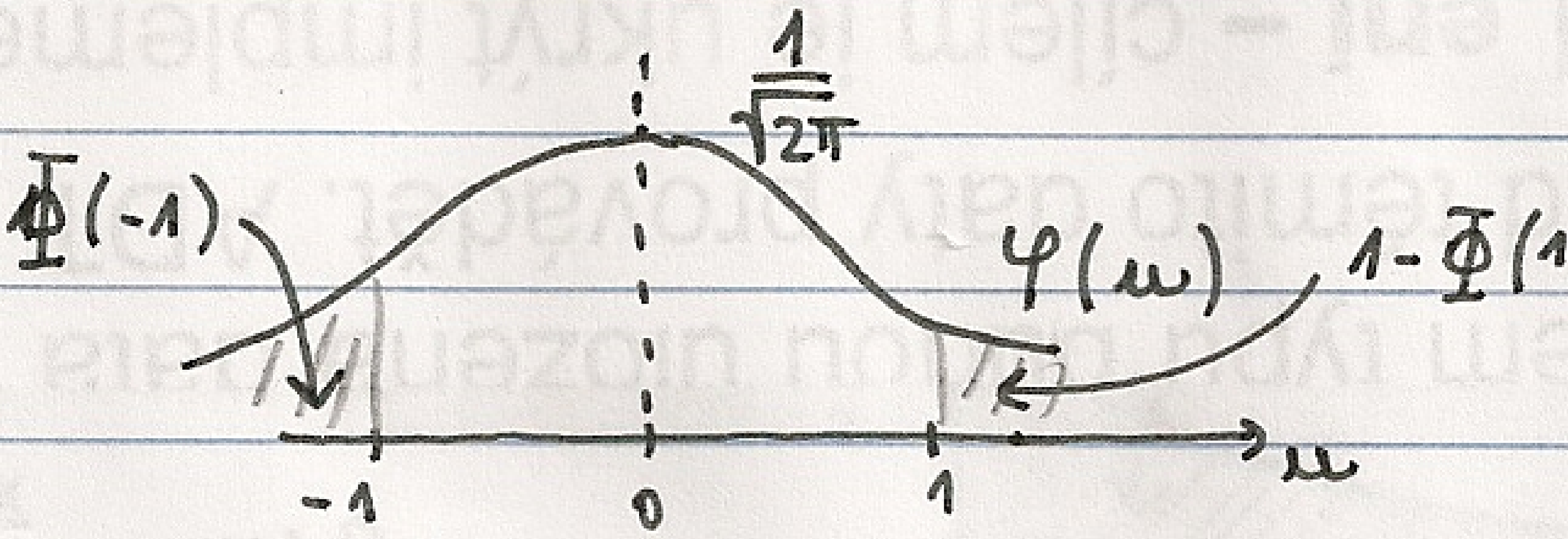
Je-li $F(x)$ distribuční funkce rozdělení $N(\mu, \sigma^2)$, pak

$$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

$X \sim N(\mu, \sigma^2)$

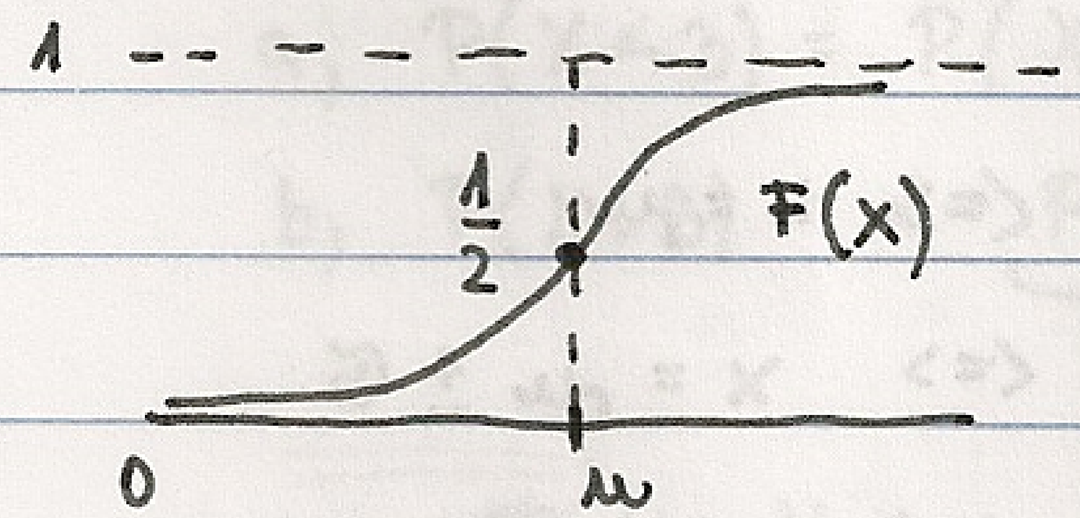


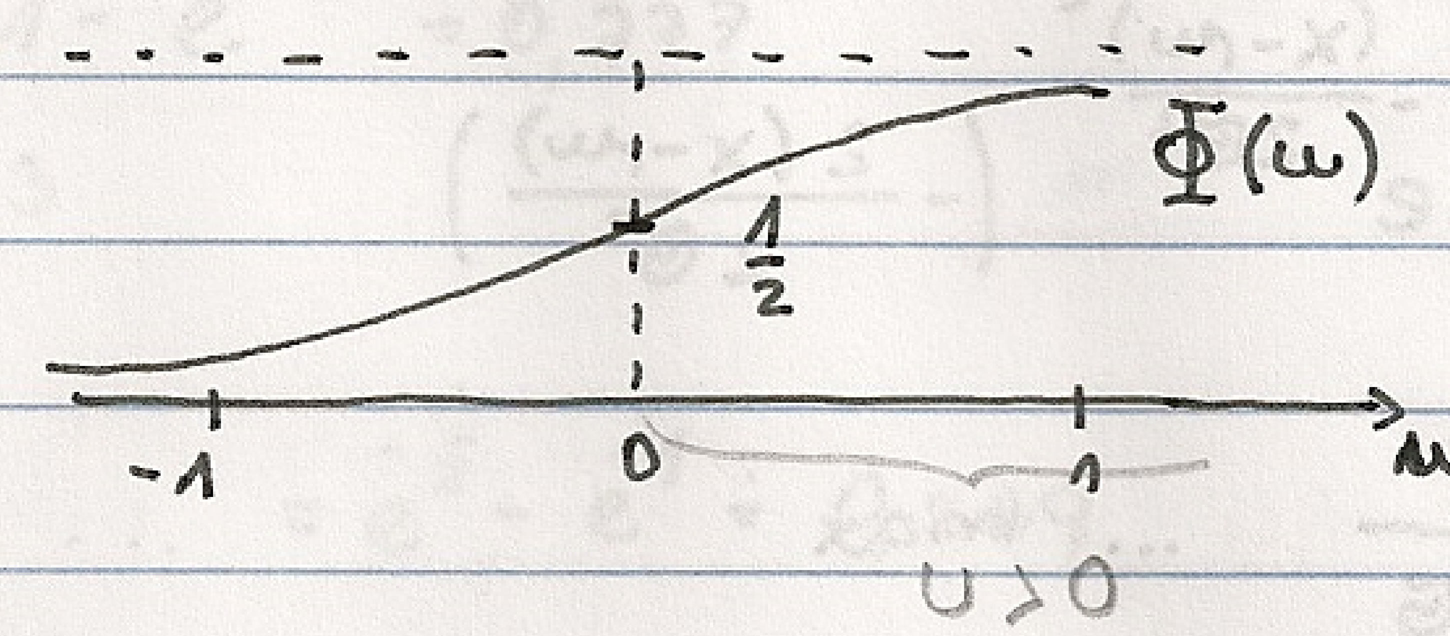
$U \sim N(0,1)$



$$u = \frac{x-\mu}{\sigma}$$

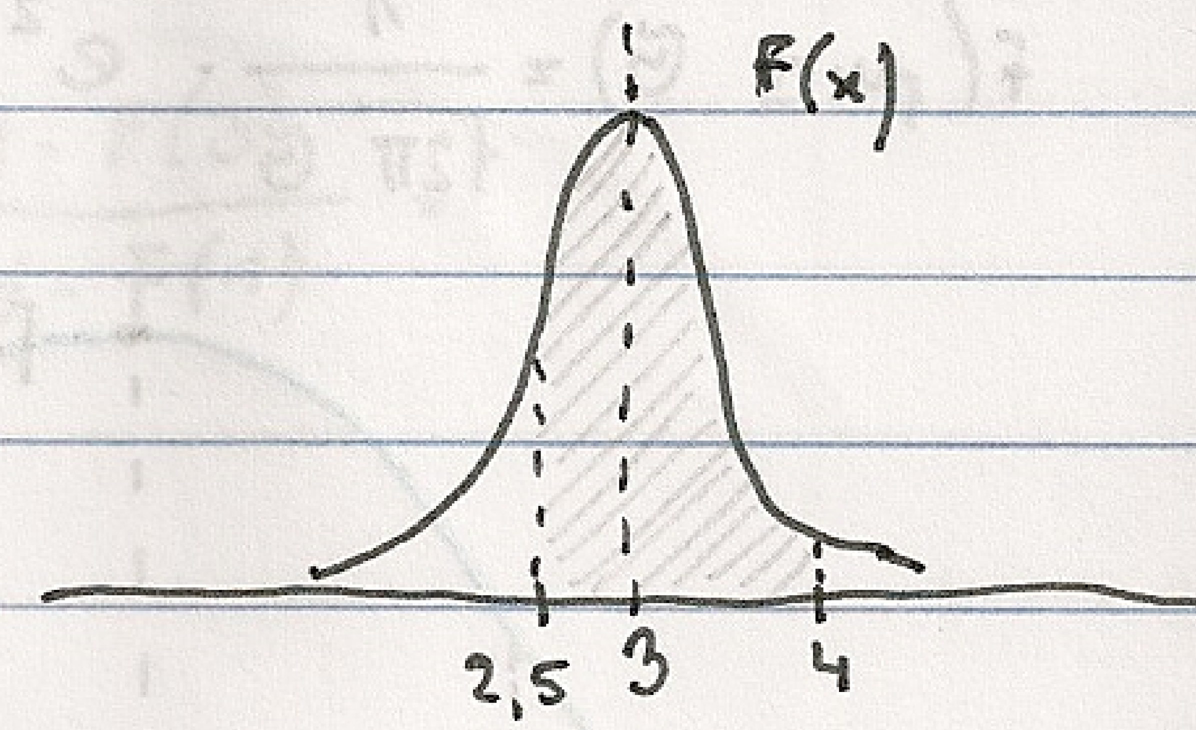
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$





př: Nechť $X \sim N(3;\ 0{,}25)$. Pomocí tabulek určete:

a) $P(2{,}5 < X < 4)$ b) $P(X < 4 \mid X > 3)$



a) $P(2{,}5 < X < 4) = F(4) - F(2{,}5) =$

$$= \Phi\left(\frac{4-3}{\sqrt{0{,}25}}\right) - \Phi\left(\frac{2{,}5-3}{\sqrt{0{,}25}}\right) =$$

$$= \Phi(2) - \underbrace{\Phi(-1)}_{1-\Phi(1)} = 0{,}9972 - (1 - 0{,}841) = \underline{0{,}818}$$

b) $P(X<4 \mid X>3) = \dfrac{P(3<X<4)}{P(X>3)} = \dfrac{F(4)-F(3)}{1-F(3)} = \dfrac{\Phi(2)-\Phi(0)}{1-\Phi(0)} = \dfrac{0{,}9772-0{,}5}{0{,}5} \doteq 0{,}954$

př: Vypočtěte $P(|X-\mu| < k\cdot\sigma)$, kde $k>0$ konst

$$P(\mu - k\cdot\sigma < X < \mu + k\cdot\sigma) = F(\mu + k\cdot\sigma) - F(k\mu + k\sigma) =$$

$$= \Phi\left(\frac{\mu + k\cdot\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) = \Phi(k) - \underbrace{\Phi(-k)}_{1-\Phi(k)} = 2\cdot\Phi(k-1)$$

<u>statistika:</u> pro $k=2$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 2\underbrace{\Phi(2)}_{0{,}977} - 1 \doteq 0{,}954$$

tzv. "pravidlo dvou sigma" - očekáváme cca 95% hodnot v intervalu $\mu \pm 2\sigma$

pro $k=3$ $\Rightarrow$ 0,997 (99,7%)

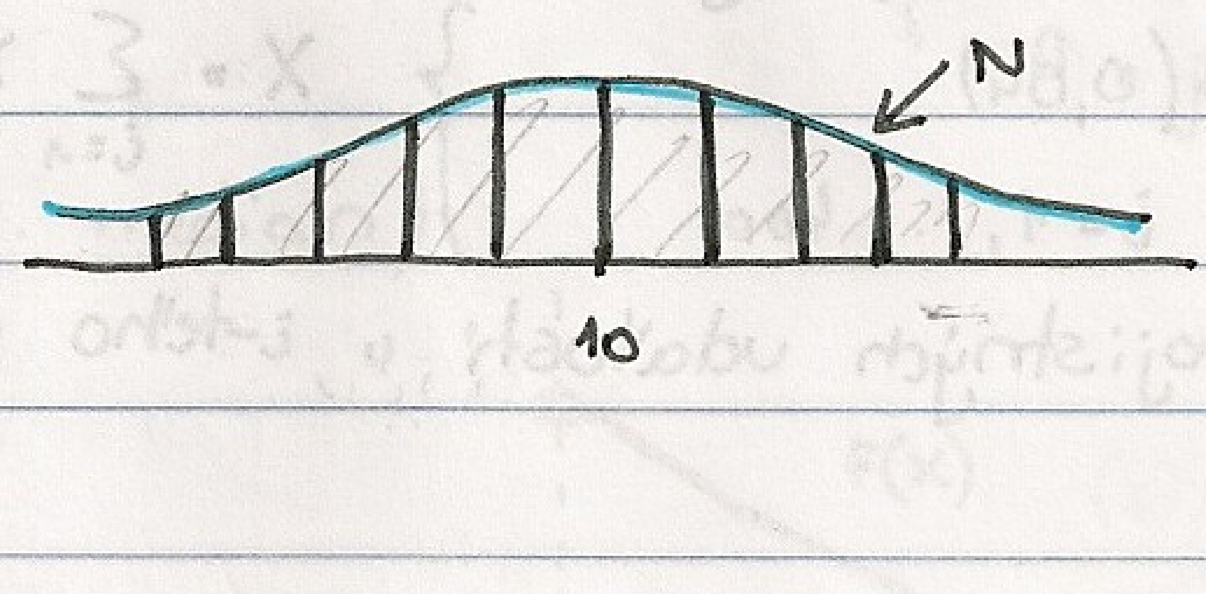
Aproximace normálním rozdělením

př: Zmetkovitost výrobní linky jsou 2%. Spočtěte pravděp., že mezi 500 kontrolovanými výrobky (s vracením zpět → zajišťuje nezávislost) bude nejvýše 15 zmetků.

X ... počet zmetků z 500

$X \sim Bi(\underset{n}{500};\ \underset{p}{0{,}02})$ $\qquad P(X \le 15) \doteq F_N(15) = \Phi\left(\frac{15-10}{\sqrt{9{,}8}}\right) \doteq$

$\approx N(\underset{n\cdot p}{10};\ \underset{n\cdot p(1-p)}{9{,}8})$ $\qquad \doteq \Phi(1{,}60) = 0{,}945$



6. APROXIMACE NORMÁLNÍM ROZDĚLENÍM

Je-li $n \in N$ tak velké, že $\boxed{np\,(1-p) \geq 9}$, pak

$$\boxed{Bi(n,p) \approx N(np,\, np\,(1-p))}$$

Je-li $\boxed{\lambda \geq 9}$, lze použít aproximaci:

$$\boxed{Po(\lambda) \approx N(\lambda,\, \lambda\,)}$$

Jsou-li $X_1, X_2, \ldots, X_n$ **nezávislé** náhodné veličiny se stejným rozdělením, $E(X_i) = \mu_0$, $D(X_i) = \sigma_0^2$, pak pro "dost velké n" platí:

$$\boxed{\sum_{i=1}^{n} X_i \approx N(n\mu_0,\, n\sigma_0^2\,)}$$

Odtud pak: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(\mu_0,\, \frac{\sigma_0^2}{n}\right)$

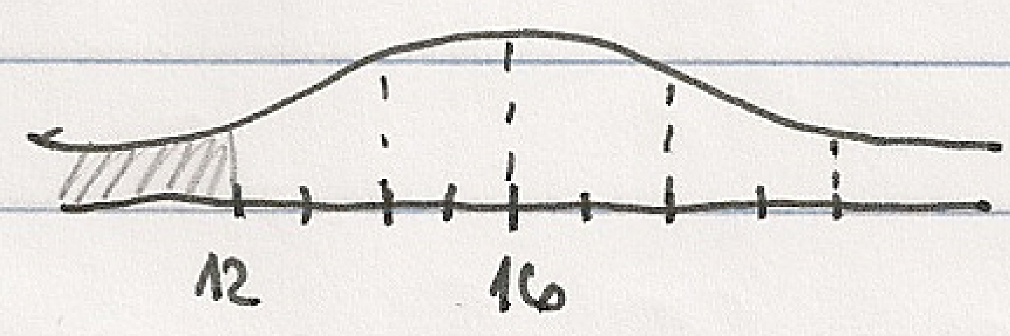
Mají-li $X_1, X_2, \ldots, X_n$ navíc normální rozdělení, má i jejich součet a průměr (přesně) normální rozdělení.

Př: Pravděp. pojistné události u 1 pojištěnce se řídí Poissonovým rozdělením se střední hodnotou 0,04. Spočtěte pravděp., že u 400 pojištěnců nastane za rok nejvýše 12 pojistných událostí.

$400 \cdot 0{,}04 = 16$

$$\left.\begin{array}{l} X_i \sim Po(0{,}04) \\ \uparrow \qquad i = 1, \ldots 400 \end{array}\right\} \quad X = \sum_{i=1}^{400} X_i \sim Po(\underbrace{16}_{\geq 9}) \approx N(16;16)$$

počet pojistných událostí u i-tého pojištěnce

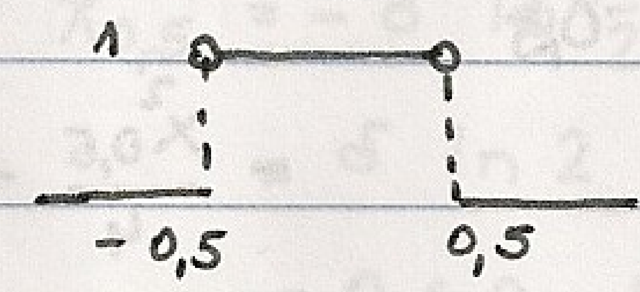


$P(X \leq 12) \doteq F_N(12) = \Phi\left(\frac{12-16}{\sqrt{16}}\right) = \Phi(-1) = 1 - \Phi(1) = 1 - 0{,}841 \doteq \boxed{0{,}159}$

př: Spočtěte pravděp., že po zaokrouhlení 25 čísel na celé jednotky bude celková zaokrouhlovací chyba v absolutní hodnotě menší než 3.

X_i ... zaokr. chyba při zaokr. i-tého čísla

$X_i \sim R(-0{,}5;\ 0{,}5)$



$(\mu_0 =)\, E(X_i) = 0$

$(\sigma_0^2 =)\, D(X_i) = \frac{1}{12}$

X .. celková zaokr. chyba

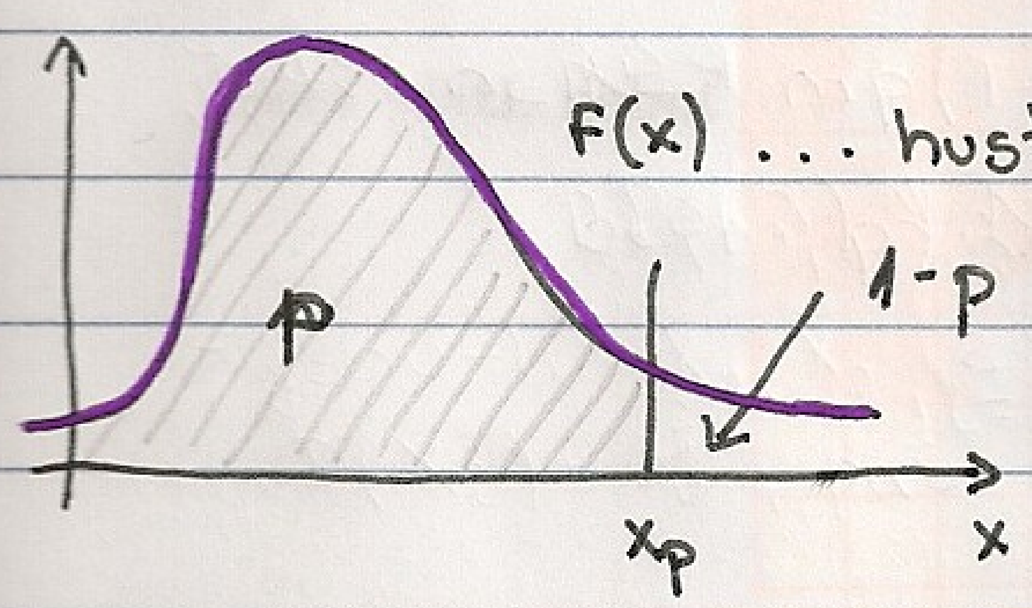
$X = \sum_{i=1}^{25} X_i \approx N\left(0;\ \frac{25}{12}\right)$

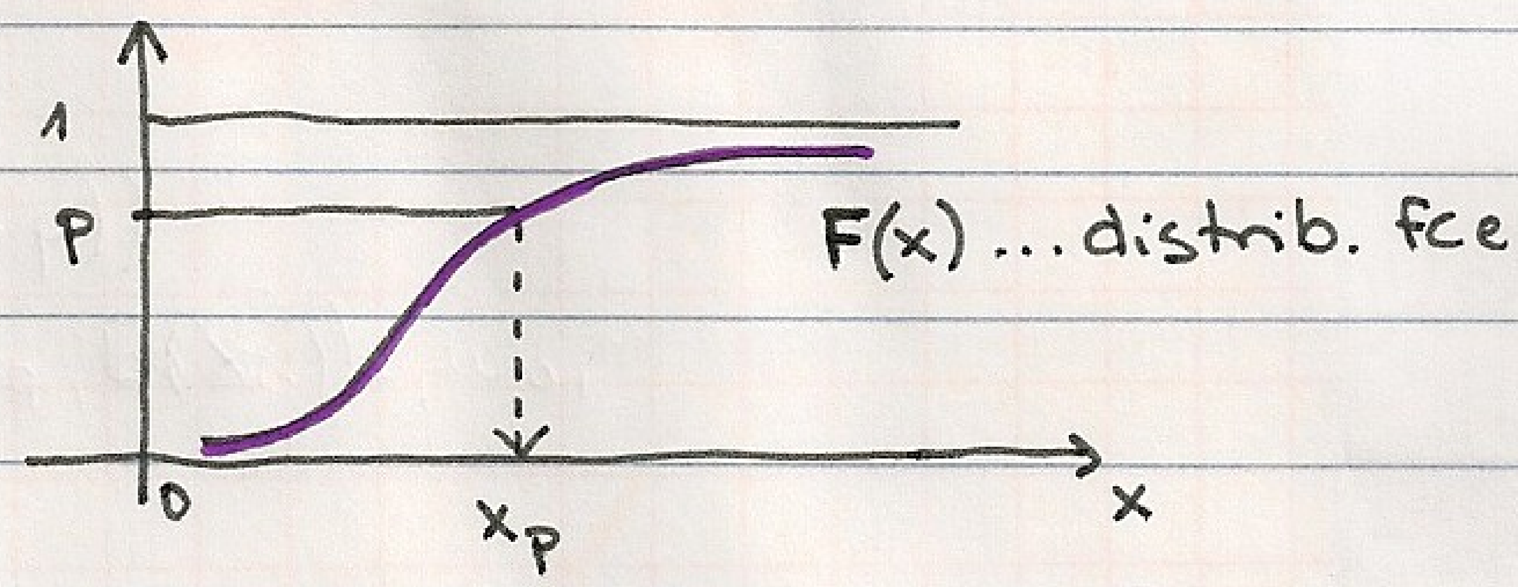
$P(-3 < X < 3) = F(3) - F(-3) = \Phi\left(\frac{3}{\sqrt{\frac{25}{12}}}\right) - \Phi\left(\frac{-3}{\sqrt{\frac{25}{12}}}\right) =$

$\doteq 2 \cdot \Phi(2{,}08) - 1 = 2 \cdot 0{,}9812 - 1 = 0{,}9624 \quad (\approx 96\,\%)$

7. KVANTILY SPOJITÝCH ROZDĚLENÍ

21.10.



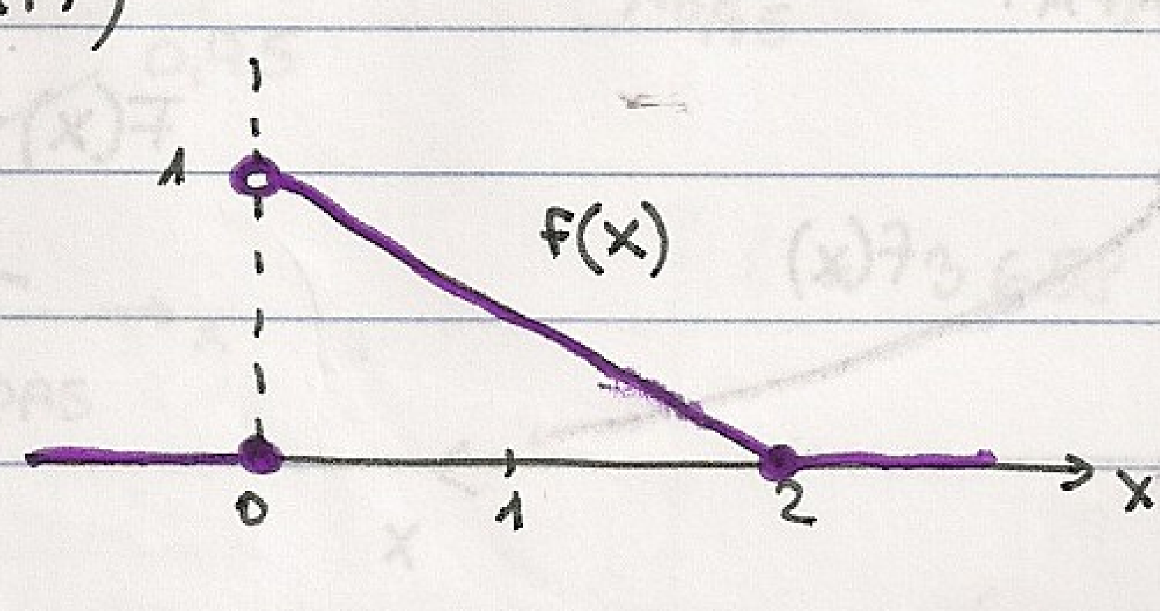


př: Nechť náh. veličina X má hustotu pravděp. $f(x) = \begin{cases} 1 - \frac{x}{2} & \text{.. pro } x \in (0,2) \\ 0 & \text{jinde} \end{cases}$

a, určete distrib. fci $F(x)$ + graf

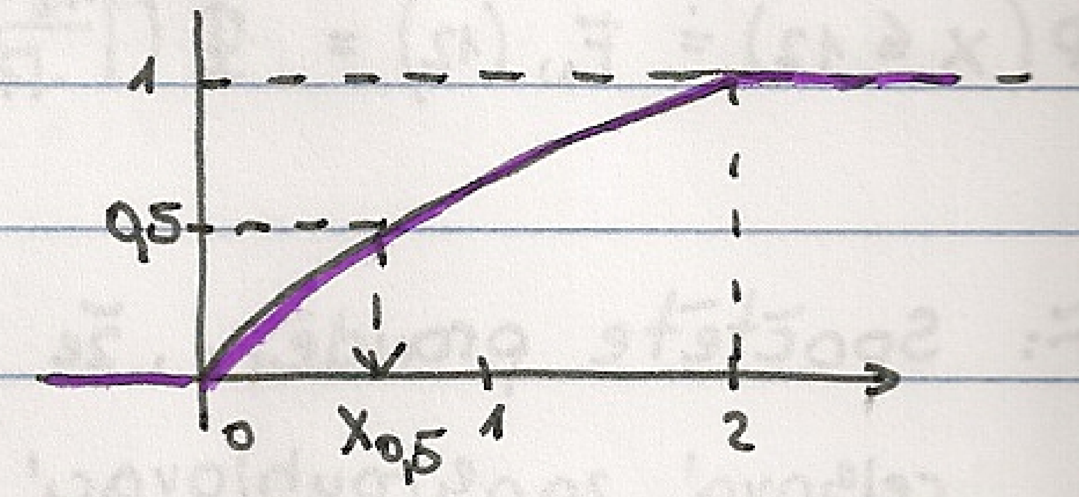
b, spočtěte 50 % kvantil $x_{0,5}$ (tj. medián)

c, spočtěte střední hodnotu $E(X)$



a,

$$F(x) = \begin{cases} 0 \;\dots\; x \le 0 \\ \int_0^x \left(1-\frac{t}{2}\right)dt = \left[t - \frac{t^2}{4}\right]_0^x = -\frac{x^2}{4} + x \;\; \dots \text{pro } x \in (0,2) \\ 1 \;\dots\; x \ge 2 \end{cases}$$



b, $F(x_{0,5}) = 0{,}5$

$$-\frac{x_{0,5}^2}{4} + x_{0,5} = \frac{1}{2} \quad / \cdot (-4)$$

$$x_{0,5}^2 + 4x_{0,5} + 2 = 0$$

$$(x_{0,5})_{1,2} = \frac{4 \pm \sqrt{8}}{2} = \begin{cases} 2+\sqrt{2} \notin (0,2) \\ \boxed{2-\sqrt{2}} = x_{0,5} \end{cases}$$

$$\doteq 0{,}586$$

c, $$E(X) = \int_0^2 x \cdot \left(1 - \frac{x}{2}\right) dx = \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 = 2 - \frac{4}{3} = \underline{\underline{\frac{2}{3}}} \doteq 0{,}667$$

kvantily spojitých rozdělení:

Nechť X je *spojitá* náhodná veličina, $p \in (0,1)$.

Reálné číslo x_p se nazývá **100 p%-NÍ KVANTIL**, je-li

$$\boxed{P(X \le x_p) = p}$$

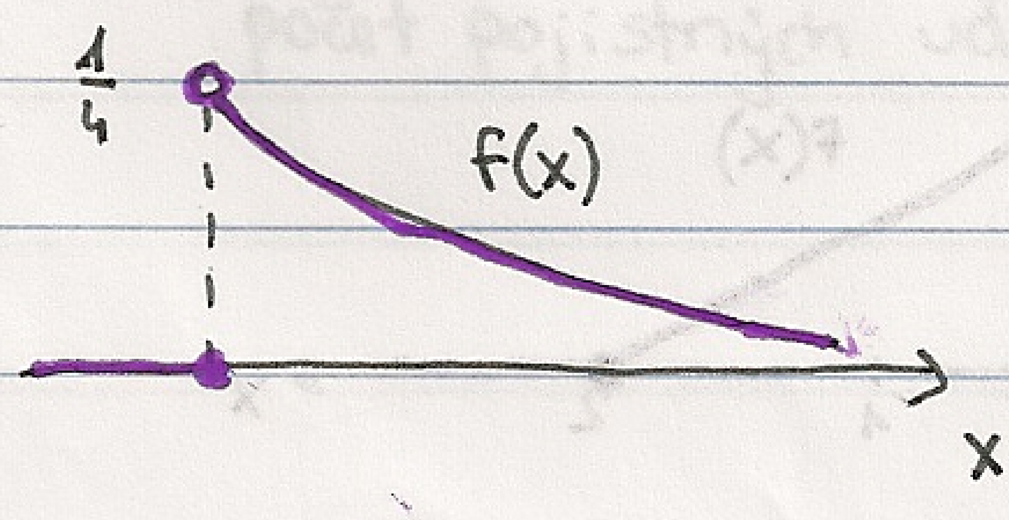
$x_{0,5}$ ... **medián**
$x_{0,25}$ ... **dolní kvartil**
$x_{0,75}$ ... **horní kvartil**

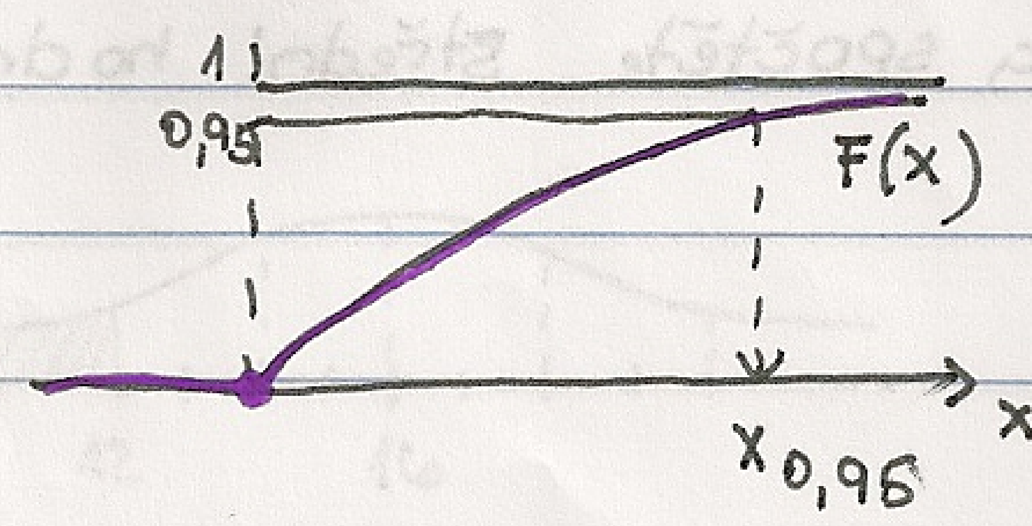
Pomocí distribuční funkce $F(x)$ lze definici kvantilu x_p přepsat takto:

$$\boxed{F(x_p) = p}$$

př: Nechť $X \sim Exp(4)$. Určete $x > 0$ tak, aby $P(X \le x) = 0{,}95$. + graficky znázornit

$$F(x) = \begin{cases} 0 \;\dots\; x \le 0 \\ 1 - e^{-\frac{x}{4}} \quad x > 0 \end{cases}$$





$1 - e^{-\frac{x}{4}} = 0{,}95 \Rightarrow x_{0,95} = 4 \cdot \ln 0{,}05 \doteq 11{,}98$

obecně: $X \sim Exp(\delta)$, $\delta > 0$

$F(x_p) = p$

$x_p = -\delta \ln(1-p)$

speciálně: $x_{0,5} = -\delta \ln \frac{1}{2} =$

$= \delta \ln 2 \doteq 0{,}69 \cdot \delta$

$= 0{,}69 \cdot E(X)$

Určení kvantilů normálního rozdělení

Kvantily x_p rozdělení $N(\mu, \sigma^2)$ lze vyjádřit pomocí kvantilů u_p rozdělení $N(0,1)$, neboť pro $p \in (0,1)$ platí

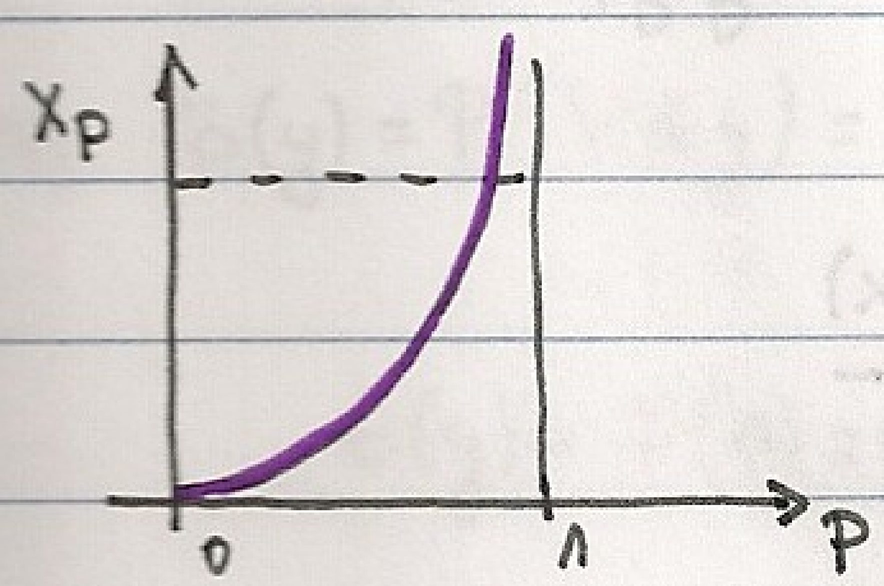
$$\boxed{x_p = \mu + \sigma \cdot u_p}$$

K určení u_p používáme **tabulky**, pro $p < 0,5$ navíc rovnost

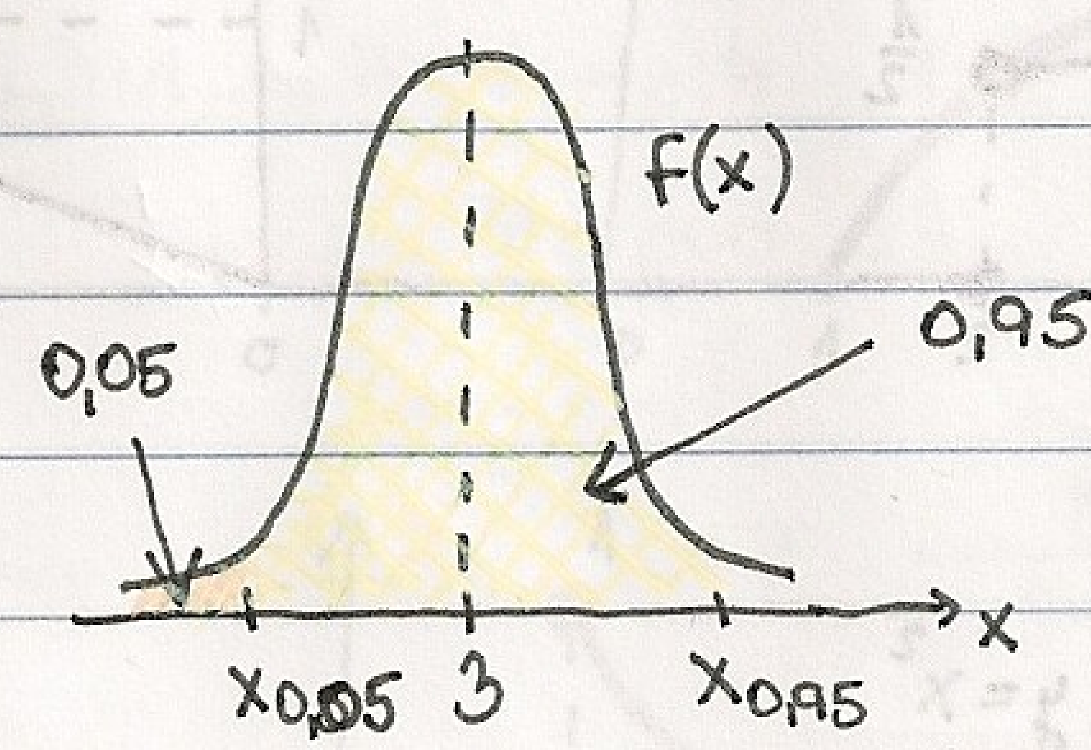
$$\boxed{u_p = -u_{1-p}}$$

[tedy např. $u_{0,05} = -u_{0,95}$]

která platí, neboť hustota $\varphi(u)$ je sudá funkce.



př: pomocí tabulek určete 5%-ní a 95%-ní kvantil náh. veličiny $X \sim N(3;\, 0{,}16)$ + graf



$$x_{0,95} = \overset{\mu}{3} + \overbrace{\sqrt{0{,}16}}^{\sqrt{\sigma^2}} \cdot \underbrace{u_{0,95}}_{\text{tabulky}}$$

$$= 3{,}658$$

$x_{0,05} = 3+\sqrt{0,16} \cdot u_{0,05} = 2,342$ $\qquad u_{0,05} = -u_{0,95} = -1,645$

8. TRANSFORMACE NÁHODNÉ VELIČINY

Dáno: X ... **spojitá** náh. veličina
$f(x)$ ... hustota X.
$F(x)$ ... distrib. fce X

Nechť $Y = h(X)$ (funkce náh. veličiny X).

Pak Y je náh. veličinou, označme:
$g(y)$ ... hustota Y,
$G(y)$ ... distrib. fce Y

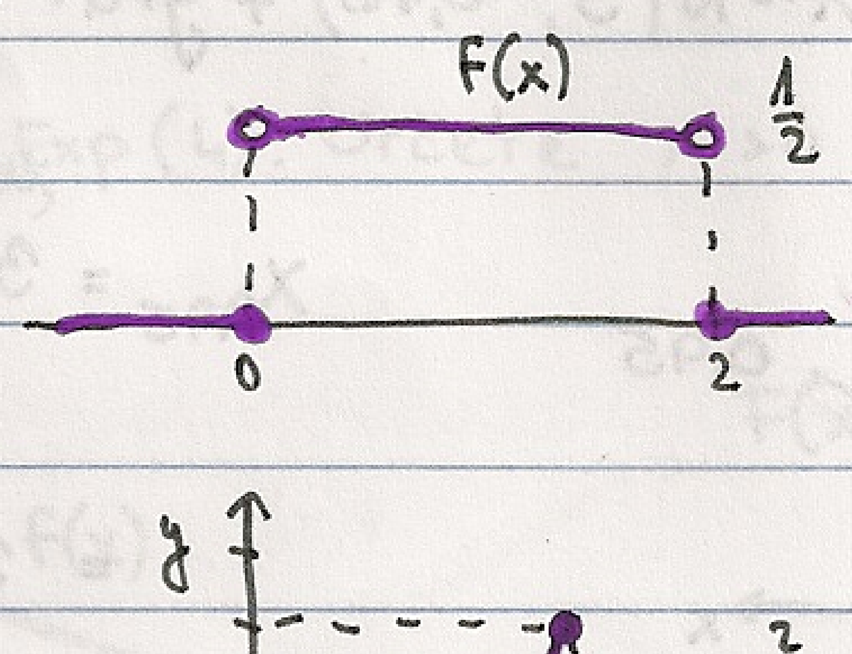
$$g(y) = ?$$
$$G(y) = ?$$

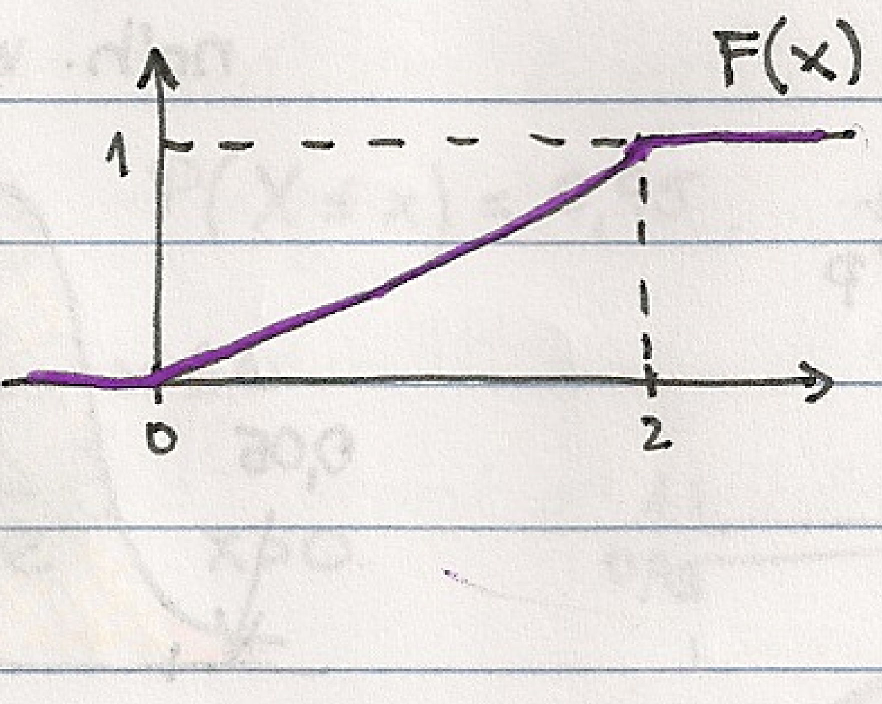
Pozn.: Je-li fce h **konvexní** nebo **ryze monotónní**, je hledání $g(y)$, resp. $G(y)$ jednodušší - viz příklady. Některá takto vzniklá rozdělení ppsti mají své názvy a označení - např. normované normální rozdělení, logaritmicko-normální rozdělení, χ^2-rozdělení.

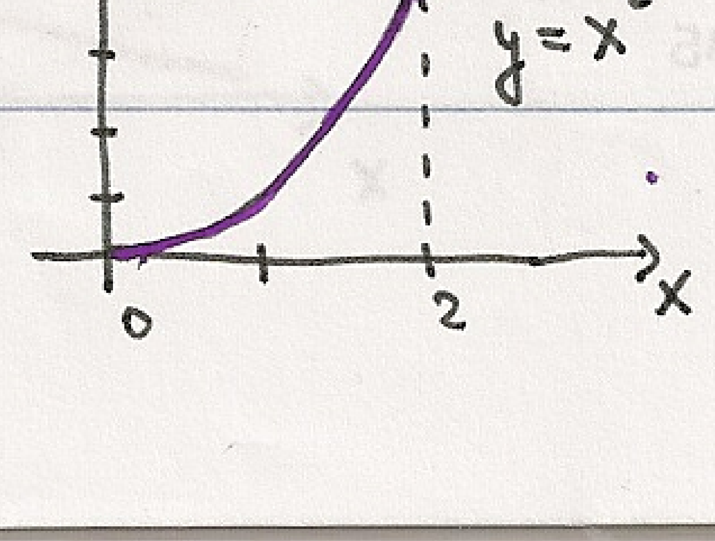
$$G(y) = P(Y \leq y) = P(h(X) \leq y)$$

př.: Nechť $X \sim R(0,2)$. Určete distrib. fci $G(y)$ a hustotu $g(y)$ náh. veličiny. $Y = X^2$ + graf G, g

<u>Řešení:</u>

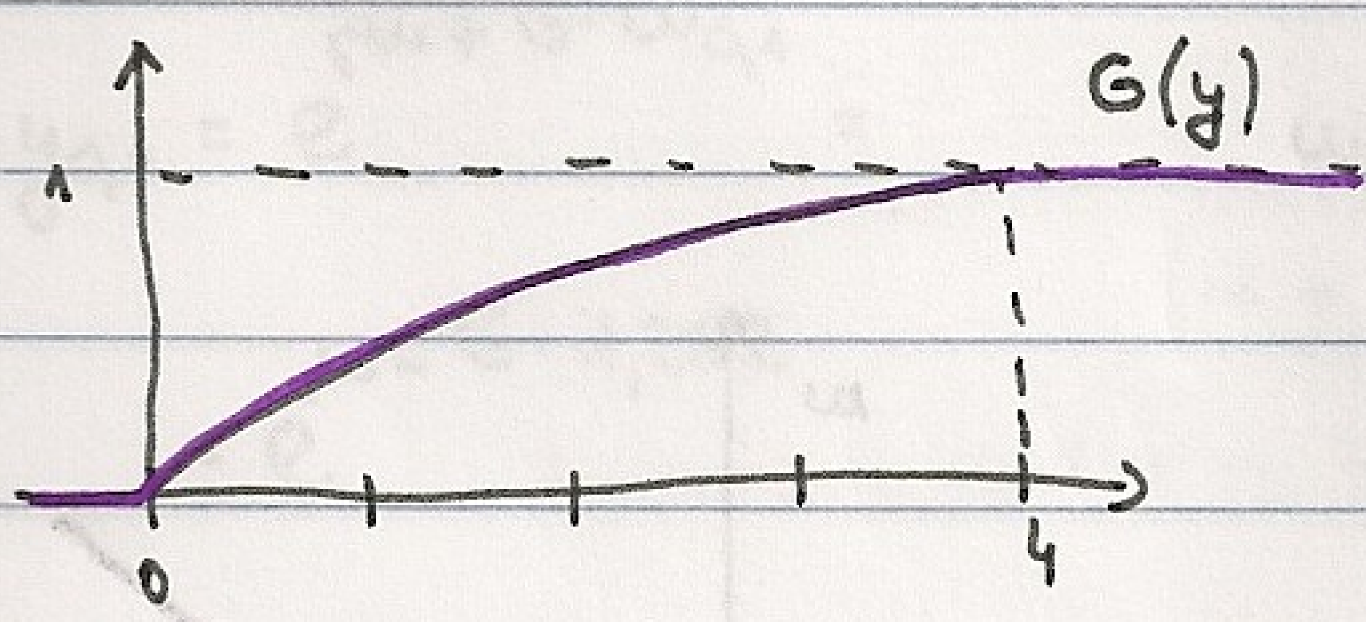






$$G(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F(\sqrt{y}) = \frac{\sqrt{y}}{2} \quad \dots \text{ pro } y \in (0,4)$$

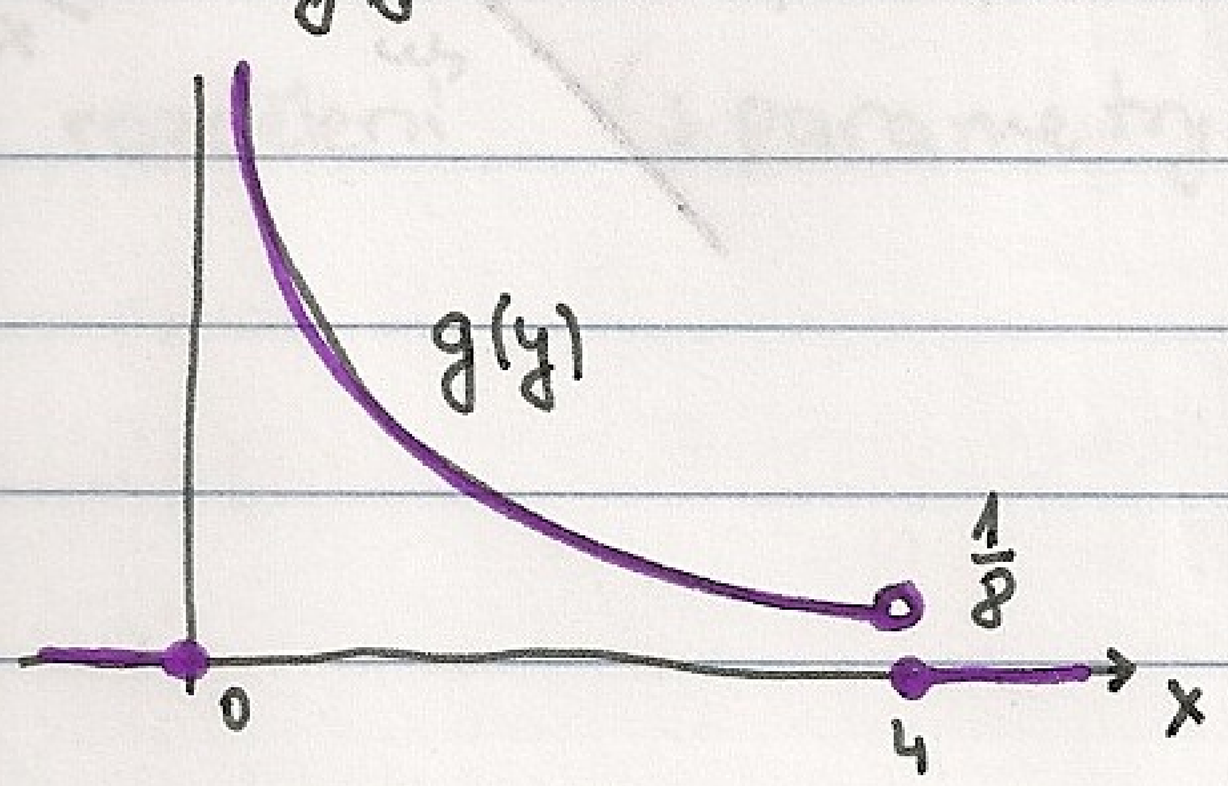
$G(y) = 0$ ... pro $y \le 0$ $\qquad\qquad$ $G(y) = 1$ ... pro $y \ge 4$



$$g(y) = \frac{dG}{dy} = \begin{cases} 0 \dots y \notin (0,4) \\ \frac{1}{4\sqrt{y}} \dots y \in (0,4) \end{cases}$$

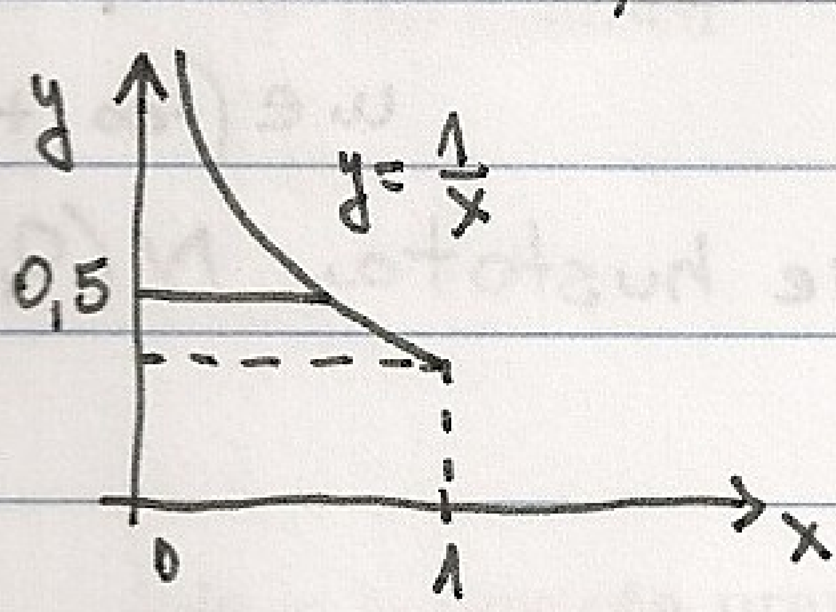
$y \to 0+ \Rightarrow g(y) \to +\infty$

$y \to 4- \Rightarrow g(y) \to \frac{1}{8}$



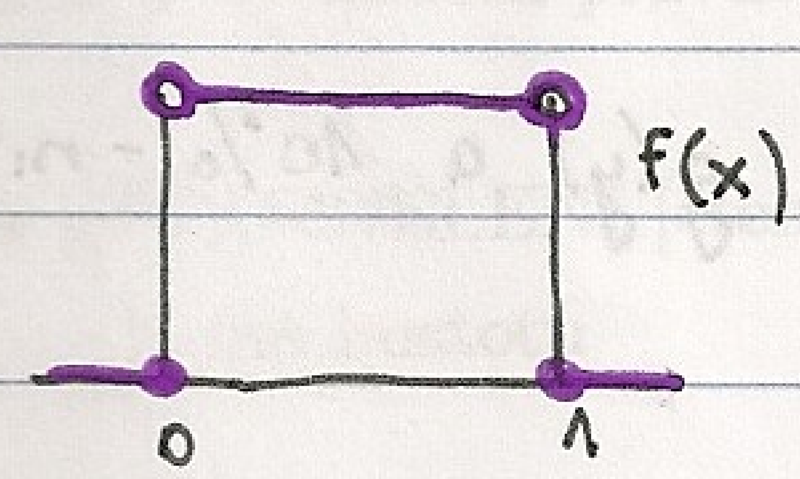
Př: Nechť $X \sim R(0,1)$, $Y = \frac{1}{X}$

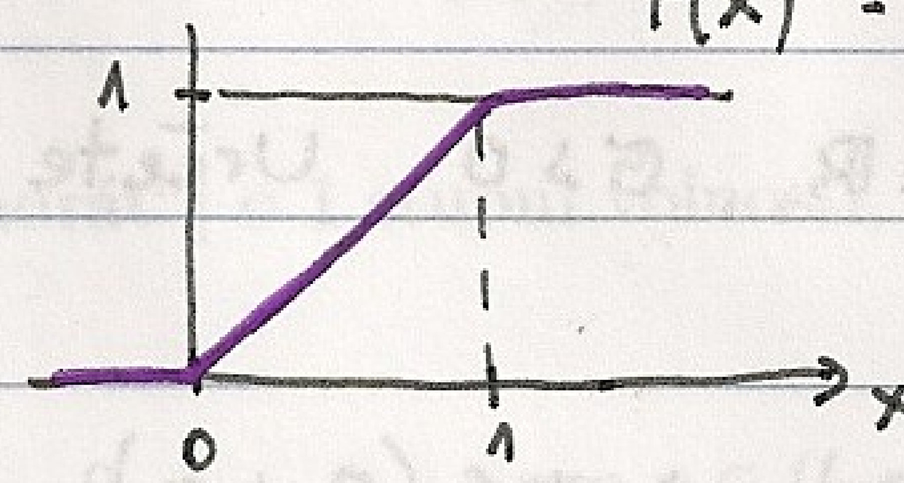
Určete a) $G(y)$ b) $g(y)$ c) $E(Y)$ d) $y_{0,5}$



$x \in (0,1) \Rightarrow y = \frac{1}{x} \in (1, +\infty)$

X:

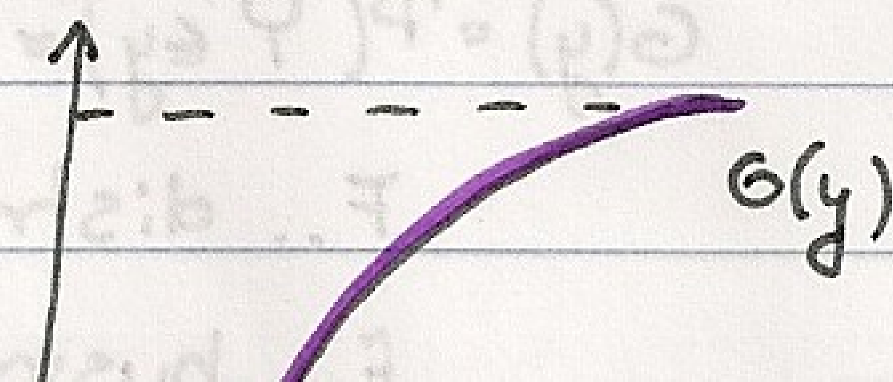




$$F(x) = \begin{cases} 0 \dots x \le 0 \\ x \dots x \in (0,1) \\ 1 \dots x \ge 1 \end{cases}$$

a) $G(y) = P(Y \le y) = P\left(\frac{1}{X} \le y\right) = P\left(\frac{1}{y} \le X\right) = 1 - P\left(X < \frac{1}{y}\right) = 1 - F\left(\frac{1}{y}\right) = 1 - \frac{1}{y}$

$G(y) = P(0) = 0$ .. $y \le 1$



c) $E(Y) = \int_1^{+\infty} y \cdot \frac{1}{y^2}\, dy = \ln y \Rightarrow$ <u>neexistuje</u> $= +\infty$

d) $y_{0,5} = 0,5$

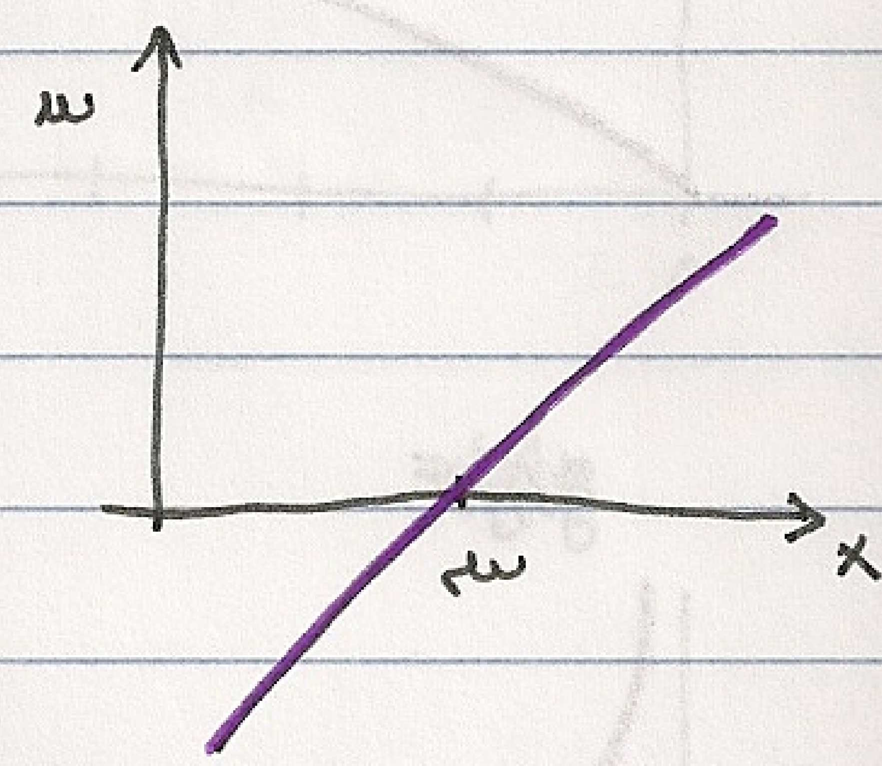
$$1 - \frac{1}{y} = 0,5 \quad \Rightarrow \quad y_{0,5} = 2$$

př.: Nechť $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$

Určete hustotu $\varphi(u)$ náh. veličiny $U = \frac{X - \mu}{\sigma}$

$$U = \frac{X}{\sigma} - \frac{\mu}{\sigma}$$

lineární fce



$X:$ $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in (-\infty, +\infty)$

$F(x) = \int_{-\infty}^{x} f(t)\, dt$

$U:$ $\varphi(u)$ ... hustota

$\Phi(u)$ ... distrib. fce

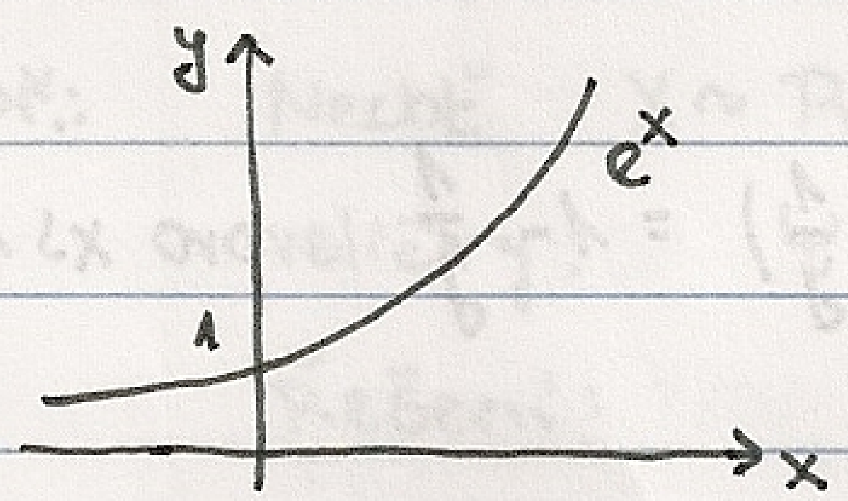
$$U: \quad \Phi(u) = P(U \leq u) = P\left(\frac{X - \mu}{\sigma} \leq u\right) = P(X \leq \mu + \sigma \cdot u) = F(\mu + \sigma \cdot u),$$

$u \in (-\infty, +\infty)$

$$\varphi(u) = \frac{d\Phi}{du} = f(\mu + \sigma \cdot u) \cdot \sigma = \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}} \cdot e^{-\frac{u^2}{2}} \cdot \cancel{\sigma}$$ což je hustota $N(0,1)$

derivace Φ podle u

př: Nechť $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Určete hustotu $g(y)$ a 10%-ní kvantil náh. veličiny $Y = e^X$



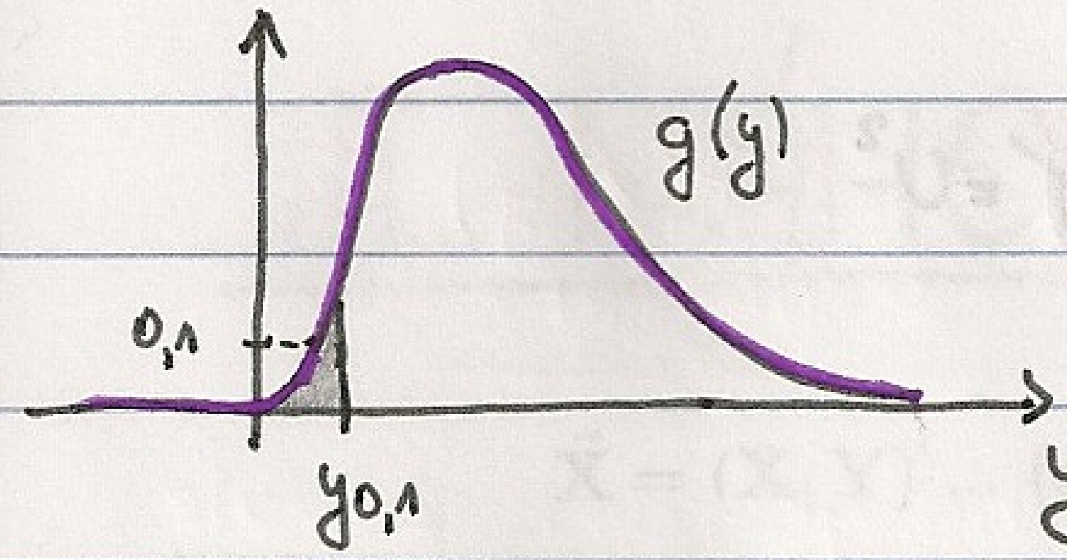
$x \in (-\infty, +\infty) \Rightarrow y \in (0, +\infty)$

$$G(y) = P(Y \leq y) = P(\underbrace{e^X}_{=e^x} \leq y) = P(X \leq \ln y) = F(\ln y) \qquad y > 0$$

F.. distrib fce $N(\mu, \sigma^2)$

f.. hustota

$$g(y) = \frac{dG}{dy} = f(\ln y) \cdot \frac{1}{y} = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot y} \cdot e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \quad \dots \text{ pro } \underline{\underline{y > 0}}$$



$$G(y_{0,1}) = F(\ln y_{0,1}) = \Phi\left(\frac{\ln y_{0,1} - \mu}{\sigma}\right) = y_{0,1} \quad (*)$$

$$\Uparrow$$

$$\Phi(w_{0,1})$$

Φ ... prostá $\Rightarrow \dfrac{\ln y_{0,1} - \mu}{\sigma} = w_{0,1}$ ↖ v tabulkách

$$y_{0,1} = e^{\mu + \sigma \cdot w_{0,1}} = \qquad w_{0,1} = -w_{0,9} = -1{,}282$$

$$= e^{\mu - \sigma \cdot 1{,}282}$$

Y ... s hustotou pravděp. $(*)$ nazýváme LOGARITMICKO-NORMÁLNÍ rozdělení (s parametry $\mu \in \mathbb{R}$, $\sigma > 0$)

Pozn.: Y... spojitá náhodná veličina

- LOGARITMICKO-NORMÁLNÍ rozdělení (s parametry $\mu \in R$, $\sigma > 0$), píšeme $Y \sim N(\mu, \sigma^2)$:

  Y má hustotu

$$g(y) = \begin{cases} 0 & \text{pro } y \in (-\infty, 0] \\ \dfrac{1}{\sigma y} \cdot \varphi\left(\dfrac{\ln y - \mu}{\sigma}\right) & \text{pro } y \in (0, +\infty), \end{cases}$$

kde φ je hustota pravděpodobnosti rozdělení $N(0,1)$.

- χ^2-ROZDĚLENÍ [čteme:"chí-kvadrát"] o 1 stupni volnosti, píšeme $Y \sim \chi^2(1)$:

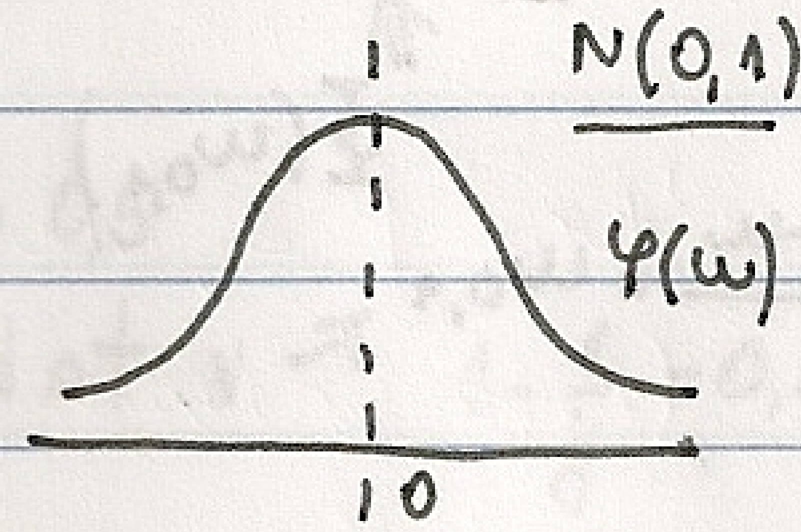
  Y má hustotu

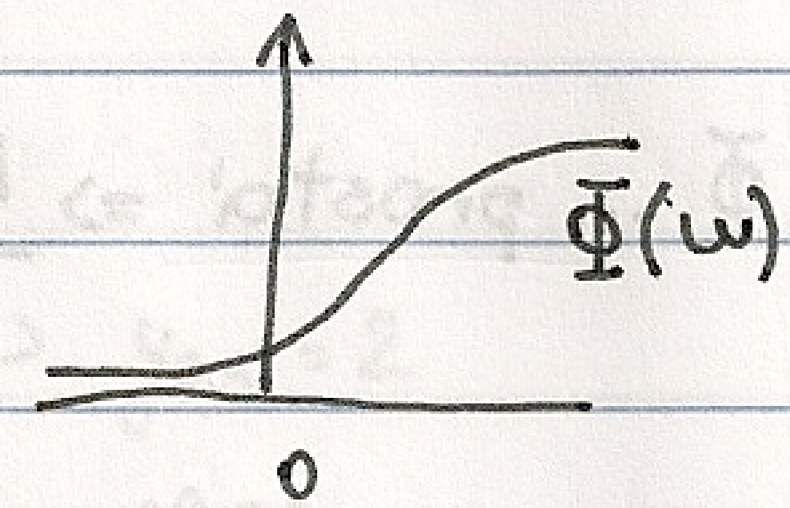
$$g(y) = \begin{cases} 0 & \text{pro } y \in (-\infty, 0] \\ \dfrac{1}{\sqrt{2\pi y}} \cdot e^{-\frac{y}{2}} & \text{pro } y \in (0, +\infty), \end{cases}$$

[Později ve statistice bude $\chi^2(\nu)$, kde $\nu \in N$.]

př: Nechť $U \sim N(0,1)$. Určete hustotu $g(y)$ náh. veličiny $Y = U^2$

U:



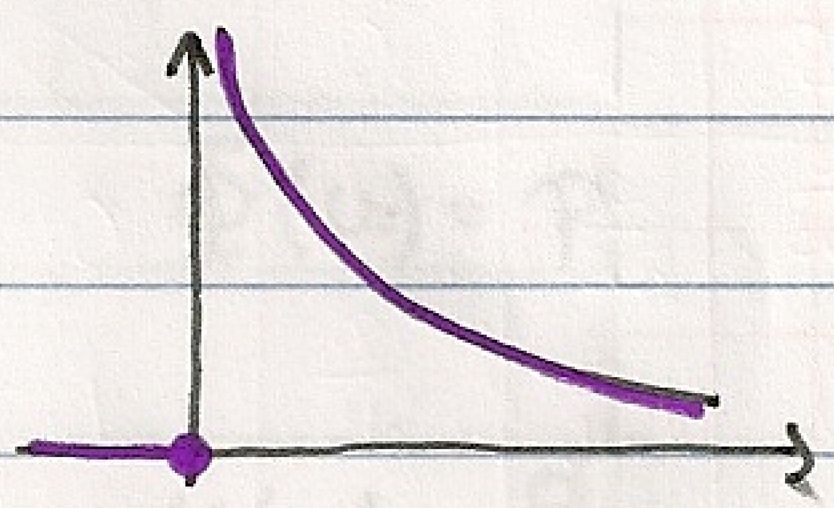


$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \quad \dots \; u \in (-\infty, +\infty) \Rightarrow y = u^2 \in (0, +\infty)$$

$$\Phi(u) = \int_{-\infty}^{u} \varphi(t)\,dt \qquad (\Phi' = \varphi)$$

$$G(y) = P(Y \leq y) = P(U^2 \leq y) = P(-\sqrt{y} \leq U \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-y)$$

$$g(y) = \frac{dG}{dy} = \begin{cases} \varphi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} - \underbrace{\varphi(-\sqrt{y})}_{\varphi(\sqrt{y})} \cdot \left(-\frac{1}{2\sqrt{y}}\right) = \frac{\varphi\sqrt{y}}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} \cdot e^{-\frac{y}{2}} & \dots \; y > 0 \\ 0 & \dots \text{jinde} \end{cases}$$



III. NÁHODNÝ VEKTOR

$\vec{\mathbf{X}} = (\mathbf{X}, \mathbf{Y})$ ... (dvourozměrný) **náhodný vektor**

X, Y ... náh. veličiny (buď obě diskr. nebo obě spoj.)

$\boxed{F(x,y) = P(X \le x), Y \le y)}$ kde $x, y \in (-\infty, +\infty)$

... (sdružená) **distribuční funkce** náh. vektoru $\vec{X}$

$F_1(x) = P(X \le x)$, resp. $F_2(y) = P(Y \le y)$

... tzv. **marginální** distrib.funkce X, resp.Y

$\vec{X} = (X, Y)$ **diskrétní**:

$P(x,y)$ ... (sdružená) **pravděpodobnostní funkce**

$P_1(x)$, $P_2(y)$ ... **marginální** ppstní fce X, resp. Y

$\vec{X} = (X, Y)$ **spojitý**:

$f(x,y)$ ... (sdružená) **hustota** ppsti

$f_1(x)$, $f_2(y)$ ... **marginální** hustoty X, resp. Y

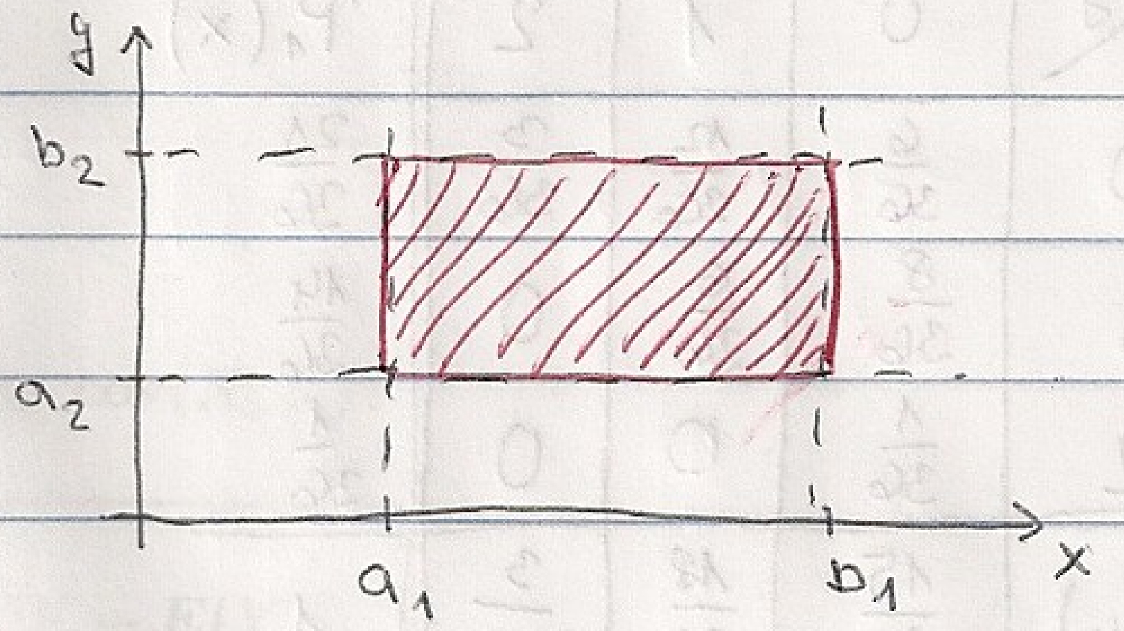
1, $0 \le F(x,y) \le 1$

2, je-li $I = (a_1, b_1\rangle \times (a_2, b_2\rangle$,

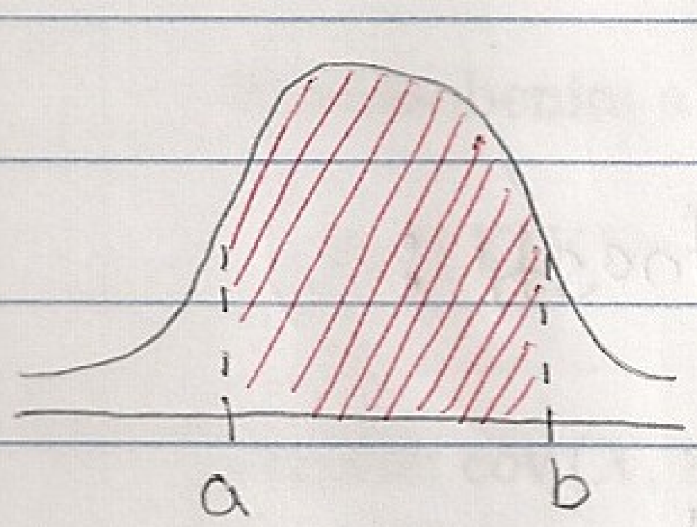
paž $P(\vec{X} \in I) =$

$= F(b_1, b_2) - F(a_1, b_1) -$

$F(a_2, b_1) + F(a_1, a_2)$



$F(b_1, b_2) = P(X \le b_1, Y \le b_2)$



$F(b) - F(a)$

Definice: Řekneme, že náh. ~~veličina~~ vektor $\vec{X} = (X, Y)$ je diskrétního typu, existuje-li konečná nebo spočetná množina $M \subset \mathbb{R}^2$ taková, že

1, $P(\vec{X} = \vec{x}) > 0$ pro $\vec{x} = [x, y] \in M$

$P(\vec{X} = \vec{x}) = 0$ pro $\vec{x} \in \mathbb{R}^2 \setminus M$

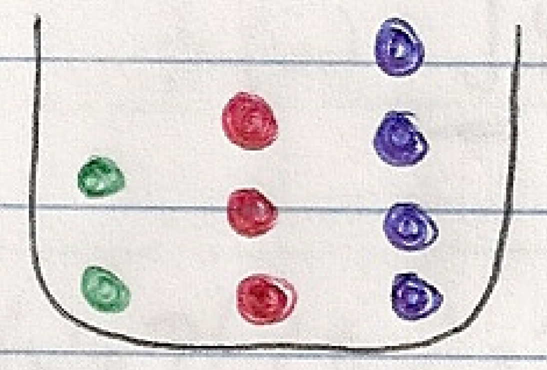
2, $\sum_{\vec{x} \in M} P(\vec{X} = \vec{x}) = 1$ $\qquad$ $A \subset \mathbb{R}^2 \Rightarrow P(\vec{X} \in A) = \sum_{[x,y] \in A \cap M} P(x,y)$

Př: V osudí je 9 koulí: 2 žluté, 3 červené, 4 modré. Náh. vybereme 2 koule (bez vracení zpět).

Označme: X ... počet žlutých (ze 2 vybraných)

Y ... počet červených (ze 2 vybraných)

Stanovte pravděp. fci $P(x,y)$ náh. vektor $\vec{X} = (X, Y)$ + graf

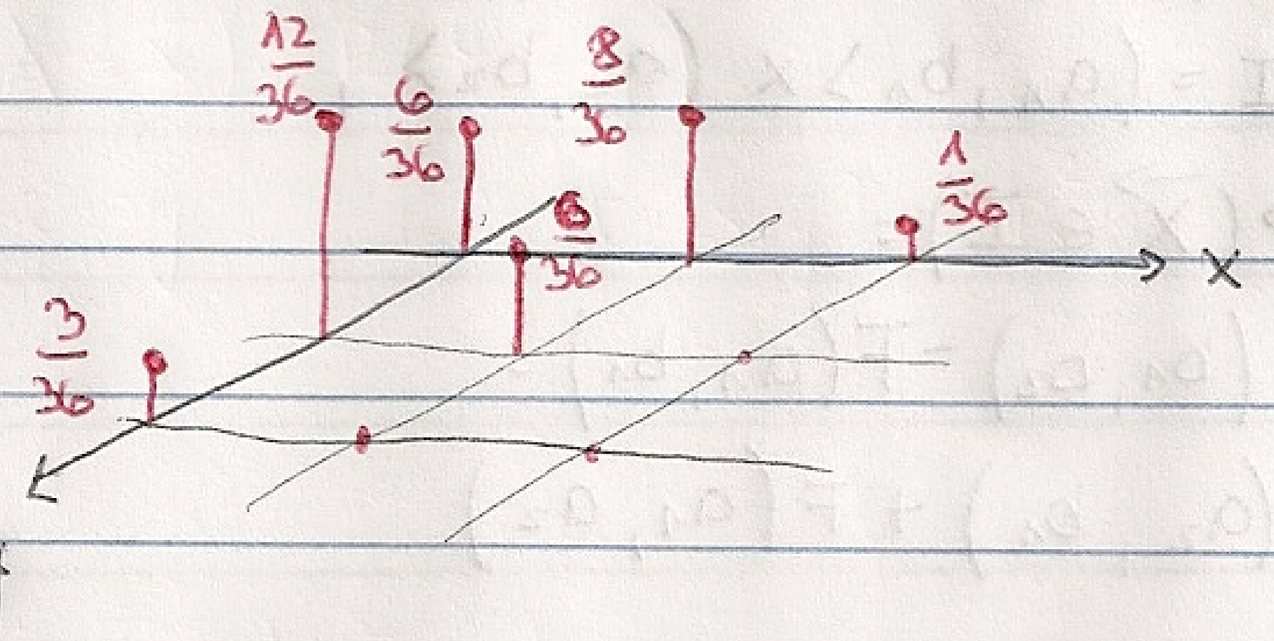


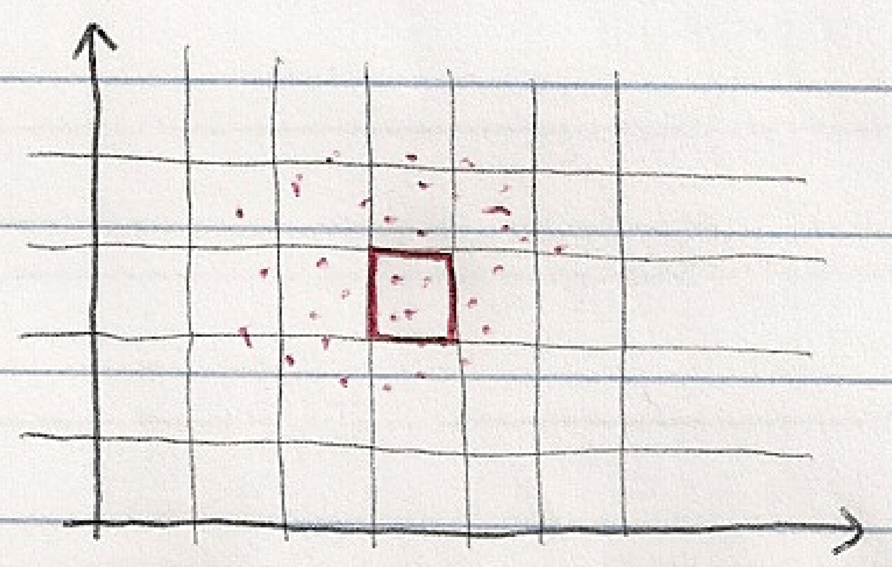
$$P(0,0) = \frac{\binom{4}{2}}{\binom{9}{2}} \qquad P(0,1) = \frac{\binom{3}{1}\binom{4}{1}}{\binom{9}{2}}$$

$$P(1,0) = \frac{\binom{2}{1}\binom{4}{1}}{\binom{9}{2}} \qquad P(1,1) = \frac{\binom{3}{1}\binom{2}{1}}{\binom{9}{2}}$$

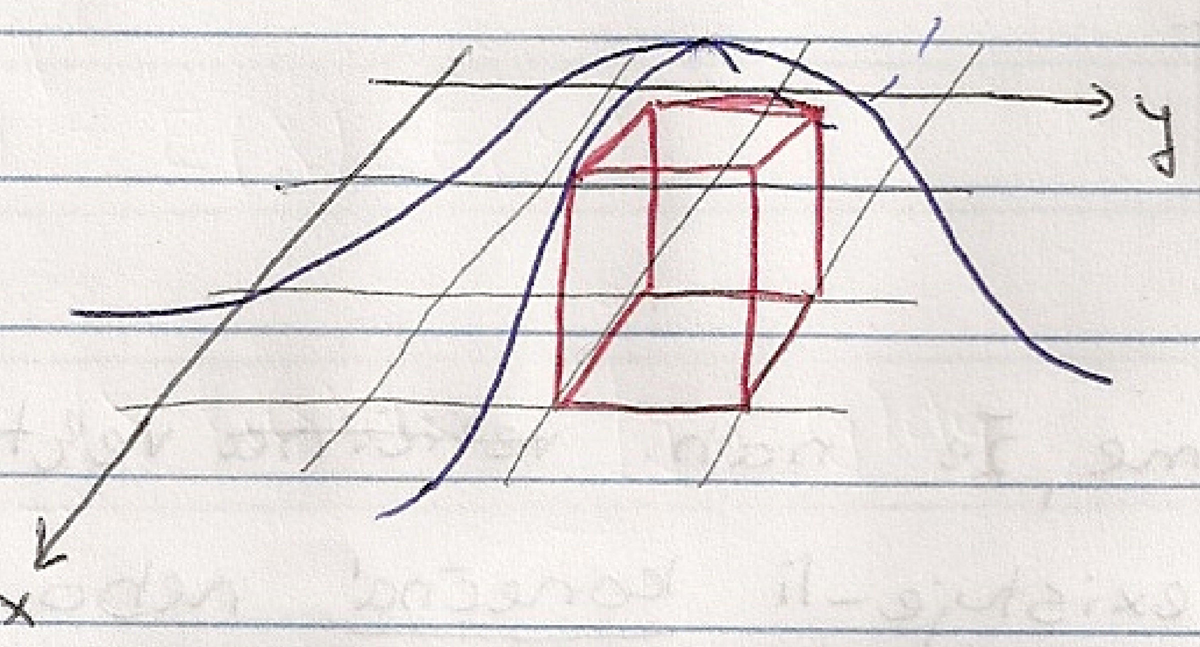
$$P(2,0) = \frac{\binom{2}{2}}{\binom{9}{2}} \qquad P(0,2) = \frac{\binom{3}{2}}{\binom{9}{2}}$$

$x \backslash y$	0	1	2	$P_1(x)$
0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
1	$\frac{8}{36}$	$\frac{6}{36}$	0	$\frac{14}{36}$
2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$P_2(y)$	$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	1





Histogram relativních četností:



1, $f(x,y) \geq 0 \quad \forall [x,y] \in \mathbb{R}^2$

2, $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

objem pod grafem hustoty

• Velikost kvádru je přímoúměrný počtu teček na $\square$

• součet objemů kvádrů $\sum_i \sum_j X_{ij} = 1$

Definice: Řekneme, že náh. vektor $\vec{X} = (X, Y)$ je spojitého typu, existuje-li nezáporná fce $f: \mathbb{R}^2 \to \mathbb{R}$ taková, že pro $x, y \in (-\infty, \infty)$ platí:

$$\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(t_1, t_2)\, dt_2\, dt_1 = F(x_1, x_2) \quad [= F(x, y)],$$

kde F je sdružená distrib. funkce.

Fci $f = f(x, y)$ nazveme (sdružená) hustota pravděpodobnosti

Pozn.: • je-li dána $F(x, y)$, pak $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$ (pro $[x, y]$ kde $\exists$ derivace)

• je-li $A \subset \mathbb{R}^2$, pak $P(\vec{X} \in A) = \iint_A f(x, y)\, dx\, dy$

$E(\vec{X}) = (E(X), E(Y))$ ... vektor středních hodnot

Vztah mezi X a Y udává tzv. **KOVARIANCE**, píšeme $\text{cov}(X, Y)$:

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

Roznásobením a úpravou dostaneme tzv. **výpočetní tvar kovariance**:

$$\text{cov}\,(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Zřejmě: $\text{cov}\,(X, Y) = \text{cov}\,(Y, X)$ (nezáleží na pořadí)

$\text{cov}\,(X, X) = D(X)$

$\text{cov}\,(Y, Y) = D(Y)$

Pro (nekonstantní) náh. veličiny X, Y definujeme tzv.
KORELAČNÍ KOEFICIENT, píšeme $\varrho(X,Y)$:

$$\boxed{\varrho = \varrho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\,\sigma(Y)},}$$

kde $\sigma(X)$, $\sigma(Y)$ jsou směrodatné odchylky veličin X, Y.

Platí :
- $-1 \leq \varrho \leq 1$
- $\varrho(X,Y) = \varrho(Y,X)$
- $\varrho(X,X) = \varrho(Y,Y) = 1$

ϱ je mírou statistické **lineární** závislosti veličin X, Y

$\varrho = +1 \Leftrightarrow Y = aX + b$, kde $a > 0$, $b \in R$

$\varrho = -1 \Leftrightarrow Y = aX + b$, kde $a < 0$, $b \in R$

Je-li $\varrho = 0$, říkáme, že X, Y jsou **nekorelované**.
X, Y nezávislé $\Rightarrow X, Y$ jsou nekorelované
(implikace "$\Leftarrow$"obecně neplatí)

$\rho = 0$ může doprovázet ~~ne~~ lineární nezávislost X a Y

$$P(X \leq x \cap Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$
X, Y nezávislé náh. veličiny $\quad \forall [x,y] \in \mathbb{R}^2$

$\vec{X} = (X, Y)$

X, Y jsou <u>nezávislé</u> $\Leftrightarrow F(x,y) = F_1(x) \cdot F_2(y)$

(resp. $f(x,y) = f_1(x) \cdot f_2(x)$, $\vec{X}$ spoj.
$P(x,y) = P_1(x) \cdot P_2(x)$, $\vec{X}$ spoj.)

$$\Downarrow$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

$$E(X \cdot Y) = \iint_{\mathbb{R}^2} x \cdot y \cdot \underbrace{f(x,y)}_{f_1(x) \cdot f_2(y)} \, dx\, dy = \underbrace{\int_{-\infty}^{\infty} x \cdot f_1(x)\, dx}_{E(X)} \cdot \underbrace{\int_{-\infty}^{\infty} y \cdot f_2(y)\, dy}_{E(Y)}$$

Pro náh. vektor $\vec{X} = (X_1, X_2, \ldots, X_k)$ lze kovariance, resp. korelační koeficienty uspořádat do čtvercových matic $\Sigma = (\sigma_{ij})$, resp. $R = (\varrho_{ij})$, a to označením

$$\sigma_{ij} = \operatorname{cov}(X_i, X_j)\ , \text{ resp. } \quad \varrho_{ij} = \varrho(X_i, X_j)$$

Pak

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}$$

resp.

$$R = \begin{bmatrix} \varrho_{11} & \varrho_{12} & \cdots & \varrho_{1k} \\ \varrho_{21} & \varrho_{22} & \cdots & \varrho_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \varrho_{k1} & \varrho_{k2} & \cdots & \varrho_{kk} \end{bmatrix}$$

Σ ... **kovarianční matice**, je symetrická, na hlavní diagonále jsou rozptyly $D(X_i)$

R ... **korelanční matice**, je symetrická, na hlavní diagonále jsou 1

př: Náh. vektor $\vec{X} = (X, Y)$ je určen pravděpodobnostní funkcí

$P(x,y)$: $P(0,0) = 0{,}1$ $\quad P(0,1) = 0{,}2$

$P(1,0) = 0{,}2$ $\quad P(1,1) = 0{,}3$

$P(2,0) = 0{,}1$ $\quad P(2,1) = 0{,}1$

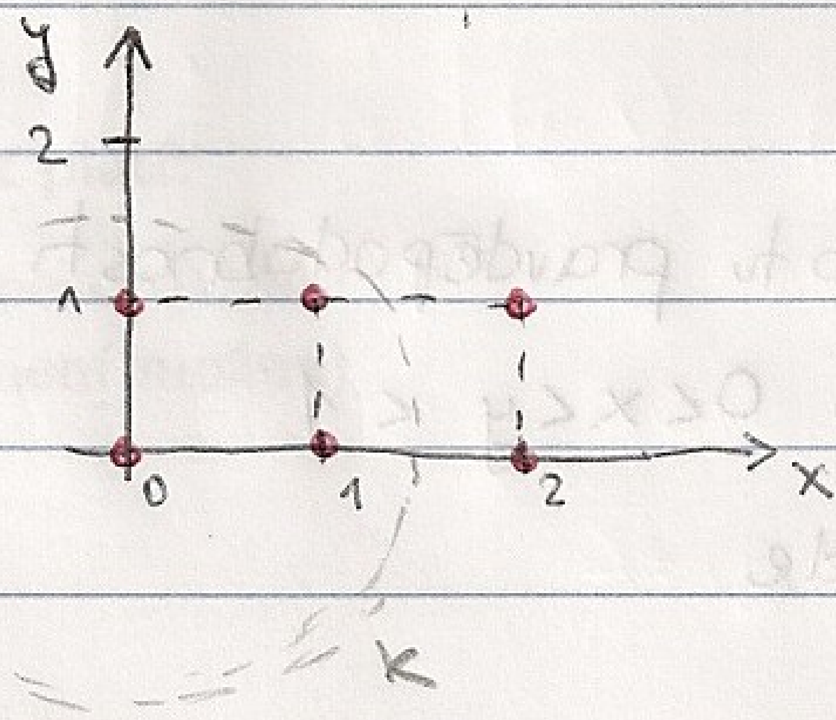
a, určete $P(\vec{X} \in k)$, kde $k = \{[x,y] \in \mathbb{R}^2 ;\ x^2 + y^2 < 3\}$ kruh s poloměrem $\sqrt{3}$

b, vypočtěte vektor středních hodnot $E(\vec{X})$

c, vypočtěte kovariační matici Σ

d, vypočtěte korelační matici R

Řešení: a,



- v kruhu K jsou jen 4 body s nenulovou pravděp

$P(\vec{X} \in k) = 0{,}1 + 0{,}2 + 0{,}2 + 0{,}3 = 0{,}8$

b,

x \ y	0	1	$P_1(x)$
0	0,1	0,2	0,3
1	0,2	0,3	0,5
2	0,1	0,1	0,2
$P_2(y)$	0,4	0,6	1

$E(\vec{X}) = (E(X), E(Y)) = (0{,}9\ ;\ 0{,}6)$

$E(X) = \sum_x x\,P_1(x) = 0 \cdot 0{,}3 + 1 \cdot 0{,}5 + 2 \cdot 0{,}1 = 0{,}9$

$E(Y) = \sum_y y \cdot P_2(y) = 0 \cdot 0{,}4 + 1 \cdot 0{,}6 = 0{,}6$

c, $\underset{\sigma_{12}=\sigma_{21}}{\overset{\parallel}{cov}}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y) = 0{,}5 - 0{,}9\cdot 0{,}6 = \underline{-0{,}04}$

$E(X\cdot Y) = \sum_x \sum_y x\cdot y\, P(x,y) = 1\cdot 1\cdot 0{,}3 + 2\cdot 1\cdot 0{,}1 = \underline{0{,}5}$
(ostatní členy $=0$)

$\sigma_{11} = D(X) = E(X^2) - E^2(X) = 1{,}3 - 0{,}9^2 = \underline{0{,}49}$

$E(X^2) = \sum_x x^2 P_1(x) = 0^2\cdot 0{,}3 + 1^2\cdot 0{,}5 + 2^2\cdot 0{,}2 = 1{,}3$

$\sigma_{22} = D(X) = E(Y^2) - E^2(Y) = 0{,}6 - 0{,}6^2 = \underline{0{,}24}$

$E(Y^2) = \sum_y y^2 P_2(y) = 0{,}6$

$$\Sigma = \begin{bmatrix} 0{,}49 & -0{,}04 \\ -0{,}04 & 0{,}24 \end{bmatrix}$$

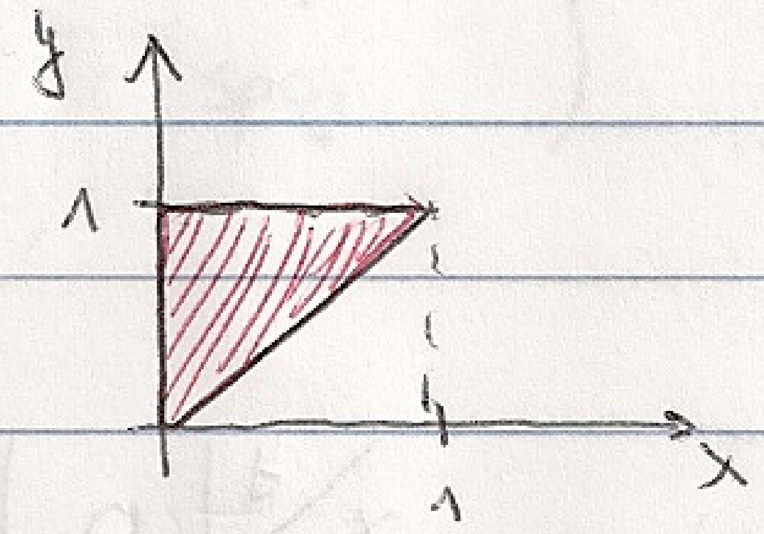
d, $\rho_{11} = \rho_{22} = 1$

$\rho_{12} = \rho_{21} = \dfrac{cov(X,Y)}{\sigma(X)\cdot\sigma(Y)} = \dfrac{-0{,}04}{\sqrt{0{,}49}\cdot\sqrt{0{,}24}} \doteq -0{,}12$

$$R = \begin{bmatrix} 1 & -0{,}12 \\ -0{,}12 & 1 \end{bmatrix}$$

př: Nechť $\vec{X} = (X,Y)$ má hustotu pravděpodobnosti

$f(x,y) = \begin{cases} c & \dots \text{ pro } 0 < x < y < 1 \\ 0 & \dots \text{ jinde} \end{cases}$

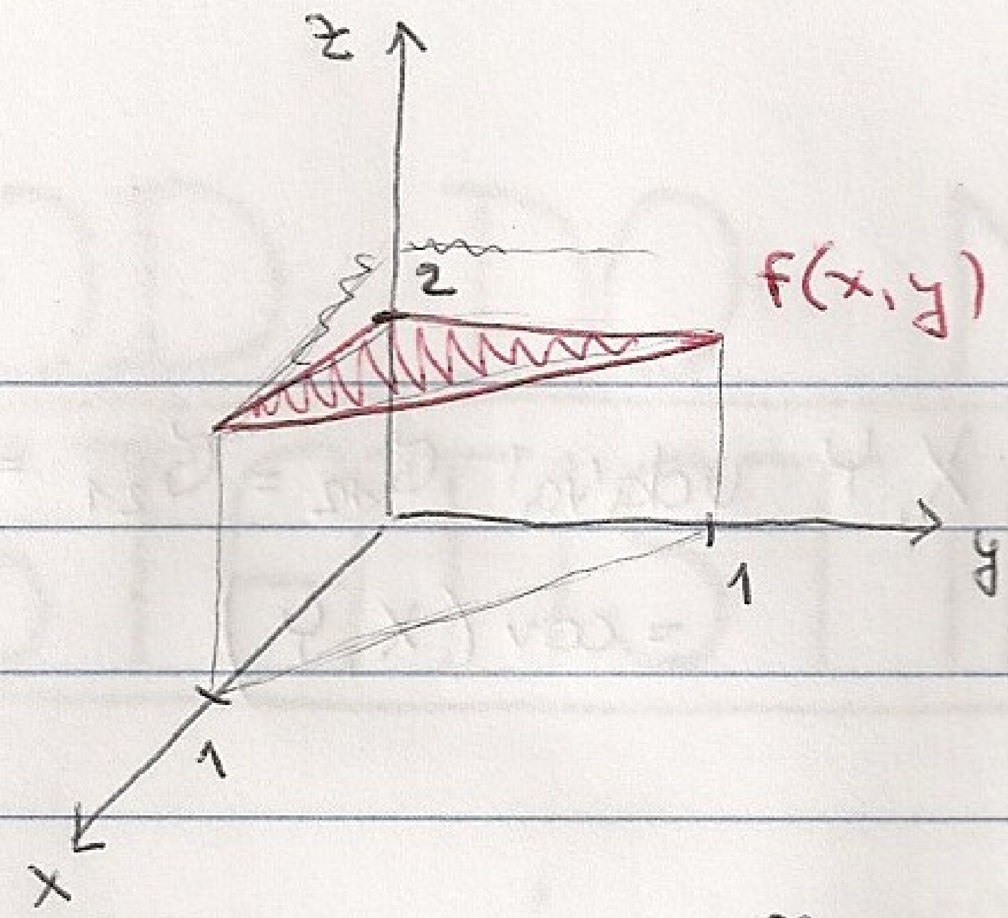


a, určete konstantu $\underline{c}$

b, určete podmíněnou ~~pravděp.~~ stř. hodnotu
$E(Y \mid X = x)$ tzv. regresivní fce

Řešení: a, $1 = \iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = \int_0^1 \left(\int_x^1 c\,dy\right) dx =$

$= c\int_0^1 (1-x)\,dx = c\cdot\left[x - \frac{x^2}{2}\right]_0^1 = c\cdot\frac{1}{2} \Rightarrow \underline{\underline{c=2}}$



b) $E(Y \mid X=x) = \int_{-\infty}^{\infty} y \cdot \underbrace{f(y \mid x)}_{\text{podmíněná hustota}} dy = \int_{x}^{1} y \cdot \frac{1}{1-x} dy = \frac{1}{1-x} \cdot \left[\frac{y^2}{2}\right]_x^1 = \frac{x+1}{2}$ pro $x \in (0,1)$

$$f(y \mid x) = \frac{F(x,y)}{f_1(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}$$

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \begin{cases} \int_x^1 2\, dy = 2(1-x) & \text{pro } x \in (0,1) \\ 0 & \text{jinde} \end{cases}$$

Dvourozměrné normální rozdělení

s vektorem středních hodnot $\vec{\mu} = (\mu_1, \mu_2)$ a kovarianční maticí Σ má hustotu

$$f(x,y) = \frac{1}{2\pi\sqrt{det\,\Sigma}} \cdot e^{-\frac{1}{2}P(x,y)},$$

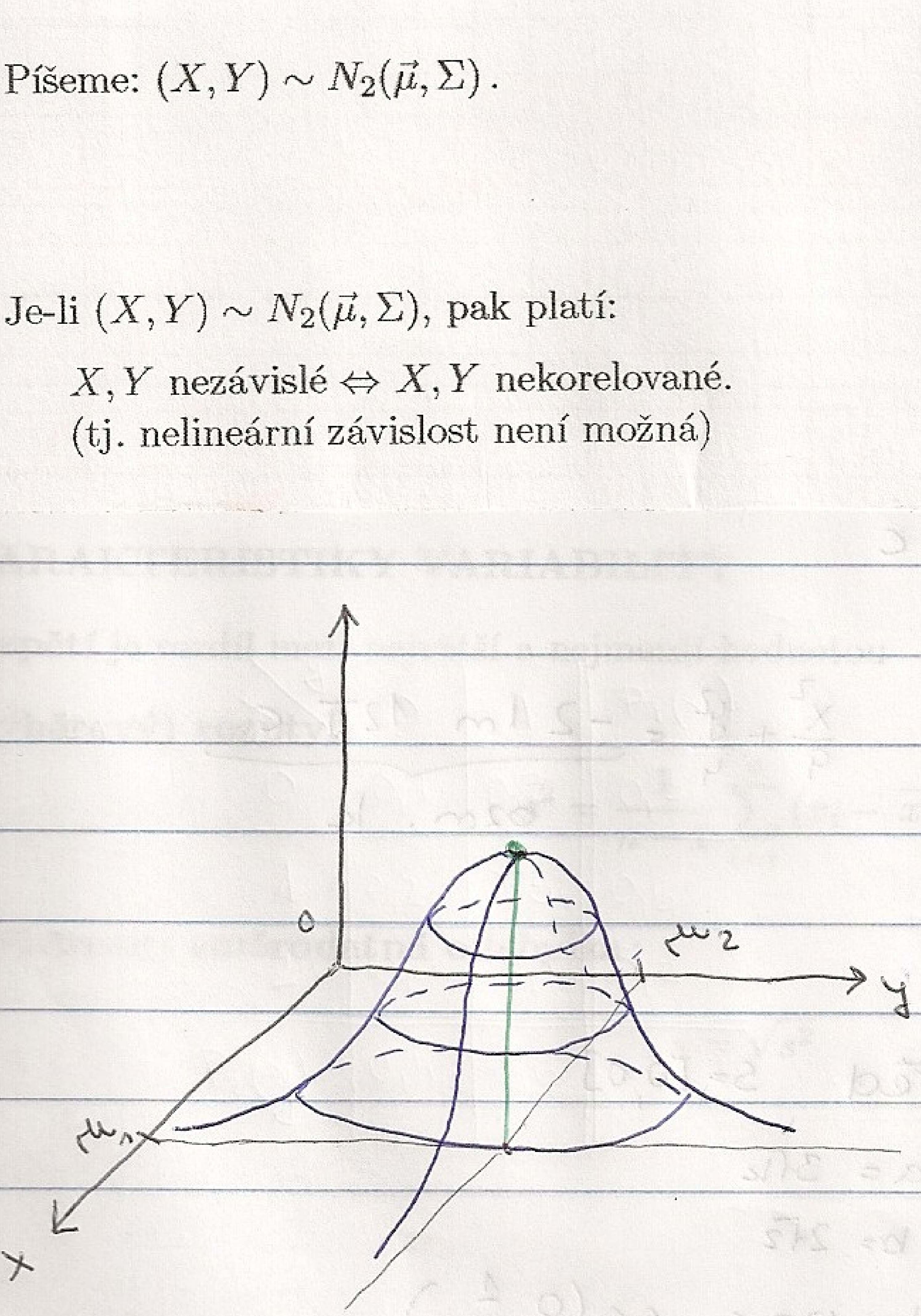
kde

$$P(x,y) = (x-\mu_1, y-\mu_2)\,\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}.$$

Píšeme: $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$.

Je-li $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$, pak platí:

X, Y nezávislé $\Leftrightarrow$ X, Y nekorelované.
(tj. nelineární závislost není možná)



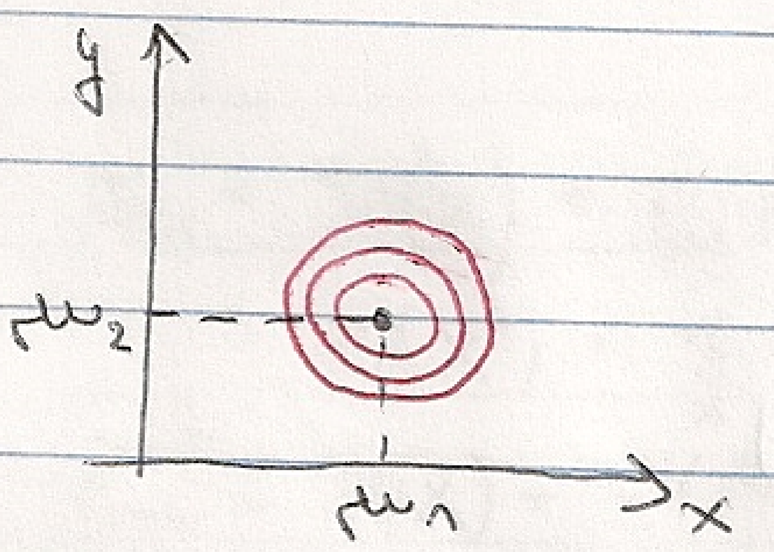
- v bodě $[\mu_1, \mu_2]$ je max fce
- když $x = \mu_1$ dostaneme řez rovinnou YZ
- řezy rovinami $x = \mu_1$, resp. $y = \mu_2$ jsou Gaussovy křivky
- vrstevnice jsou elipsy

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

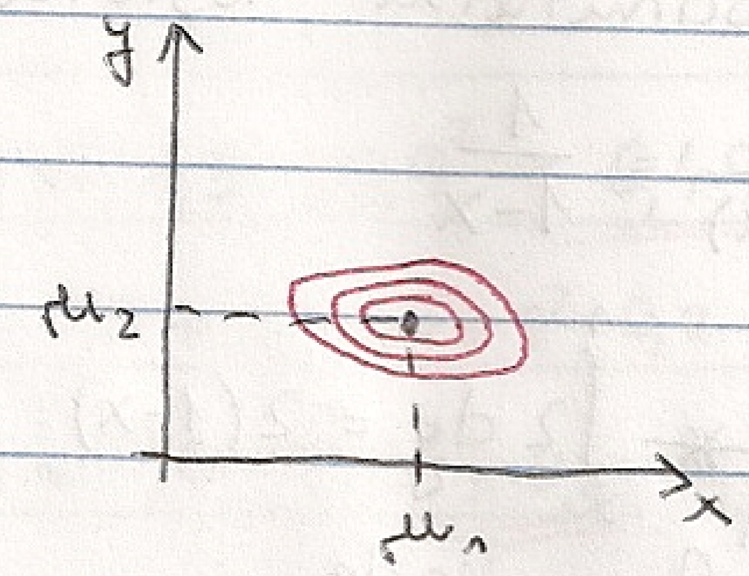
- Vztah mezi X, Y udává $\sigma_{12} = \sigma_{21} = \operatorname{cov}(X, Y)$

$X \sim N(\mu_1, \sigma_{11})$

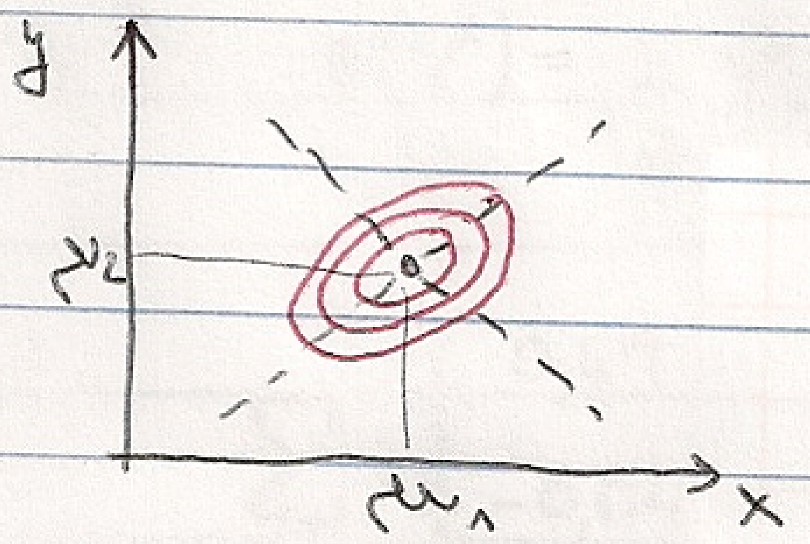
$Y \sim N(\mu_2, \sigma_{22})$



$\sigma_{12} = 0$
$\sigma_{11} = \sigma_{22}$



$\sigma_{12} = 0$
$\sigma_{11} > \sigma_{22}$



$\sigma_{12} > 0$
$\sigma_{11} > \sigma_{22}$

př: je-li $\vec{\mu} = (0,0)$
$\Sigma = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$, pak hustota pravděpodobnosti $N_2(\vec{\mu}, \Sigma)$ je

$$f(x,y) = \frac{1}{2\pi\sqrt{36}} e^{-\frac{1}{2}\left(\frac{x^2}{9} + \frac{y^2}{4}\right)}$$

$\det \Sigma = 9 \cdot 4 = 36$

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{9} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \qquad P(x,y) = (x,y)\begin{bmatrix} \frac{1}{9} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{9}x^2 + \frac{1}{4}y^2$$

zvolme $c \in \left(0, \frac{1}{12\pi}\right)$, pak $f(x,y) = c$

$$\frac{1}{12\pi} \cdot e^{-\frac{1}{2}\left(\frac{x^2}{9} + \frac{y^2}{4}\right)} = c \qquad \frac{x^2}{9} + \frac{y^2}{4} = \underbrace{-2 \ln 12\pi c}_{\text{ozn. } k}$$

zřejmě $k > 0$

$$\frac{x^2}{9k} + \frac{y^2}{4k} = 1$$

elipsa střed $S = [0,0]$

poloosy $a = 3\sqrt{k}$
$b = 2\sqrt{2}$

kde $k = -2\ln 12\pi c$, $c \in \left(0, \frac{1}{12\pi}\right)$

ÚVOD DO MATEMATICKÉ STATISTIKY

1. POPISNÁ STATISTIKA

(statistický) **soubor**: $x_1, x_2, \ldots, x_n$

[$\approx$ hodnoty (diskr. nebo spoj.) náh. veličiny]

$x_i \in R \ldots$ prvek souboru, hodnota v souboru

$n \in N \ldots$ rozsah souboru

četnost, resp. relativní četnost prvku

uspořádaný soubor: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

CHARAKTERISTIKY POLOHY

- **(aritmetický) průměr**:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Je-li n_i četnost i-té hodnoty x_i, $(i = 1, 2, \ldots, k)$ souboru, pak $n = n_1 + n_2 + \ldots + n_k$, tj. průměr lze napsat jako $\overline{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n} = \sum_{i=1}^{k} x_i w_i$, kde $w_i = \frac{n_i}{n}$ $(i = 1, 2, \ldots, k)$.

- **medián** $\tilde{x}$ (uspořádaného) souboru je jeho:
  prostřední hodnota (je-li n liché), resp.
  aritm. průměr dvou prostředních hodnot (je-li n sudé)

- **modus** $\hat{x}$ je hodnota(-y) s nejvyšší četností

CHARAKTERISTIKY VARIABILITY

- **rozpětí** je rozdíl mezi největší a nejmenší hodnotou
- **(výběrový) rozptyl**:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

- **(výběrová) směrodatná odchylka**:

$$s = \sqrt{s^2}$$

př: Je-li dán soubor: 6,4,7,7,5,4,9,7,3,8 pak např.

- četnost prvku 4 je 2, četnost 7 je 3
- rozsah = 10
- relativní četnost 4 je 0,2, rel. četnost 7 je 0,3
- průměr: $\bar{x} = \frac{1}{10}(6+4+\ldots+3+8) = 6.0$
- soubor uspořádaný vzestupně: 3,4,4,5,6,7,7,7,8,9
- medián: $\tilde{x} = \frac{1}{2}(6+7) = 6.5$
- modus: $\dot{x} = 7$

pozn.: aritmetický průměr je zavádějící, když je v souboru položka s abnormální hodnotou

př: Jestliže v souboru z prvního příkladu máme 55 místo 5, pak vzestupně uspořádaný soubor je: 3,4,4,6,7,7,7,8,9,55

$\Rightarrow$ průměr: $\bar{x} = 11.0$

medián: $\tilde{x} = \frac{1}{2}(7+7) = 7$

modus: $\dot{x} = 7$

def.: CHARAKTERISTIKA VARIABILITY

- rozpětí je rozdíl mezi největší a nejmenší hodnotou
- (výběrový) rozptyl:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{(průměrný čtverec)}$$

- (výběrová) směrodatná odchylka: $s = \sqrt{s^2}$ (průměrná vzd. od středu)

př: Najděte rozptyl a směrodatnou odchylku souboru: 2,3,5,6,9,17

Řešení: Průměr $\bar{x} = 7.0$

$$s^2 = \frac{1}{6-1}\left((2-7)^2 + (3-7)^2 + \ldots + (17-7)^2\right) = 30.0$$

$$s = \sqrt{30} \doteq 5,48$$

2. ODHADY PARAMETRŮ

náhodný výběr rozsahu n z rozdělení náhodné veličiny X je posloupnost *nezávislých* náh. veličin $X_1, X_2, \ldots, X_n$, které mají stejné rozdělení jako X

$x_1, x_2, \ldots, x_n$ je tzv. **realizace** náh. výběru (= statistický soubor)

neznámou hodnotu parametru v rozdělení X nahrazujeme vhodným reálným číslem, spočteným z realizace náh. výběru - tzv. **bodový odhad** parametru

- stř. hodnotu $E(X)$ odhadujeme aritm. průměrem:

$$E(X) \approx \bar{x}$$

odtud: $p \approx \bar{x}$ pro $X \sim A(p)$

$\lambda \approx \bar{x}$ pro $X \sim Po(\lambda)$

$\delta \approx \bar{x}$ pro $X \sim Exp(\delta)$

$\mu \approx \bar{x}$ pro $X \sim N(\mu, \sigma^2)$

- rozptyl $D(X)$ odhadujeme výběrovým rozptylem:

$$D(X) \approx s^2$$

odtud: $\sigma^2 \approx s^2$ pro $X \sim N(\mu, \sigma^2)$

intervalové odhady (neznámého) parametru θ rozdělení náh. veličiny X:

pro $\alpha \in (0,1)$ [často bývá $\alpha = 0.01,\ 0.05$ nebo 0.10]

interval (a,b) nazveme **100 (1−α) %-ní** (oboustranný) **interval spolehlivosti**, je-li

$$\boxed{P(a < \theta < b) = 1 - \alpha}$$

[a, resp. b je tzv. dolní, resp. horní mez spolehlivosti]

jednostranné intervaly spolehlivosti: $P(\theta < b) = 1 - \alpha$, resp. $P(a < \theta) = 1 - \alpha$

- je-li $x_1, x_2, \ldots, x_n$ realizace náhodného výběru z rozdělení $X \sim A(p)$, pak $100\,(1-\alpha)$%-ní **interval spolehlivosti pro parametr p** je:

$$\left(\hat{p} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

kde $\hat{p} \approx \bar{x}$, přitom n je tak velké, aby $\boxed{n\hat{p}(1-\hat{p}) \geq 9}$,

$u_{1-\frac{\alpha}{2}}$ jsou kvantily rozdělení $N(0,1)$

Idea odvození: $\underbrace{0,1,0,0,\dots 0}$

$\underline{n}$ - nezávislých pokusů

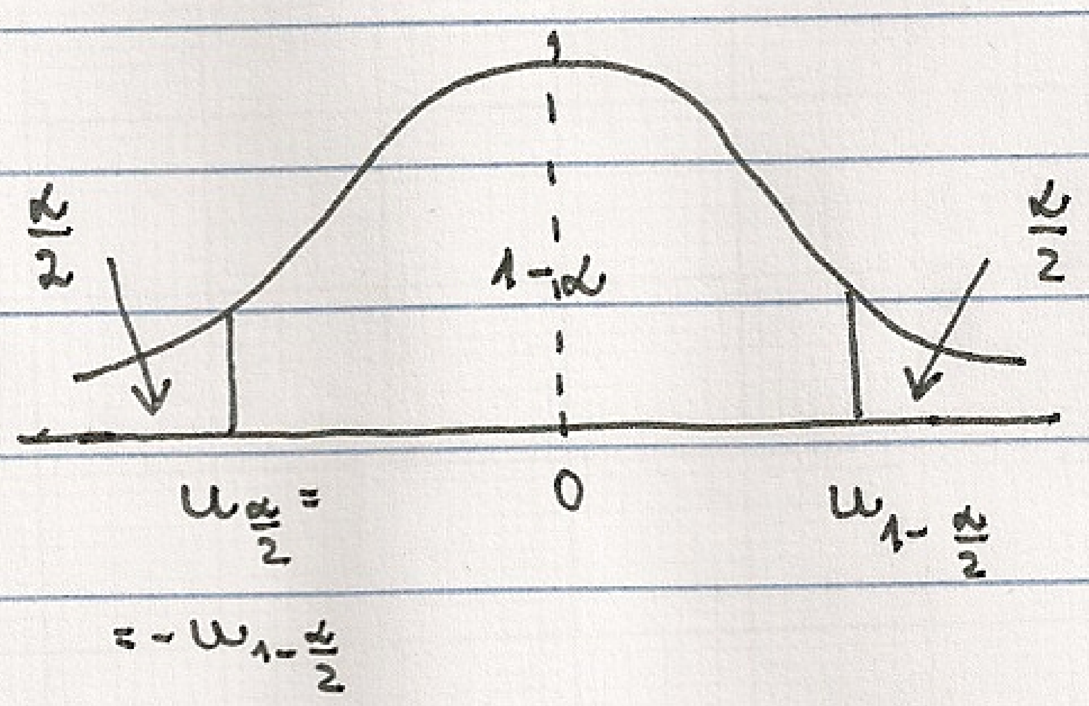
$\underline{k}$ - krát „úspěch"

$$k \sim Bi(n,p) \approx N(n\cdot p, \underbrace{n\cdot p(1-p)}_{\geq 9})$$

$$\hat{p} = \frac{k}{n} = \tilde{x} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

$$\hat{p} \approx p$$

$$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}}\cdot\sqrt{n} \sim N(0,1)$$



z def. kvantilu:

$$P\left(-u_{1-\frac{\alpha}{2}} < \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}}\cdot\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P(\hat{p} - E < p < \hat{p} + E) = 1-\alpha \quad , \text{kde} \quad E = u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

př: V nah. výběru rozsahu $n = 600$ bylo zjištěno 12 nekvalitních výrobků

Stanovte: a) 95%-ní b) 99%-ní

interval spolehlivosti pro skutečný podíl nekvalitních výrobků

Řešení: $\underbrace{0,0,1,\dots 0}_{n=600}$ (12-krát "1") $\hat{p} \pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\hat{p} = \frac{12}{600} = 0,02$ (2%) ... bodový odhad

$$\left(n\cdot\hat{p}(1-\hat{p}) = 600\cdot 0,02\cdot 0,98 = 11,76 \geq 9\right)$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0,02\cdot 0,98}{600}} \doteq 0,0057$$

a) $\alpha = 0,05 \Rightarrow 1-\frac{\alpha}{2} = 0,975$

$u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96 \Rightarrow 0,02 \pm \underbrace{1,96\cdot 0,0057}_{0,011}$ $\quad$ 2% ± 1,1%

tj. 0,9% ~ 3,1%

s 95%-ní pravděp.

b) $\alpha = 0{,}01$ $\qquad$ $2\% \pm 1{,}5\%$

$u_{1-\frac{\alpha}{2}} = u_{0{,}995} = 2{,}576 \Rightarrow 0{,}02 \pm \underbrace{2{,}576 \cdot 0{,}0057}_{0{,}015}$ $\qquad$ $0{,}5\% \sim 3{,}5\%$

s 99%-ní ppstí

př: V anketě odpovědělo z 50 respondentů

20 osob ANO (1)

30 osob NE (0)

Označme p parametr příslušného rozdělení $A(p)$

a) určete 95%-ní interval spolehlivosti

b) určete počet respondentů takový, aby délka 95%-ního int. spolehlivosti byl menší než 0,06 $(\hat{p} \pm 3\%)$

<u>Řešení:</u> a) $\hat{p} = \frac{20}{50} = 0{,}4$

$m \cdot \hat{p}(1-\hat{p}) = 50 \cdot 0{,}4 \cdot 0{,}6 = 12 \ (\geq 9)$

TAB.

$\alpha = 0{,}05 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}975 \Rightarrow u_{0{,}975} = 1{,}96$

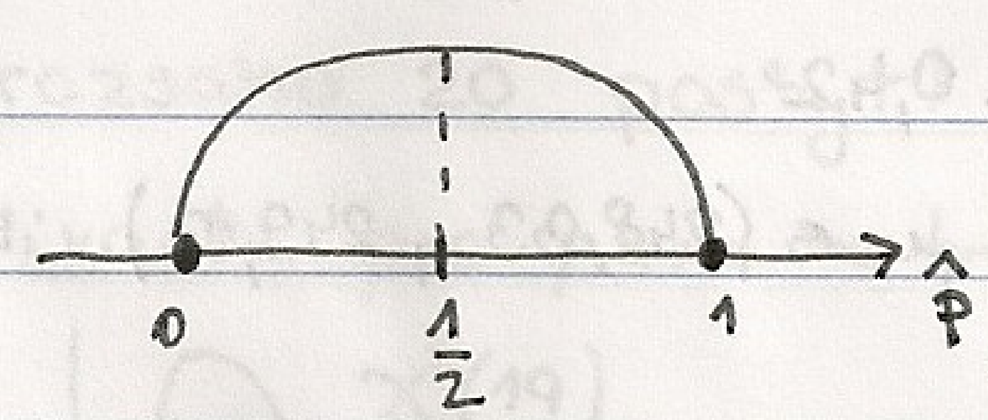
$$\left(0{,}4 - 1{,}96 \cdot \sqrt{\frac{0{,}4 \cdot 0{,}6}{50}} \ ; \ 0{,}4 + 1{,}96 \cdot \sqrt{\frac{0{,}4 \cdot 0{,}6}{50}}\right) = (0{,}26 ; 0{,}54)$$

tj. $0{,}4 \pm 0{,}134$ s 95% ppstí

b) $\hat{p}$ se může s rostoucím $\underline{m}$ měnit

$\underbrace{\hat{p}(1-\hat{p})}_{\hat{p} - \hat{p}^2}$ odhadneme maximem

to nastává pro $\hat{p} = \frac{1}{2}$



$(E \Rightarrow) \ u_{0{,}975} \sqrt{\frac{\hat{p}(1-\hat{p})}{m}} < 0{,}03 \Rightarrow$

$\Rightarrow m > \left(\frac{1{,}96}{0{,}03}\right)^2 \cdot \frac{1}{2} \cdot \frac{1}{2} \doteq 1067{,}\overline{1} \doteq 1068$

Aspoň 1068 respondentů.

- je-li $x_1, x_2, \ldots, x_n$ realizace náhodného výběru z rozdělení $X \sim N(\mu, \sigma^2)$, kde $\boxed{n > 30}$, pak $100\,(1-\alpha)$%-ní **interval spolehlivosti pro parametr** μ je:

$$\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \,,\; \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

kde $u_{1-\frac{\alpha}{2}}$ jsou kvantily rozdělení $N(0,1)$

Pozn.: pro $\boxed{n \leq 30}$ tento interval spolehlivosti je:

$$\left(\bar{x} - t_{\nu,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \,,\; \bar{x} + t_{\nu,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

$t_{\nu,1-\frac{\alpha}{2}} \ldots$ kvantily t-rozdělení s ν stupni volnosti, kde $\nu = n - 1$

t-rozdělení, též **Studentovo rozdělení**, $t(\nu)$, je typem spojitého rozdělení ppsti. Křivka, která je grafem hustoty t-rozdělení, má "obdobný" tvar a symetrii jako normální rozdělení (kde $\mu = 0$), ale odráží větší "tvárnost" prostřednictvím *stupně volnosti*, označovaném ν.

Tabulky obsahují jen kvantily t-rozdělení pro běžné volby α.

Pro $n > 30$ se používá aproximace $t(\nu) \approx N(0,1)$.

50-ti

př: převážením údajně čtvrtkilových balíčků soli jsme dostali $\bar{x} = 248{,}75$ [g]

$s = 1{,}52$ [g]

Stanovte 95%-ní interval spolehlivosti pro skutečnou hmotnost výrobků. (Předpokládáme výběr $N(\mu, \sigma^2)$)

Řešení: $n = 50$ (>30)

$\alpha = 0{,}05 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0{,}975} = 1{,}96$ $\qquad u_{0{,}975} \cdot \frac{s}{\sqrt{n}} = 1{,}96 \cdot \frac{1{,}52}{\sqrt{50}} \doteq 0{,}42$

$248 \pm 0{,}42$

tj. $\mu \in (248{,}33 \,;\, 249{,}17)$ s 95%-ní pravděp.

$(\bar{x} - E, +\infty)$ $\quad$ $98{,}55 - 2{,}539 \cdot \dfrac{0{,}84}{\sqrt{20}} \doteq$ $\quad$ $\mu \in (98{,}07\,;\, +\infty)$

$E = 2{,}539$ $\qquad$ $= 98{,}55 - 0{,}48$ $\qquad$ aneb $\mu > 98{,}07\ [°C]$ (s 99%-ní pravděpodobností

- je-li $x_1, x_2, \ldots, x_n$ realizace náhodného výběru z rozdělení $X \sim N(\mu, \sigma^2)$,

  pak $100\,(1-\alpha)$%-ní **interval spolehlivosti pro parametr** σ^2 je:

$$\left(\frac{(n-1)\,s^2}{\chi^2_{\nu,\,1-\frac{\alpha}{2}}},\ \frac{(n-1)\,s^2}{\chi^2_{\nu,\,\frac{\alpha}{2}}} \right)$$

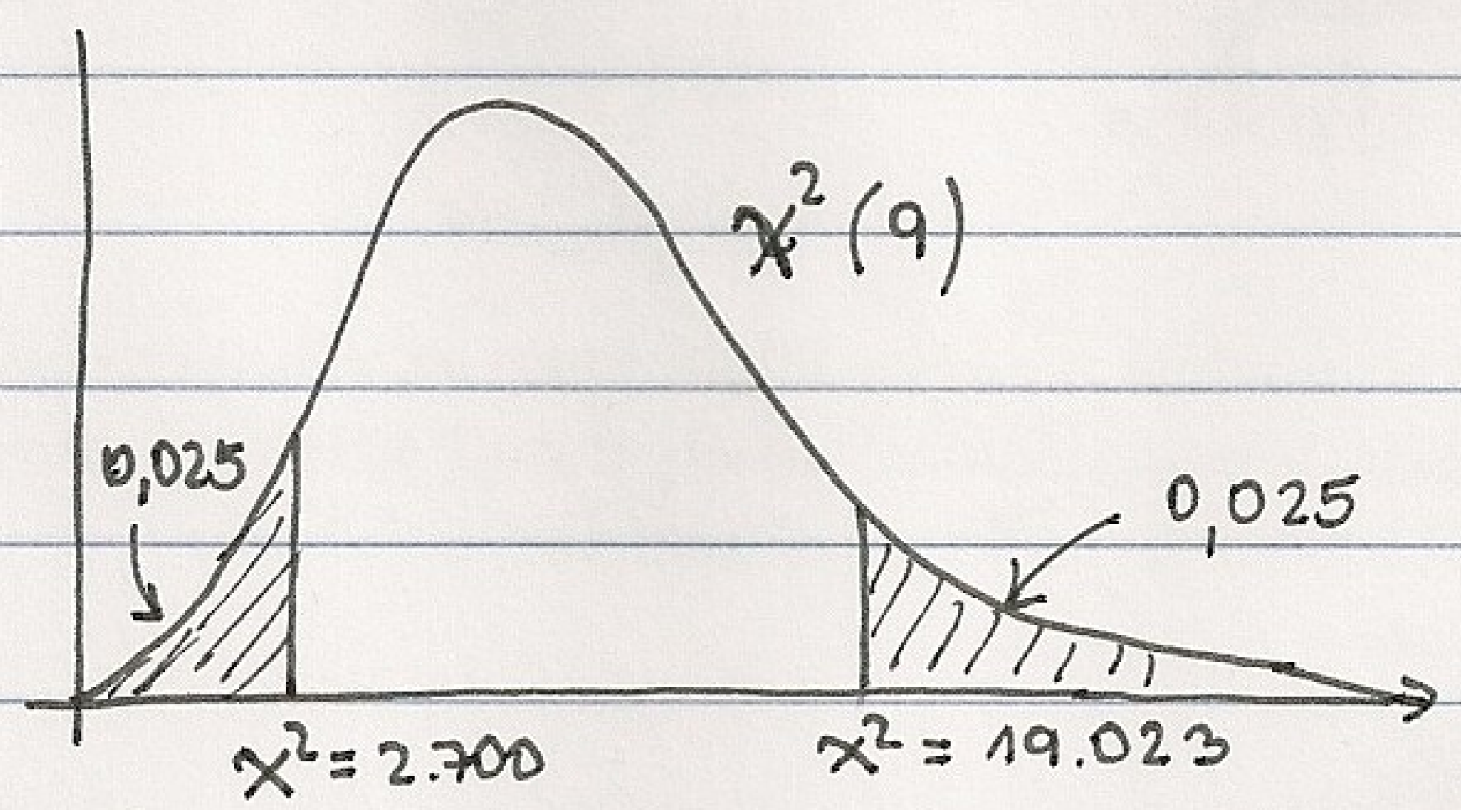
  kde $\chi^2_{\nu,\,1-\frac{\alpha}{2}}$, $\chi^2_{\nu,\,\frac{\alpha}{2}}$ jsou kvantily χ^2-rozdělení

  s ν stupni volnosti, kde $\nu = n - 1$

chí-kvadrát rozdělení, $\chi^2(\nu)$, je typem spojitého rozdělení. Křivka, která je grafem hustoty χ^2-rozdělení, je identicky nulová pro všechny nekladné argumenty, a nemá symetrický tvar. Tvar hustoty χ^2-rozdělení závisí na čísle zvaném *stupeň volnosti*, označovaném ν.

Tabulky obsahují jen kvantily χ^2-rozdělení pro běžné volby α.

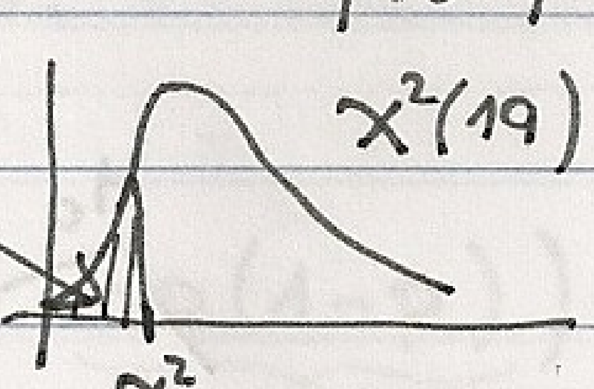
např:



př: Náh. výběr z rozdělení $N(\mu, \sigma^2)$ rozsahu 20 poskytl $s = 0{,}84$ [°C].

Stanovte 95% horní mez spolehlivosti pro parametr σ^2

Řešení: $\alpha = 0{,}05 \longleftarrow \alpha$



$n = 20 \Rightarrow \nu = n - 1 = 19$

$\chi^2_{19,\,0.05} = 10{,}1$ tabulky

horní mez: $\dfrac{(n-1)\cdot s^2}{\chi^2_\alpha} = \dfrac{19 \cdot 0{,}84^2}{10{,}1} =$

$\doteq 1{,}33$

$\sigma^2 < 1{,}33$ (s 95%-ní pravděp.)

3. TESTOVÁNÍ HYPOTÉZ

Použitím naměřených hodnot testujme určitou (statistickou) hypotézu, označme ji H_0.

Náš závěr bude vždy jeden z následujících dvou:

1. Zamítáme H_0.
2. Nezamítáme H_0.

Můžeme dopustit chyb dvou druhů:

chyba 1. druhu: H_0 je pravdivá, ale test vede k zamítnutí H_0;

chyba 2. druhu: H_0 je nepravdivá, ale test vede k nezamítnutí H_0.

Ppst chyby 1. druhu se značí α, požaduje se malá a volí se před testem (běžně: 0.1, 0.05, 0.01)

$\alpha \ldots$ tzv. **hladina významnosti**

Ppst chyby 2. druhu se označuje β (závisí na volbě α).

Rozhodnutí o případném zamítnutí H_0 provádíme pomocí testu zvaného **testové kritérium**. Množina všech hodnot test. kritéria vedoucí k zamítnutí H_0, se nazývá **kritický obor**, označuje se W. Hodnota, která odděluje W od hodnot, které vedou k nezamítnutí H_0, je tzv. **kritická hodnota**

Obecný postup:

1) Stanovíme H_0 - tzv. **nulová hypotéza**

2) Stanovíme H_1 - tzv. **alternativní hypotéza**

3) Zvolíme **hladinu významnosti** α.

4) Vybereme vhodné test. kritérium, stanovíme jeho rozdělení ppsti za předpokladu platnosti H_0. Toto rozdělení, hladina význ. a formulace H_1 určují **kritický obor** W (kritickými hodnotami jsou pak příslušné kvantily rozdělení test.kritéria). Načrtneme graf rozdělení test. kritéria, vyznačíme W.

5) Hodnota test. kritéria $\in W \Rightarrow$ **zamítáme** H_0.
Hodnota test. kritéria $\notin W \Rightarrow$ **nezamítáme** H_0.

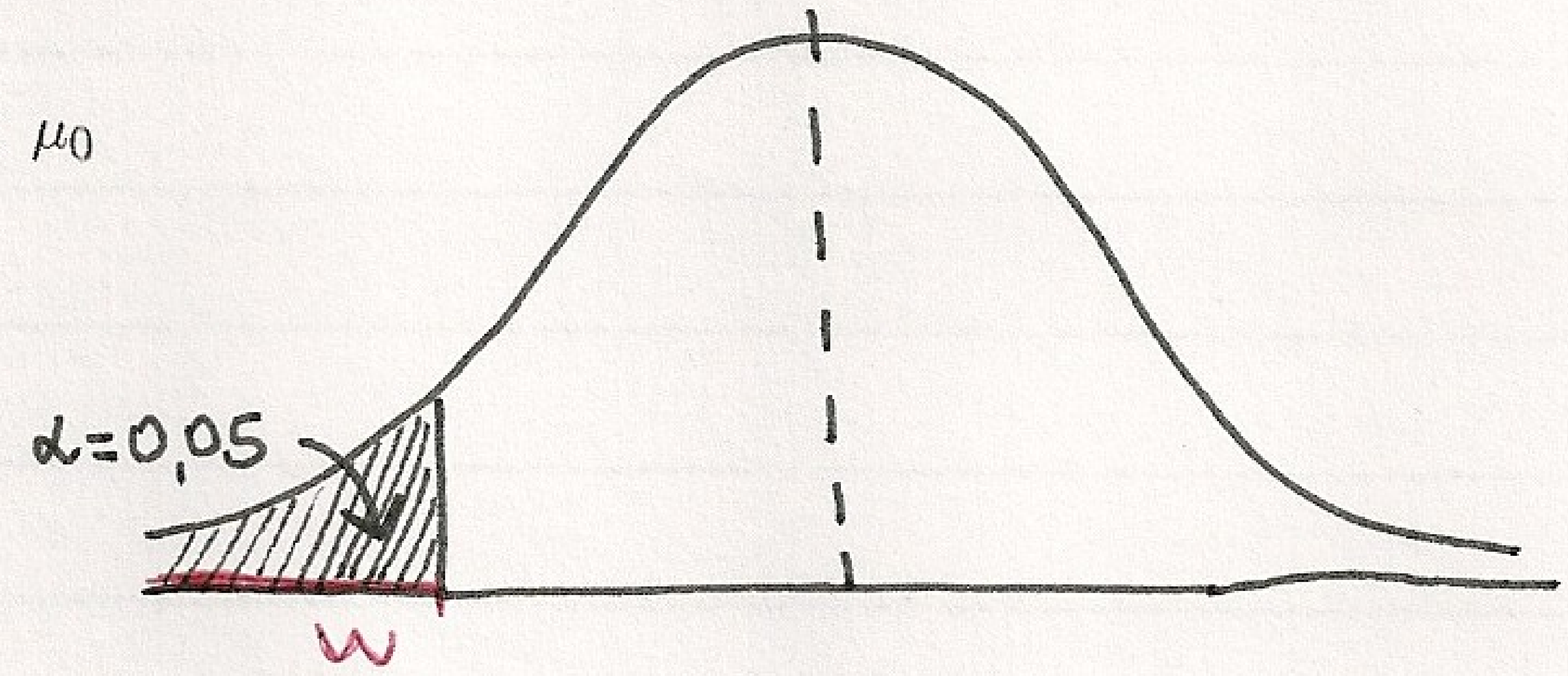
H_0

H_0		nezamítáme	zamítáme
	pravdivá	–	α
	nepravdivá	β	–

Podle formulace H_1 dostáváme testy:

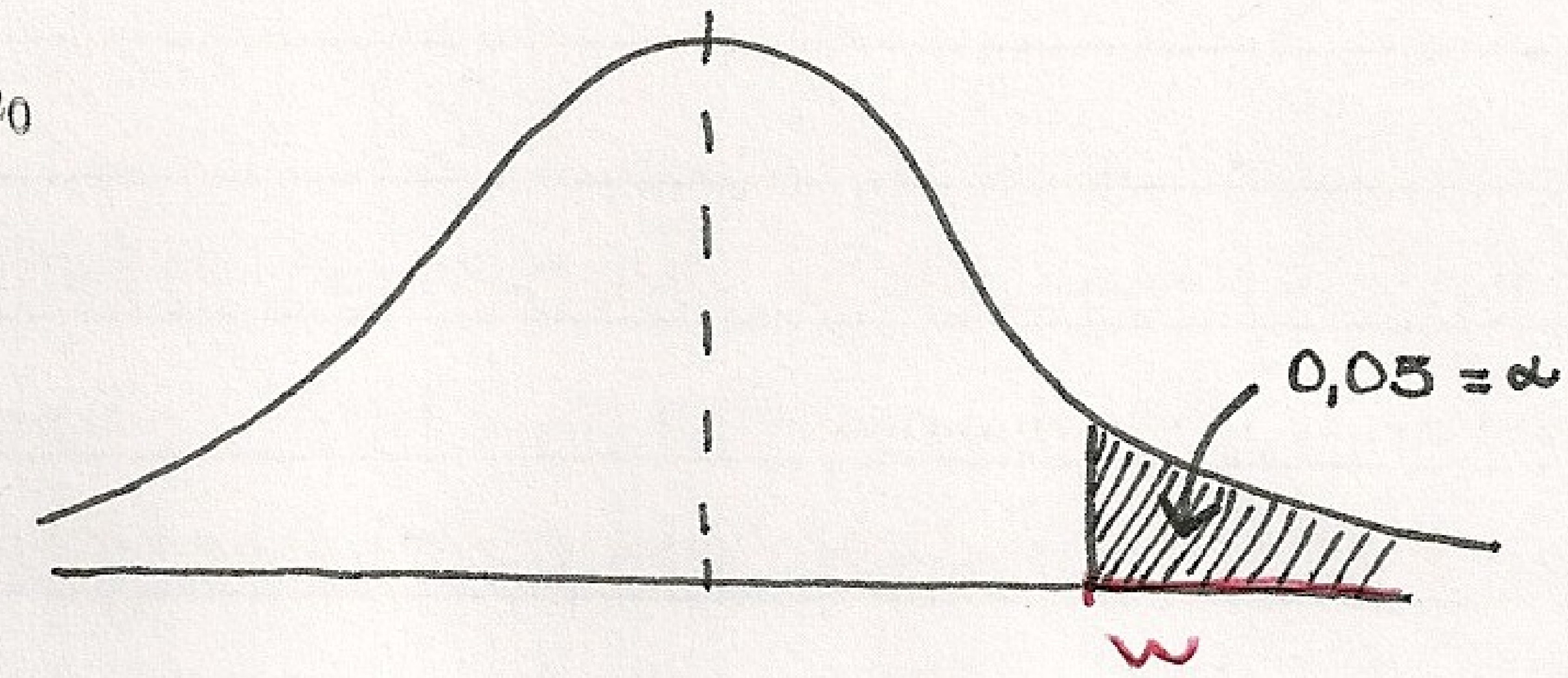
- **levostranné:**

např. $H_1 : \mu < \mu_0$



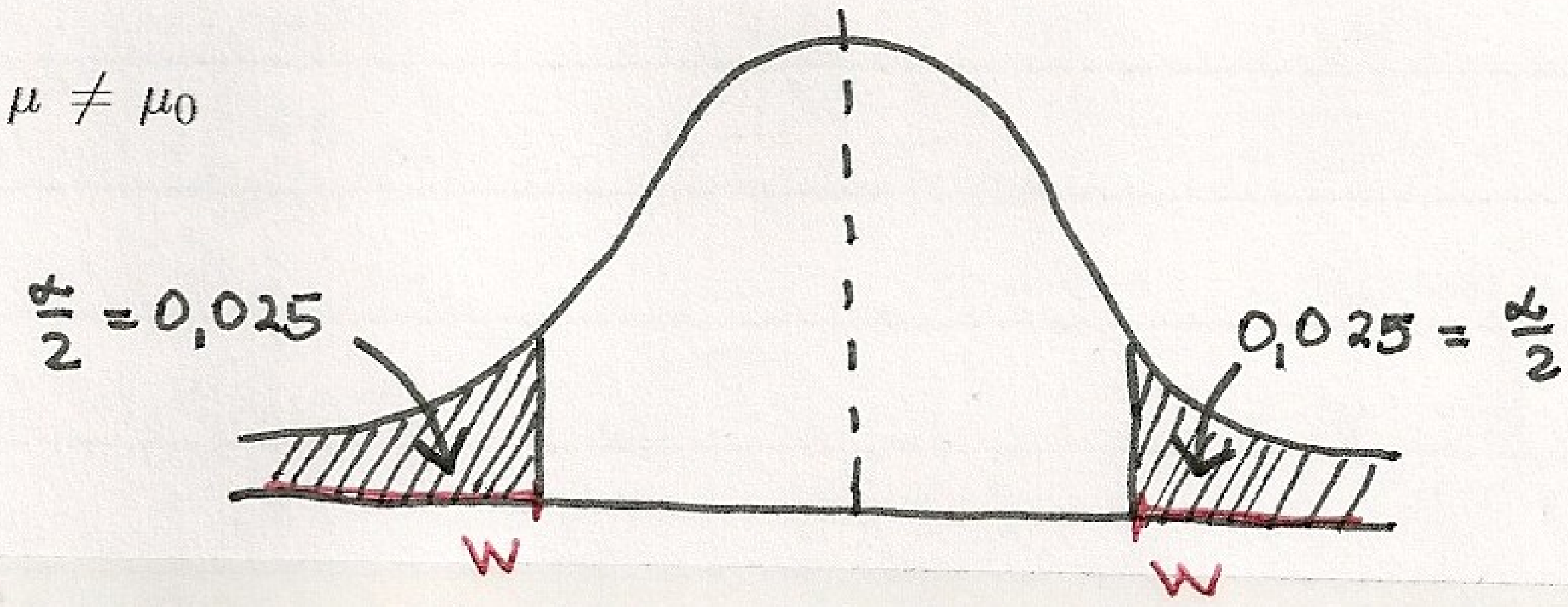
- **pravostranné:**

např. $H_1 : \mu > \mu_0$



- **oboustranné:**

např. $H_1 : \mu \neq \mu_0$



$x_1, \dots x_m$ $\quad X \sim A(p)$, nezáv.

$$X = \sum_{i=1}^{m} X_i \sim Bi(m, p) \approx N(mp, \underbrace{mp(1-p)}_{\geq 9})$$

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{m}\right) \qquad U = \frac{\bar{X} - p}{\sqrt{p(1-p)}} \cdot \sqrt{m} \sim N(0,1)$$

Předpokládejme, že známe typ rozdělení, z něhož pocházejí hodnoty v souboru, a testujeme jen neznámé hodnoty parametrů. Testujeme tzv. **PARAMETRICKÉ HYPOTÉZY**
- viz následující testy:

Test parametru p rozdělení $A(p)$

Používáme testové kritérium

$$u = \frac{\hat{p} - p}{\sqrt{p\,(1-p)}}\,\sqrt{n}\,,$$

kde $\hat{p} = \bar{x}$ je odhad parametru p.

p je testovaný parametr (daný v H_0)

n je rozsah souboru (test požaduje n tak velké, aby $n\hat{p}(1-\hat{p}) \geq 9$)

je-li H_0 pravdivá, pak $u \approx N(0,1)$

př: Výrobní linka musí být seřízena, přesáhne-li její zmetkovitost 3 %. Náhodný výběr rozsahu 400 obsahovat 18 zmetků. Na hladině významnosti $\alpha = 0{,}05$ testujte hypotézu, že skutečná zmetkovitost přesahuje 3%.

Řešení: $H_0: p \leq 0{,}03$ $\qquad n = 400$

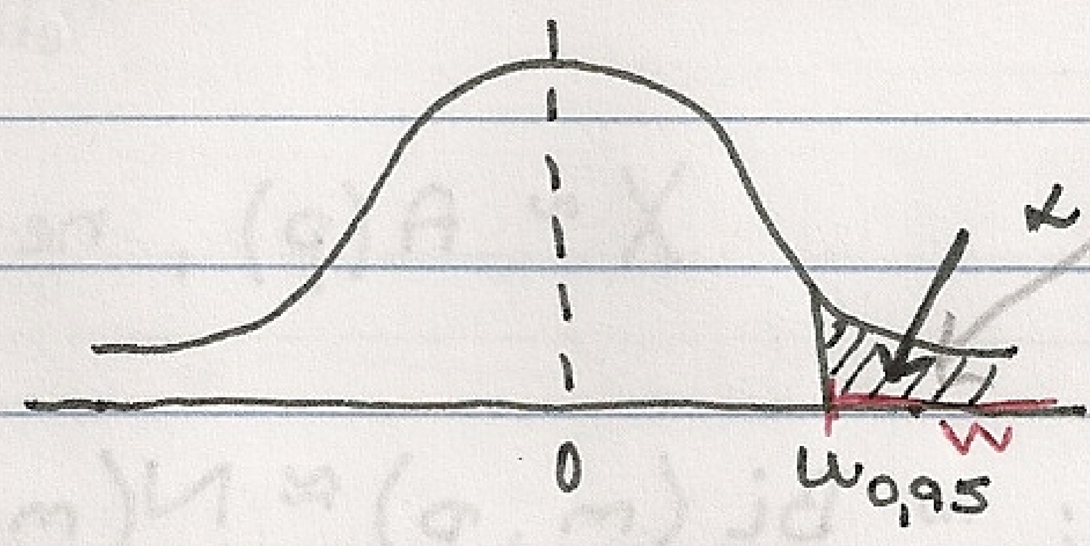
$\times$ $\qquad \hat{p} = \frac{18}{400} = 0{,}045$ (4,5%)

→ $H_1: p > 0{,}03$

→ $\alpha = 0{,}05$ $\qquad [\underbrace{400 \cdot 0{,}045}_{18} \cdot 0{,}955 \geq 9]$

hodnota testovacího kritéria: $u = \frac{0{,}045 \cdot 0{,}03}{\sqrt{0{,}03 \cdot 0{,}97}} \cdot \sqrt{400} \doteq 1{,}759$

platí-li H_0, pak $u \sim N(0,1)$



kritická hodnota $u_{0,95} = 1{,}645$

W – kritický obor

$W = (1{,}645 ; +\infty)$

$1{,}759 \in W \Rightarrow$ zamítáme H_0

(s 95%-ní pravděp. je zmetkovitost větší než 3%)

Test parametru μ rozdělení $X \sim N(\mu, \sigma^2)$, tzv. **t-test**

Testové kritérium:

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

Je-li H_0 pravdivá , pak t má t-rozdělení ppsti s $n-1$ stupni volnosti, tj.

$$t \sim t(\nu), \quad \text{kde } \nu = n - 1.$$

Pozn.: Pro velký rozsah n souboru (řekněme $n > 30$), platí $t(\nu) \approx N(0,1)$. Kritickými hodnotami jsou pak příslušné kvantily u_p rozdělení $N(0,1)$.

př.: 10 naměřených hodnot ($\approx$ realiazace náh. veličiny $N(\mu, \sigma^2)$) poskytlo $\bar{x} = 98{,}85$, $s = 1{,}77$

Na hladině významnosti $\alpha = 0{,}05$ testujte hypotézu $\mu = 100$ (proti alternativě $\mu \neq 100$).

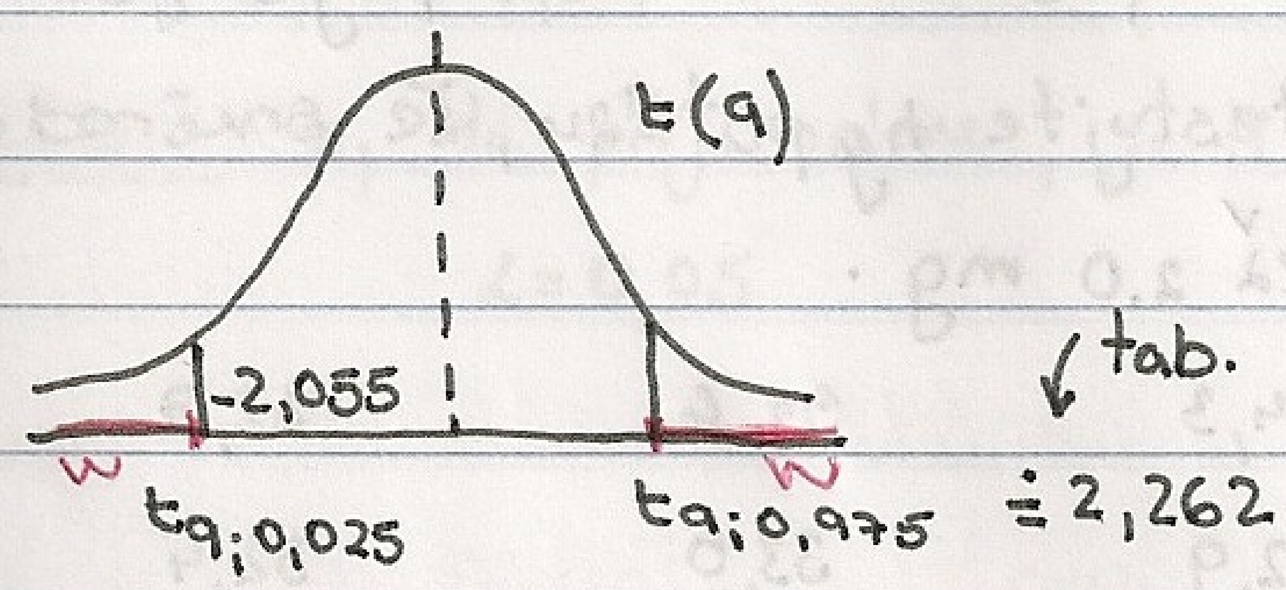
Řešení: $H_0: \mu = 100$ $\quad$ ($n = 10 < 30$)

$\times$ $\quad$ $t = \dfrac{98{,}85 - 100}{1{,}77} \cdot \sqrt{10} \doteq -2{,}055 \notin W$

$H_1: \mu \neq 100$ $\quad$ $\Rightarrow$ nezamítáme

$\alpha = 0{,}05$

platí-li H_0: pak $t \sim t(9)$ $\quad$ $\nu = n - 1 = 10 - 1 = 9$



$W = \{t;\ |t| > 2{,}262\}$

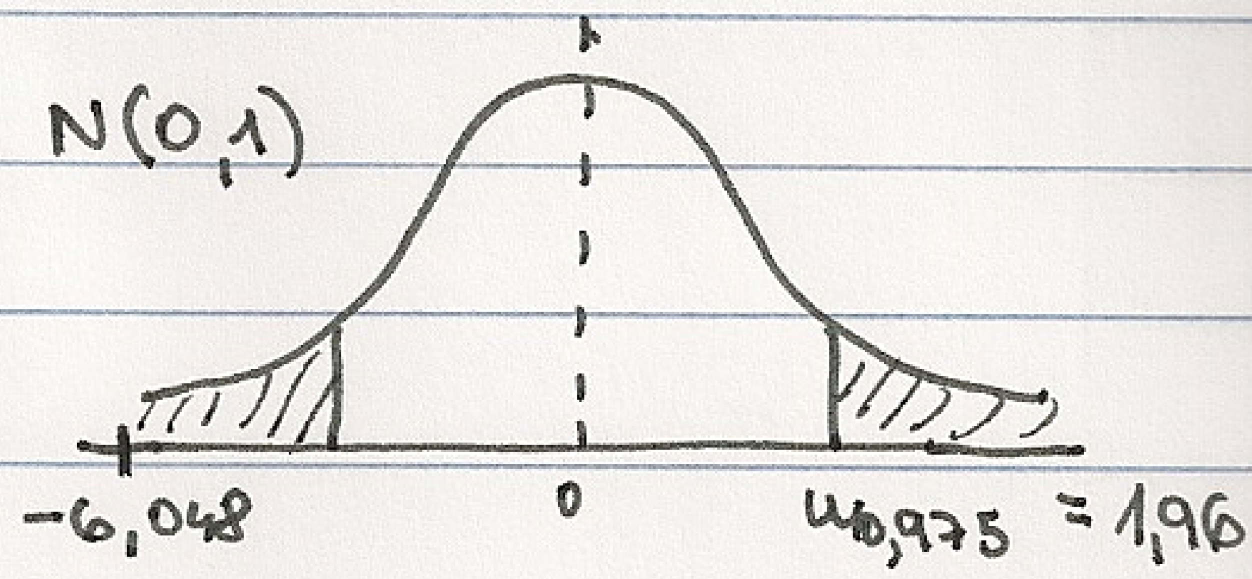
př: Náhodný výběr a převážení 50 balíčků údajně čtvrtkilových balení kávy: $\bar{x} = 248{,}7$ [g] (aritm. průměr)

$s = 1{,}52$ [g] (výběrová směrodatná odchylka)

Na hladině významnosti $\alpha = 0{,}05$ testujte hypotézu

$H_0: \mu = 250$ [g]

Řešení: $H_1: \mu \neq 250$ [g] $n = 50 > 30$ $N(0,1)$



$$u = \frac{248{,}7 - 250}{1{,}52} \cdot \sqrt{50} = \underline{-6{,}048} \in W$$

$$W = \{u;\ |u| > 1{,}96\}$$ $\Rightarrow$ zamítáme H_0 ($\alpha = 0{,}05$)

Test parametru σ^2 rozdělení $X \sim N(\mu, \sigma^2)$

Testové kritérium:

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2},$$

kde n je rozsah souboru,
s^2 je výběrový rozptyl,
σ^2 je testovaný rozptyl (daný v H_0).

Je-li H_0 pravdivá, pak χ^2 má χ^2-rozdělení ppsti s $n-1$ stupni volnosti, tj.

$$\chi^2 \sim \chi^2(\nu), \quad \text{kde } \nu = n - 1.$$

př: Zjištěné obsahy kofeinu (v mg) v 12 plechovkách nápoje jsou uvedeny níže. Na hladině význ. 0,05 testujte hypotézu, že směrodatná odchylka obsahu kofeinu je menší než 2.0 mg.

33,7	34,2	31,9	34,3	31,6	32,7
33,1	35,2	31,6	32,9	33,0	32,4

Řešení: $H_0: \sigma^2 \geq (2.0)^2$

$H_1: \sigma^2 < (2{,}0)^2$

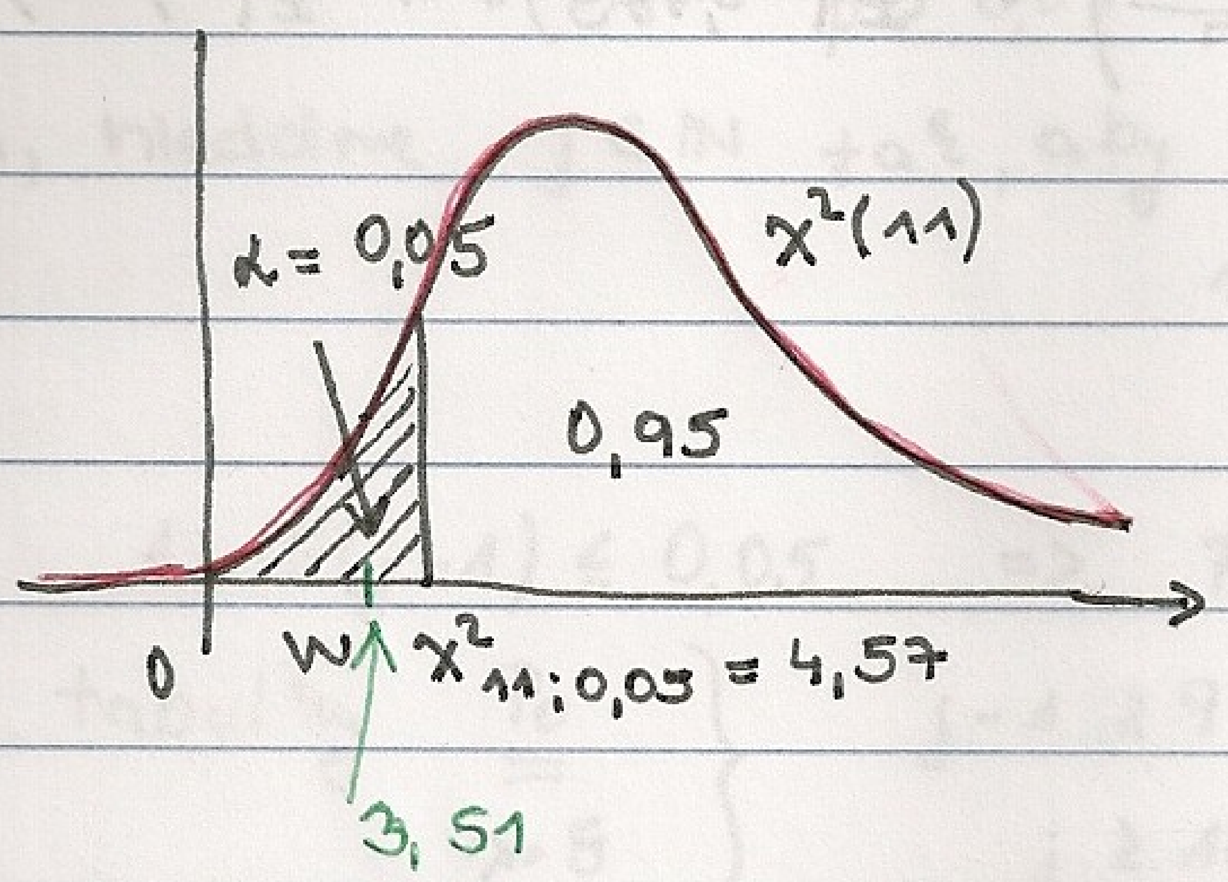
$\alpha = 0{,}05$

$n = 12$ $\bar{x} = \frac{1}{12}(33{,}7 + \ldots 32{,}4) = 33{,}05$

$$s = \sqrt{\frac{1}{11}\left[(33{,}7 - 33{,}05)^2 + \ldots + (32{,}4 - 33{,}05)^2\right]} \doteq 1{,}13$$

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2} = \frac{(12-1)(1,13)^2}{(2,0)^2} \doteq 3,51$$

Test je levostranný (H_0 zamítneme pro malé χ^2), vzhledem k $\alpha = 0,05$, $\nu = 12-1 = 11$ je kritickou hodnotou $\chi^2_{11;\,0,05} = 4,57$



kritický obor: $W = (0, 4,57)$

$3,51 \in W \Rightarrow$ zamítáme H_0

př: Mějme náh. výběr rozsahu $n > 30$ z rozdělení $N(\mu, \sigma^2)$

Nechť $n = 100$, $\quad t(\nu) \approx N(0,1)$

$s^2 = 16$.

Pro test $H_0: \mu \leq 50 \quad \times \quad H_1: \mu > 50$

spočtěte pravděp. z chyby 2. druhu, je-li ve skutečnosti $\mu = 51$

Volte: a) $\alpha = 0,05$

b) $\alpha = 0,01$

<u>Řešení</u>: a) H_0 přijímáme, je-li $u = \frac{\bar{x}-\mu}{s} \cdot \sqrt{n} < u_{0,95}$

$\alpha = 0,05$

$$\text{tj. } \frac{\bar{x}-50}{\sqrt{16}} \cdot \sqrt{100} < 1,645$$

$$\bar{x} < 50 + \frac{1,645 \cdot 4}{10}$$

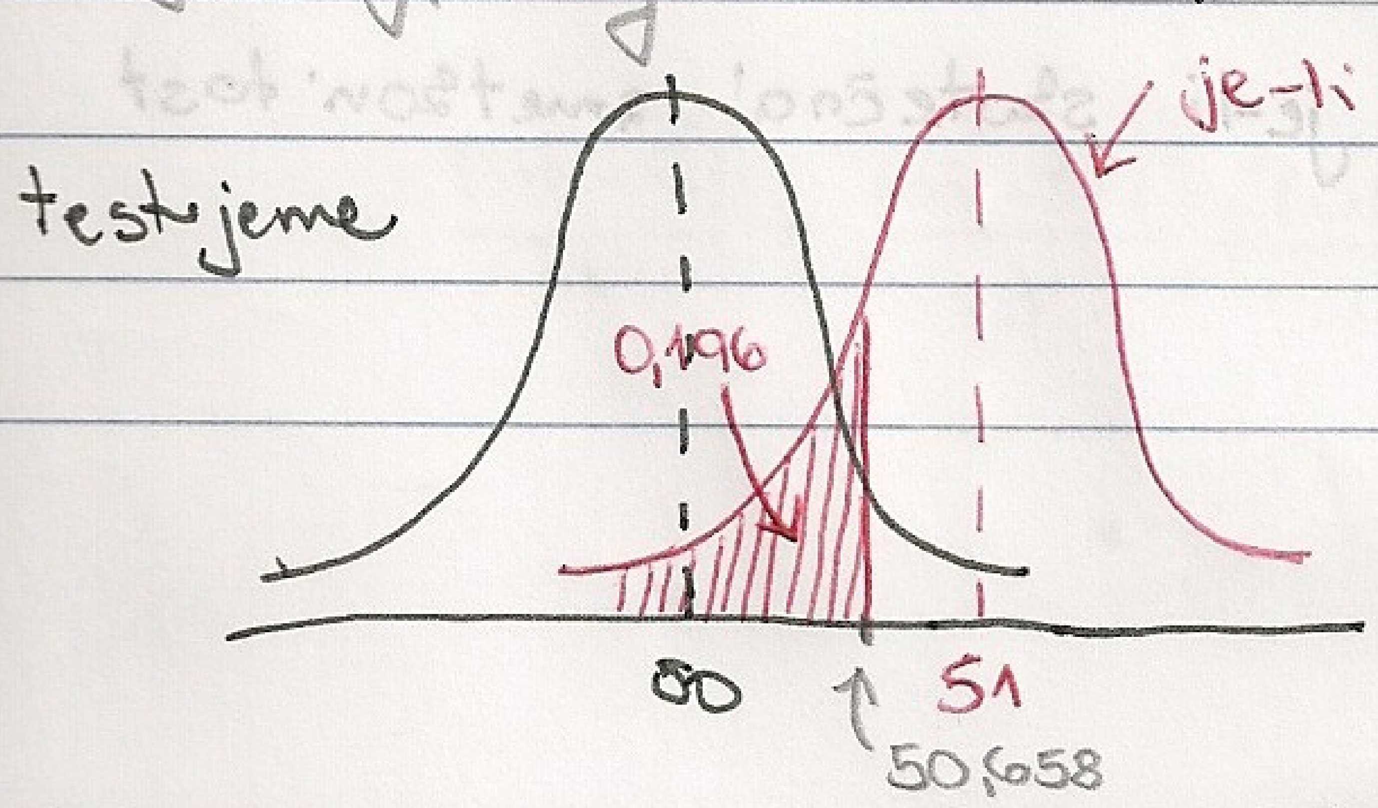
$$\bar{x} < 50,658$$

přitom $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \sim N(51;\, 0,16)$

pravděp. přijetí hypotézy

$$\beta = P(\bar{x} < 50,658) = F(50,658) = \Phi\left(\frac{50,658-51}{\sqrt{0,16}}\right) = 1 - \Phi(0,855) =$$

$$= 1 - 0,804 \doteq 0,196 \doteq 19,6\,\%$$



chyba 2. druhu je cca 20 %.

b) $\alpha = 0{,}01$ H_0 přijímáme, je-li $u = \frac{\bar{x} - \mu}{s} \cdot \sqrt{m} < u_{0,99} = 2{,}326$

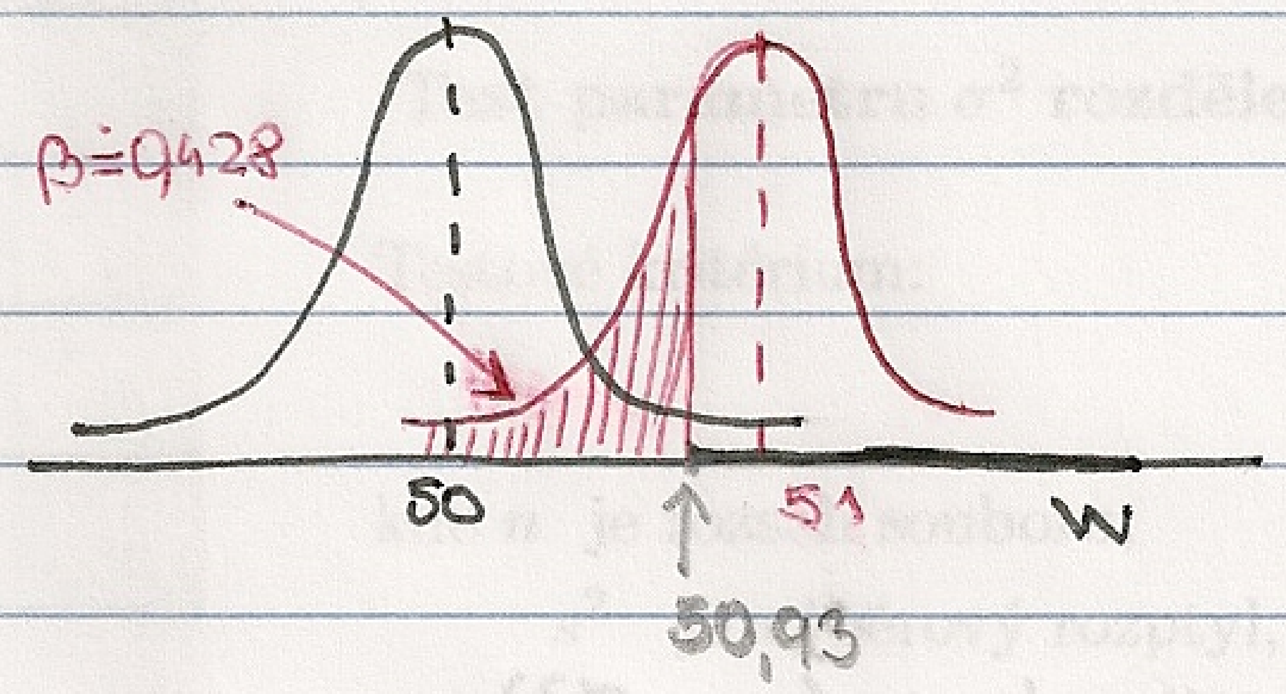
$$\bar{x} < 50 + \frac{2{,}326 \cdot 4}{10}$$

$$\bar{x} < 50{,}93$$

$\beta = P(\bar{x} < 50{,}93) \doteq F(50{,}93) = \Phi\left(\frac{50{,}93 - 51}{\sqrt{0{,}16}}\right) \doteq \Phi(-0{,}175) = 1 - \Phi(0{,}175) \doteq$

$\doteq 1 - 0{,}572 \doteq 0{,}428$

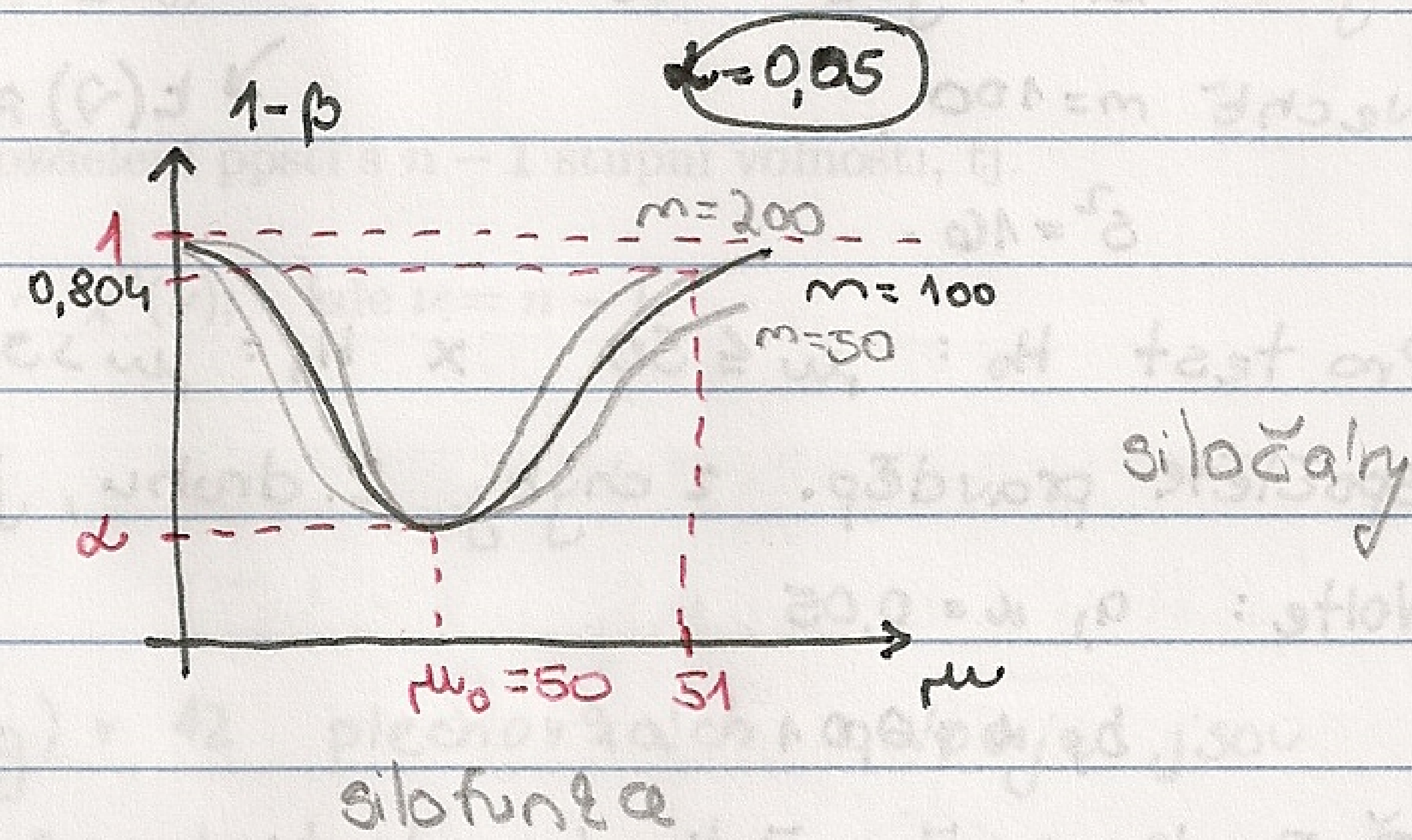
Chyba 2. druhu je cca 43%.



$1-\beta$ ... síla testu

$\alpha = 0{,}05 \Rightarrow 1-\beta = 0{,}804$

$\alpha = 0{,}01 \Rightarrow 1-\beta = 0{,}572$



př: V zásilce velkého rozsahu má být nejvýše 1% zmetků. Při příjímce bude vybráno a zkontrolováno 500 výrobků. Dodavatel vyžaduje, aby pravděp. vyhovující zásilky byla ~~[illegible]~~ $\leq 0{,}05$

a) Při jakém počtu zmetků (z 500 kontrolovaných) může odběratel zásilku odmítnout?

b) Jaká je pravděp. přijetí zásilky, je-li skutečná zmetkovitost zásilky 2%, resp. 3%?

Řešení: X … počet zmetků (z 500)

$$X \sim Bi(\underset{n\geq 30}{500};\ \underset{p<0{,}1}{0{,}01}) \approx Po(\underset{\lambda=n\cdot p}{5})$$

$p = 0{,}01$ $\quad H_0:\ p \leq 0{,}01$ $\quad \alpha = 0{,}05$ ⇐ hladina významnosti

$\times$

$H_1:\ p > 0{,}01$

a) hledáme $j \in \mathbb{N}$ tak, aby $P(X \doteq j) \leq 0{,}05$

$$1 - P(X \leq j) \leq 0{,}05$$

diskrétní rozdělení

$$1 - F(j-1) \leq 0{,}05 \Rightarrow F(j-1) \geq 0{,}95$$

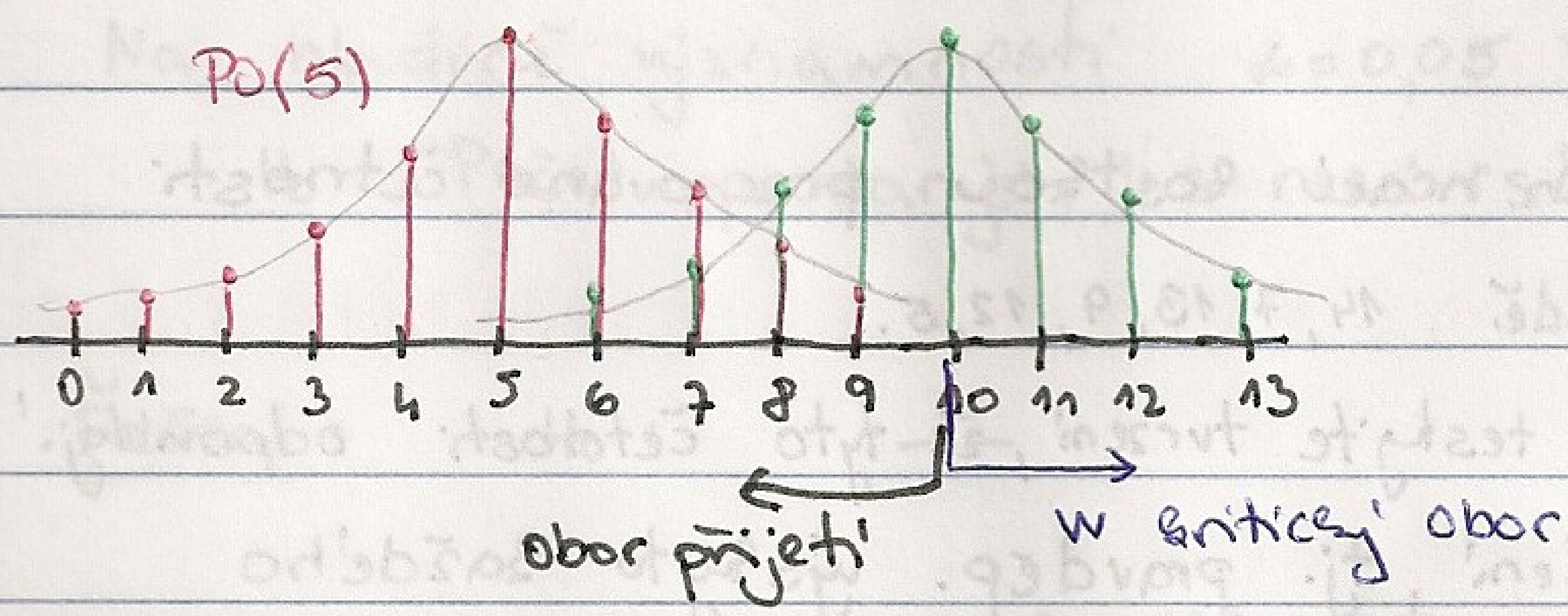
tabulky $\underset{\lambda=5}{Po}$ $\Big\}$ $j - 1 \geq 9$

$j \geq 10$

$W = \{10, 11, 12 \dots\}$

Odběratel může obj. zamítnout, pokud v ní najde 10 nebo více zmetků (z 500 kontrolovaných).

b) je-li $p = 0{,}02$, pak $X \sim Bi(500;\ 0{,}02) \approx Po(10)$



$B = P(X \leq 9) = F(9) = \boxed{0{,}458}$ (riziko odběratele)
↑ tab.

je-li $p = 0{,}03$, pak $X \sim Bi(500;\ 0{,}03) \approx Po(15) \approx N(15;\ 15)$

$$B = P(X \leq 9) = F(9) \doteq F_N(9) = \Phi\left(\frac{9-15}{\sqrt{15}}\right) = 1 - \Phi(1{,}55) \overset{\text{tab.}}{\doteq} 1 - 0{,}94 \doteq 0{,}06$$

$\doteq 6\%$

25.11.

Jiný druh hypotéz jsou tzv. **NEPARAMETRICKÉ HYPOTÉZY** - - např. testy, zda naměřené hodnoty "shodují" s určitým typem rozdělení.

Často používaný je **chí-kvadrát test dobré shody**

n ... **rozsah** souboru - tj. počet všech (ne nutně různých) naměřených hodnot experimentu

k ... **všechny možné výsledky** experimentu

pro $i = 1, 2, \ldots, k$ označme: n_i ... **pozorovaná četnost** i-tého výsledku,
n_i^O ... **očekávaná četnost** i-tého výsledku.

testové kritérium:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^O\right)^2}{n_i^O}$$

Je-li n tak velké, $n_i^O \geq 5 \;\; \forall\, i = 1, \ldots, k$, pak

$$\boxed{\chi^2 \approx \chi^2(\nu)\,, \text{ kde } \nu = k - 1}$$

K zamítnutí H_0 : *"Není (významný) rozdíl mezi pozorovanými a očekávanými četnostmi* vede "velká" hodnota χ^2, tj. test je **pravostranný**.

Př: Předpokládejme, že po 60 hodech hrací kostkou pozorované četnosti n_i čísel $i = 1,2,3,4,5,6$ byly po řadě 14, 7, 13, 9, 12, 5.
Na hladině významnosti $\alpha = 0{,}05$ testujte tvrzení, že tyto četnosti odpovídají diskrétnímu rovnoměrnému rozdělení, tj. pravděp. výskytu každého z čísel 1 až 6 je stejná.

Řešení: H_0 : „Pozorované četnosti se shodují s distr. rovn. rozdělením"

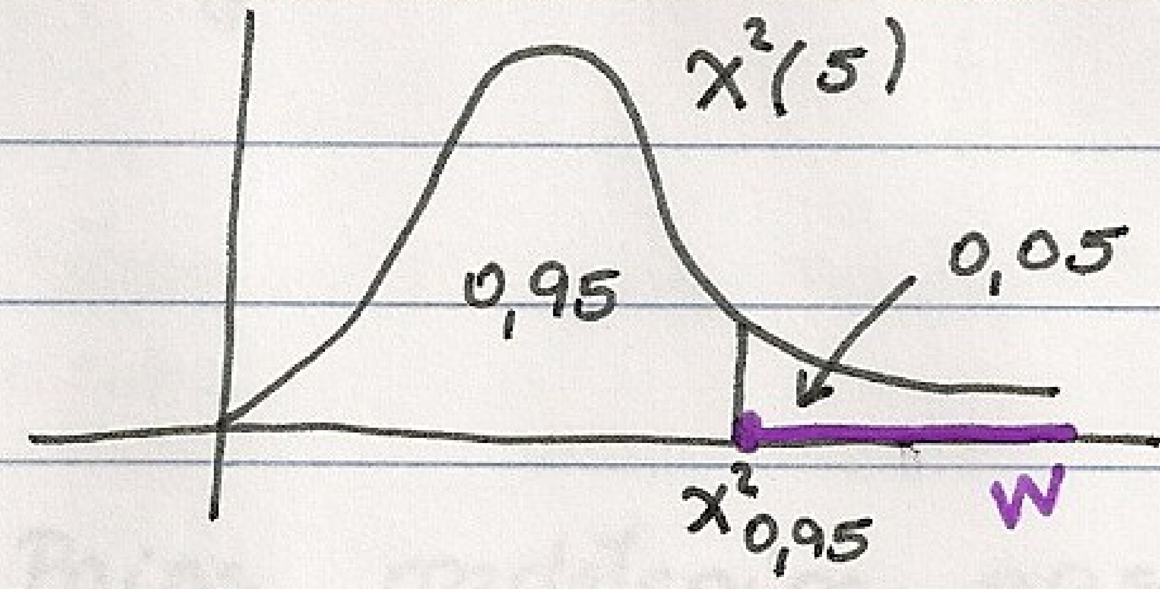
H_1 : „ –||– neshodují –||– "

$\alpha = 0{,}05$

očekávané četnosti: $n \cdot p = 60 \cdot \frac{1}{6} = 10 \quad \forall i = 1, \ldots 6$

$$\chi^2 = \frac{(14-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(5-10)^2}{10} =$$

$= \frac{1}{10}(16+9+9+1+4+25) = \frac{64}{10} = \underline{6,4}$



(platí-li H_0): $\quad \chi^2 \sim \chi^2(5)$

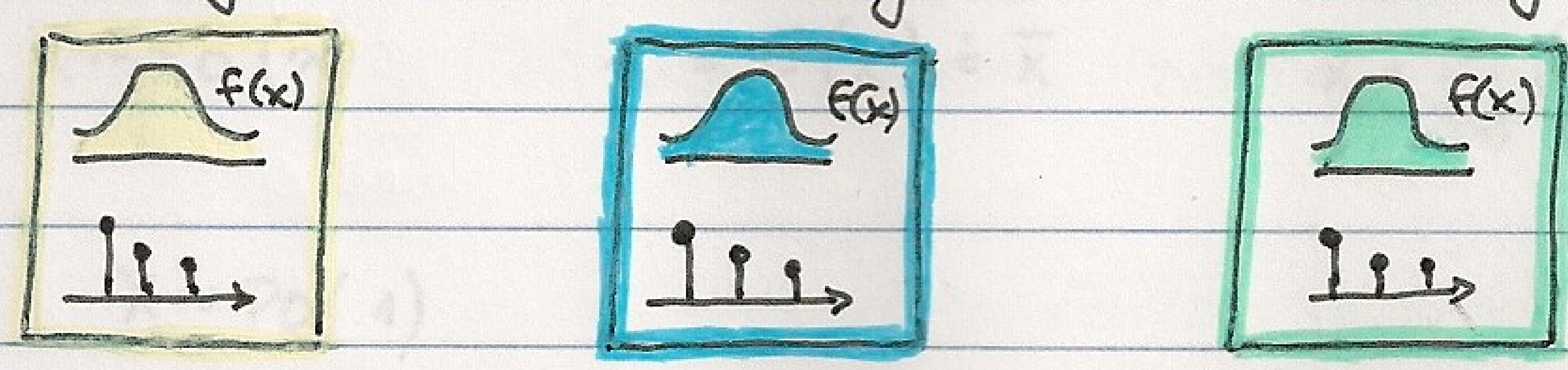
$\nu = 6-1 = 5$

$W = (11,1\ ;\ +\infty)$

př: Bestseller "Úžasný svět pravděpodobnosti"

se prodává ve třech barevných ~~přikladec~~ přebalech:

žlutý $\qquad$ modrý $\qquad$ zelený



Zatím bylo prodáno: $\quad$ 98 ks žlutých

137 ks modrých

125 ks zelených

Na hladině významnosti $\alpha = 0,05$ testujte:

H_0: "Počet prodaných ks nezávisí na barvě"

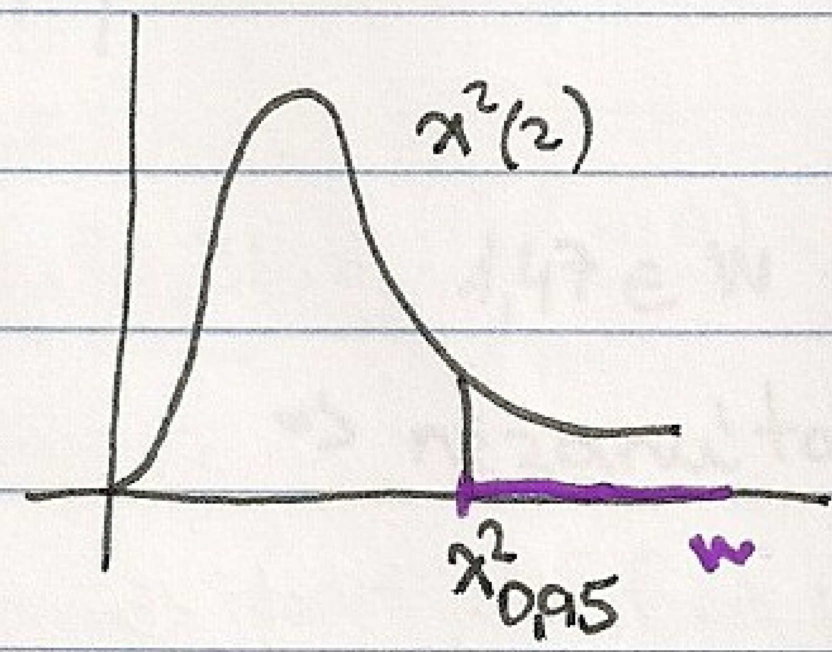
<u>Řešení</u>: $\quad H_0$: $\quad$ –||–

očekávané četnosti: $\quad \frac{98+137+125}{3} = \frac{360}{3} = 120$ — ž, m, z

$$\chi^2 = \frac{(98-120)^2}{120} + \frac{(137-120)^2}{120} + \frac{(125-120)^2}{120} = 6,65 \in W \Rightarrow$$

$\chi^2 = 2)$

$\nu = 3-1 = 2$



$W(5,99\ ;\ +\infty)$

Př: Nechť 200 realizací náh. veličiny X poskytlo četnosti:

i	0	1	2	3	4
m_i	69	81	33	14	3

Na hladině významnosti $\alpha = 0{,}05$ s Poiss. rozdělením ppsti.

Řešení: H_0: „Data se shodují s $Po(\lambda)$"

X

H_1: „ -||- neshodují -||- "

$X \sim Po(\lambda)$ $\quad \lambda = E(X) \doteq \bar{x}$ $\quad \bar{x} = \frac{1}{200}(0\cdot 69 + 1\cdot 81 + 2\cdot 33 + 3\cdot 14 + 4\cdot 3) =$

$= 1{,}005 \doteq 1$

$X \sim Po(1)$

$P(0) = F(0) = 0{,}368 \Rightarrow m_0^0 = 200 \cdot 0{,}368 = 73{,}6$ ✓

$P(1) = F(1) - F(0) = 0{,}368 \quad m_1^0 = 200 \cdot 0{,}368 = 73{,}6$ ✓

$P(2) = F(2) - F(1) = 0{,}184 \quad m_2^0 = 200 \cdot 0{,}184 = 36{,}8$ ✓

$P(3) = F(3) - F(2) = 0{,}061 \quad m_3^0 = 200 \cdot 0{,}061 = 12{,}2$

$P(4) = F(4) - F(3) = 0{,}0\cdot \quad m_4^0 = 200 \cdot 0{,}015 = 3{,}0$

! $P(X > 4) = 1 - P(X \leq 4) = 1 - F(4) = 0{,}004 \Rightarrow m_{>4}^0 = 0{,}8 < 5$

sdružíme tak, aby oč. četnosti byly ≥ 5

$m_{\geq 3}^0 = 12{,}2 + 3{,}0 + 0{,}8 = 16$ ✓

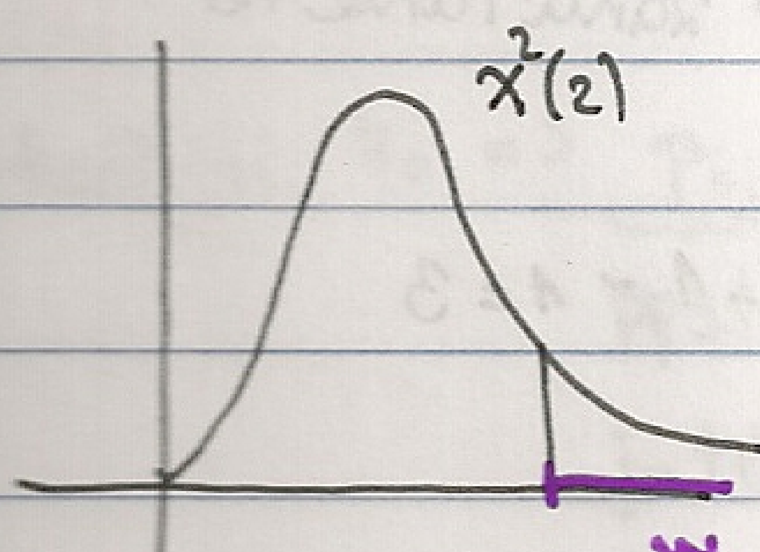
i	0	1	2	≥ 3
m_i	69	81	33	17
m_i^0	73,6	73,6	36,8	16

$\chi^2 = \frac{(69-73{,}6)^2}{73{,}6} + \dots + \frac{(17-16)^2}{16} \doteq 1{,}42$

$\nu = 4 - 1 - 1 = 2$

odhadovaný param.

$1{,}47 \in W$



$\Rightarrow$ nezamítáme

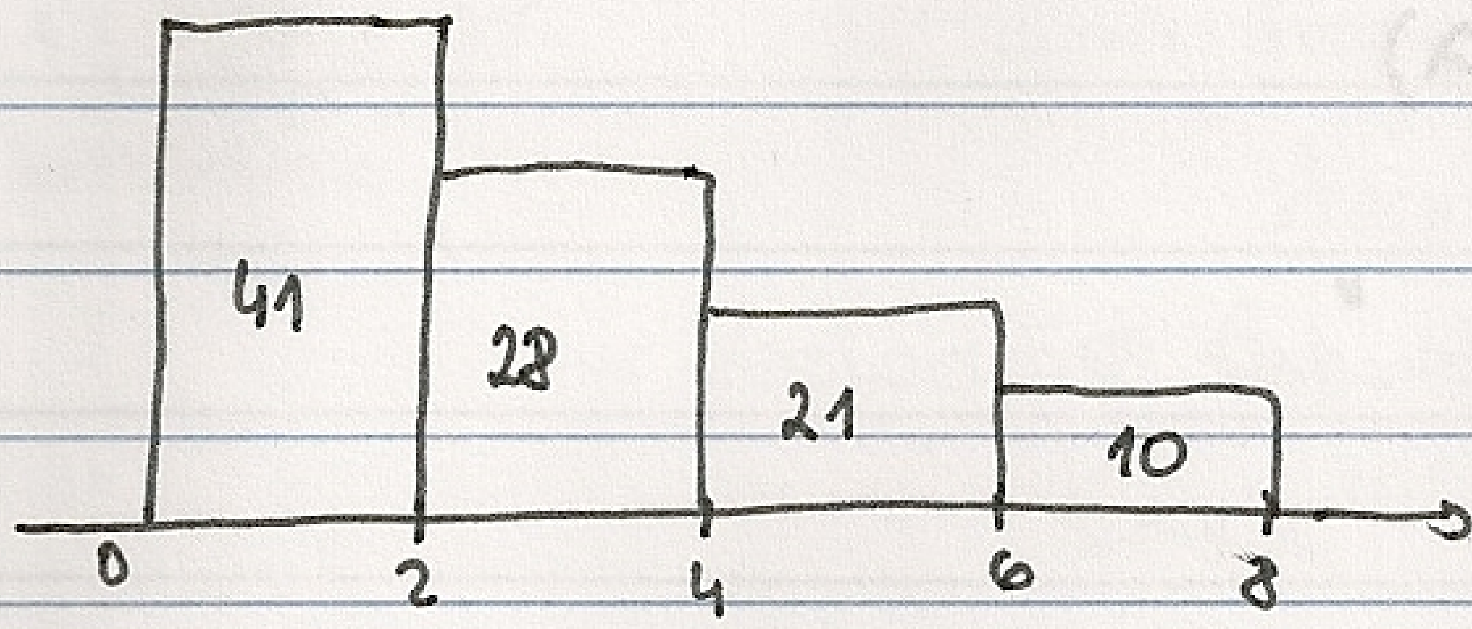
$W = (5{,}99; +\infty)$

↑ tab.

χ^2-test dobré shody lze použít i pro hodnoty získané ze *spojitých rozdělení,* jestliže jak pozorované tak očekávané četnosti náleží k nějakým disjunktním třídám (intervalům):

$\nu = k - 1 - m$, m ... počet odhadovaných parametrů

Př: Pro histogram četností dle obrázku:



nechť $\bar{x} \doteq 3{,}0$

testujte na hladině význ. $\alpha = 0{,}05$ shodu s $Exp(\delta)$

Řešení: H_0: "Data se shodují s $Exp(\delta)$"

X

H_1: " —"— neshodují —"— "

$E(X) = \delta \doteq \bar{x} = 3$

oček. ppst. : $P(0 < X < 2) = F(2) - F(0) = 1 - e^{-\frac{2}{3}} = 0{,}487$

$P(2 < X < 4) = F(4) - F(2) = 1 - e^{-\frac{4}{3}} - (1 - e^{-\frac{2}{3}}) \doteq 0{,}250$

$P(4 < X < 6) = \ldots = 0{,}128$

$P(6 < X < 8) = \ldots = 0{,}066$

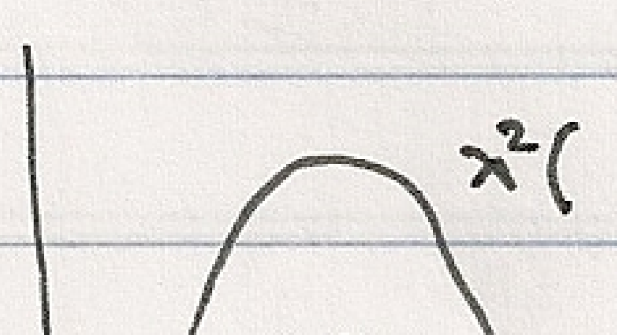
$P(X > 8) = 1 - P(X \leq 8) = 1 - F(8) \doteq 0{,}069$

$n = 41 + 28 + 21 + 10 = 100$

oček. četnosti jsou 100-krát větší než ppsti

$k = 5$ (všechny oček. četnosti jsou ≥ 5)

$$\chi^2 = \frac{(41 - 48{,}7)^2}{48{,}7} + \ldots + \frac{(10 - 6{,}6)^2}{6{,}6} + \frac{(0 - 6{,}9)^2}{6{,}9} \doteq \underline{15{,}48} \in W \Rightarrow \text{zamítáme } H_0$$



$W = (7{,}81; +\infty)$ $\nu = 5 - 1 - 1 = 3$

4. KONTINGENČNÍ TABULKY

Testujme např. zda technologie má vliv na podíl vadných výrobků. Pozorované četnosti n_{ij} $(i = 1, 2, 3;\ j = 1, 2)$ zapišme ve formě tzv. **kontingenční tabulky**

n_{ij}	Vyhovující výrobky	Vadné výrobky	
Technologie A	226	24	9,6%
Technologie B	136	14	9,3%
Technologie C	95	5	5%

kontingence ... "statistická"závislost jedné náhodné veličiny (např. technologie), ozn. ji X na jiné náhodné veličině (např. vyhovující/vadný výrobek), ozn. ji Y

Testujeme hypotézu H_0: "**X, Y jsou nezávislé.**

Testové kritérium:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - n_{ij}^O\right)^2}{n_{ij}^O},$$

kde r je počet možných hodnot veličiny X (**řádky**),

s je počet možných hodnot veličiny Y (**sloupce**),

n_{ij} ... **pozorované četnosti**

n_{ij}^O ... **očekávané četnosti** příslušné k n_{ij}

Za předpokladu platnosti H_0 je

$$n_{ij}^O = \frac{n_{i.}\ n_{.j}}{n},$$

kde $n_{i.}$ je součet četností v i-tém řádku tabulky,

$n_{.j}$ je součet četností v j-tém sloupci tabulky,

n je celkový počet pozorování v tabulce.

Jsou-li **všechny očekávané četnosti aspoň 5**, pak $\chi^2 \approx \chi^2(\nu)$, kde $\nu = (r-1)(s-1)$.

kontingenční tabulky - obdélníkové matice pozorovaných četností, kde řádky obsahují pozorované četnosti veličiny X
sloupce —〃— Y

platí-li $H_0 \Rightarrow \underbrace{p_{ij}}_{} = (p_i \cap p_j) \overset{\text{nezávislé}}{=} p_i \cdot p_j \doteq \frac{m_{i.}}{m} \cdot \frac{m_{.j}}{m}$, pak $m_{ij}^o = p_{ij} \cdot m = \frac{m_{i.} \cdot m_{.j}}{m}$

ppst, že naměřená hodnota bude v i-tém řádku a j-tém sloupci

$$m_{i.} = \sum_{j=1}^{i} m_{ij}$$

$$m_{.j} = \sum_{i=1}^{j} m_{ij}$$

př: pro kont. tabulku danou danou výše testujte na hl. významnosti 0,05 hypotézu, že podíl vadných výrobků nezávisí na použité technologii.

Řešení: H_0: podíl vadných výrobků nezávisí

X

H_1: —||— závisí

$\alpha = 0{,}05$

m_{ij}	vyhovující	vadné		$m_{i\cdot}$
A	226	24	(9,6 %)	250
B	136	14	(9,3 %)	150
C	95	5	(5 %)	100
$m_{\cdot j}$	457	43		500

Celkem testovaných výrobků: 500

$$m^0_{11} = \frac{250 \cdot 457}{500} = 228{,}5$$

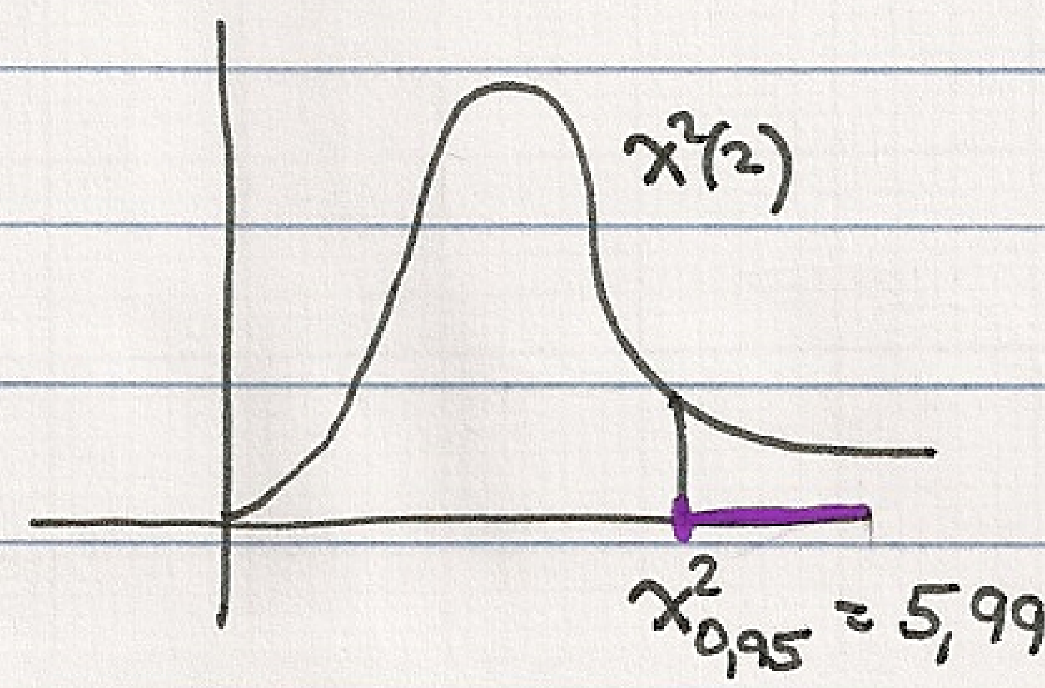
$\vdots$

$$m^0_{32} = \frac{100 \cdot 43}{500} = 8{,}6$$

m^0_{ij}	vyh.	vadné
A	228,5	21,5
B	137,1	12,9
C	91,4	8,6

$$\chi^2 = \frac{(226-228{,}5)^2}{228{,}5} + \dots + \frac{(5-8{,}6)^2}{8{,}6} \doteq 2{,}07$$

$$\nu = (3-1)(2-1) = 2$$



$2{,}07 \notin W \Rightarrow$ nezamítáme H_0 $\qquad W = (5{,}99; +\infty)$

př: Společnost "Drink Comp." hodlá nainstalovat nápojové automaty do škol. Nápojový manager společnosti udělal průzkum trhu. 200 náh. vybraných studentů dalo tyto výsledky:

x\y	SŠ	VŠ	$m_{i\cdot}$
A	33	57	90
B	30	20	50
C	5	35	40
D	12	8	20
$m_{\cdot j}$	80	120	200

Na hl. významnosti $\alpha = 0{,}01$ testujte hypotézu, že náp. preference nezávisí na typu školy.

Řešení: H_0 : „Pref nezávisí na typu školy"

x

H_1 : „ —||— závisí —||— "

m^o_{ij}

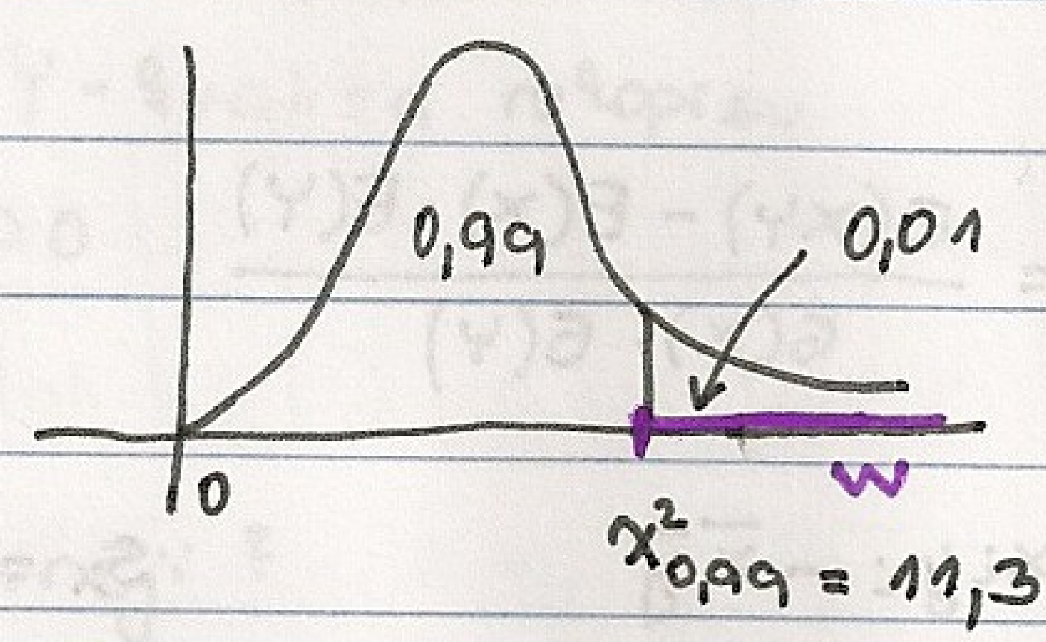
	SŠ	VŠ
A	36	54
B	20	30
C	16	24
D	8	12

$$m_{21} = \frac{50 \cdot 80}{200} = 20$$

$$m_{32} = \frac{40 \cdot 120}{200} = 24$$

$$\chi^2 = \frac{(33-36)^2}{36} + \ldots + \frac{(\overset{8}{\cancel{12}} - 12)^2}{12} = 24{,}68$$

$\nu = (4-1)(2-1) = 3$



$W = (11{,}3; +\infty)$

$24{,}68 \in W \Rightarrow$ zamítáme H_0

Nápojové automaty závisí na typu školy.

20.12. - předtermín zkoušky od 13:00

Další:

Pá 13.1. „Last minute" zápočty !

5. TESTY VÝBĚRU Z ROZDĚLENÍ

TEST NEZÁVISLOSTI

Víme (viz Náh. vektor): Pro $(X, Y) \sim N_2(\vec{\mu}, \Sigma)$ platí:
X, Y jsou nezávislé $\Leftrightarrow$ korelační koeficient $\varrho = 0$.

Jsou-li (x_i, y_i), $i = 1, 2, \dots, n$ realizace náh. výběru, odhadujeme $\varrho \approx r$, kde

$$\boxed{r = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}}{s_x \cdot s_y}}$$

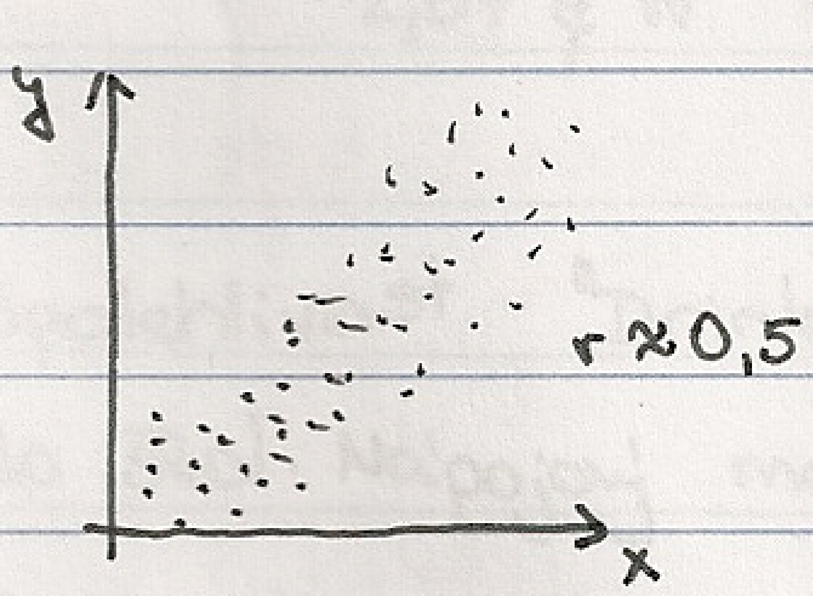
$\bar{x}$ je aritm. průměr hodnot $x_1, \dots, x_n$,
$\bar{y}$ - " - $y_1, \dots, y_n$,
s_x je výběr. směrodatná odchylka hodnot $x_1, \dots, x_n$,
s_y - " - $y_1, \dots, y_n$.

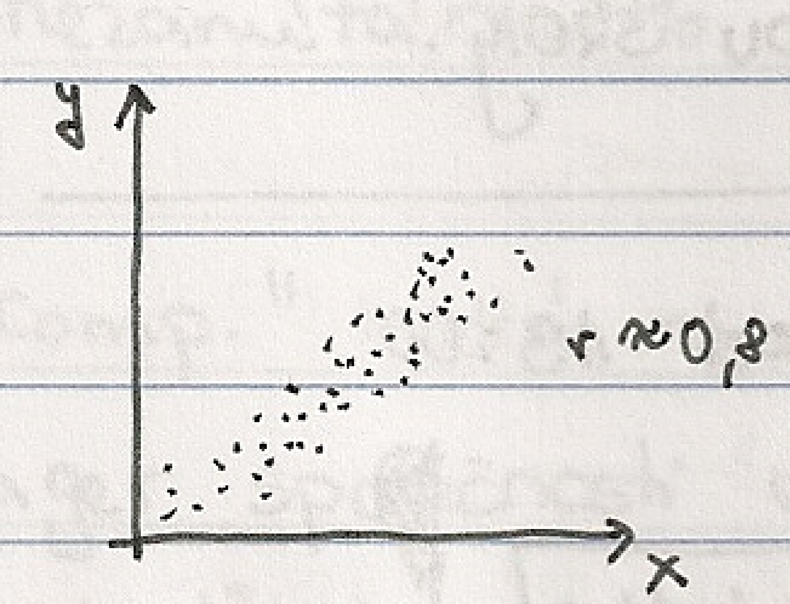
r ... **výběrový korelační koeficient**

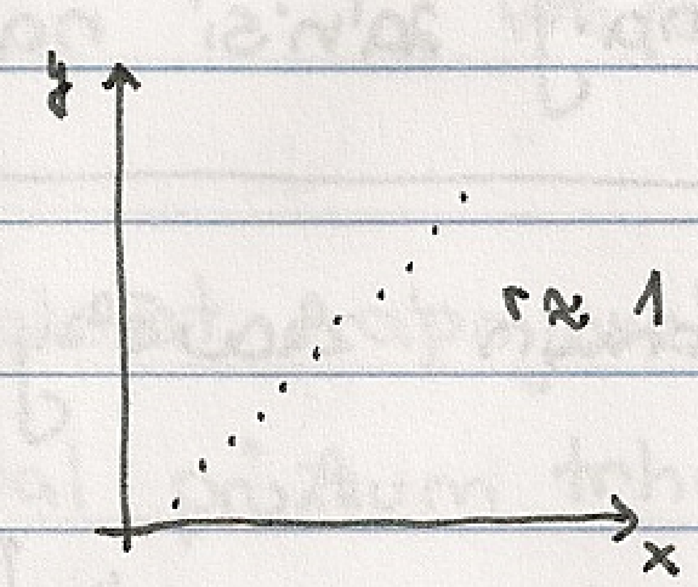
Platí: $-1 \le r \le +1$,
$r \approx 0$ naznačuje, že x_i a y_i jsou lineárně nezávislé
(resp. nezávislé - pro normální rozdělení N_2),
$r \to \pm 1$ naznačuje silnou lineární závislost
(resp. závislost pro N_2)

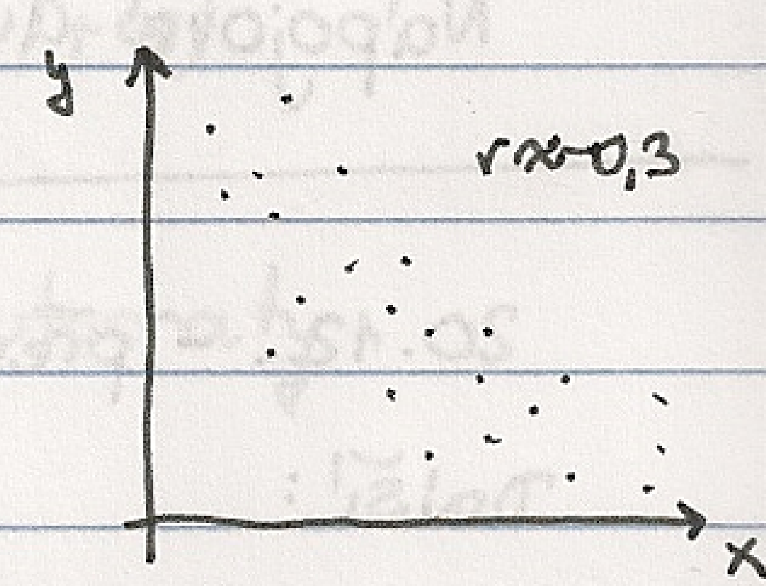
$$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)} = \frac{E(XY) - E(X)\cdot E(Y)}{\sigma(X)\cdot\sigma(Y)}$$

$$\approx r = \frac{\frac{1}{m}\sum_{i=1}^{m} x_i y_i - \bar{x}\bar{y}}{s_x \cdot s_y} \qquad s_x = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

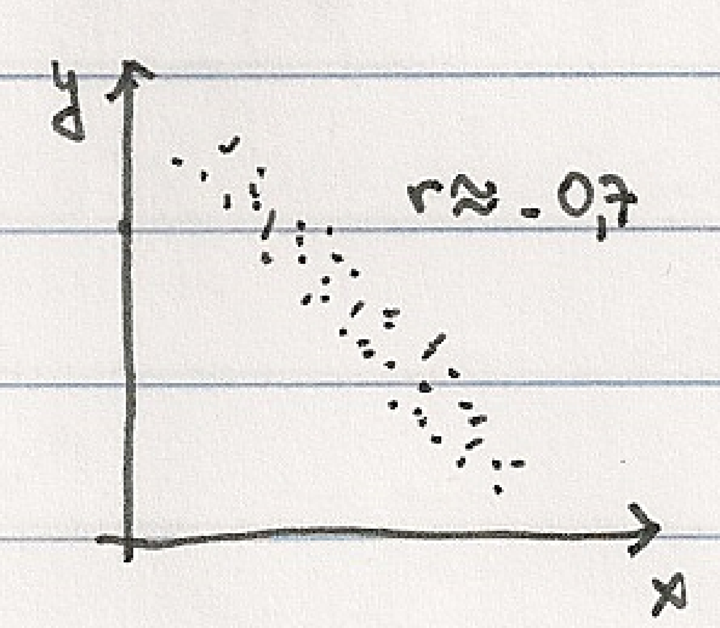


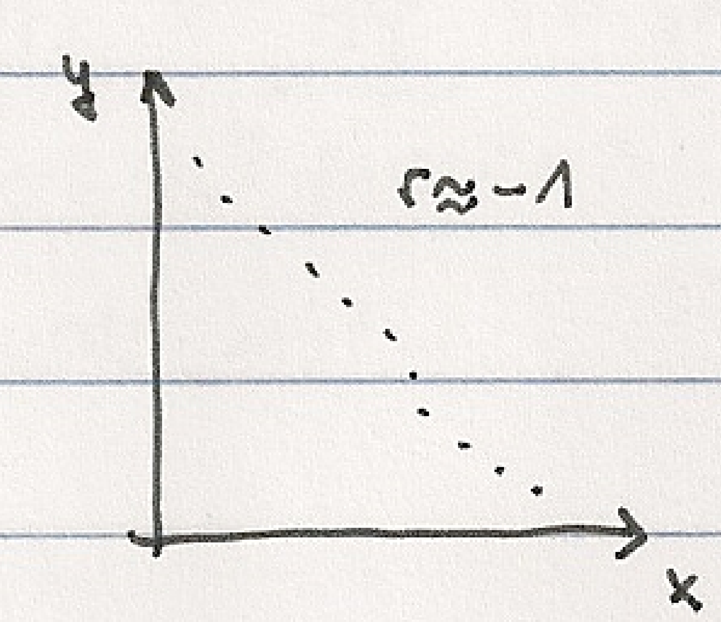


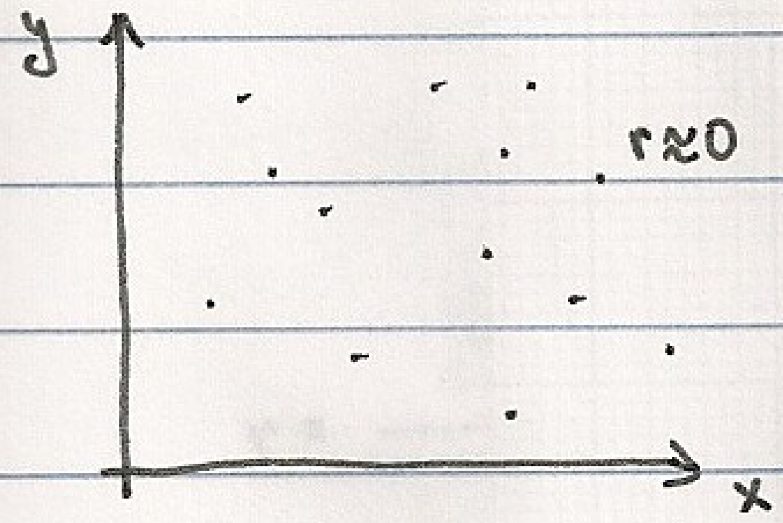


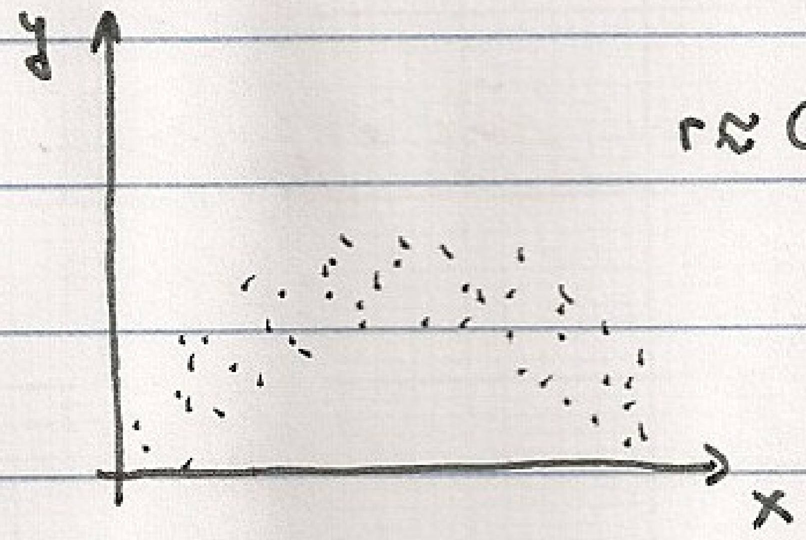


$\underline{r}$ udává míru lin. závislosti









r≈0 (není lin. závislost!)

	X	Y	r	
1.	stáří auta	cena auta	-	
2.	délka vzdělání	roční příjem	+	
3.	ujeté km	spotřeba paliva	+	
4.	číslo na 1. kostce	číslo na 2. kostce	0	nezávislé jevy! LNZ
5.	schopnost vidět ve tmě	množství snězené mrkve	?	

KLAMNÁ KORELACE

- hodnota korel. koef. mezi náh. veličinami X, Y je vysoká nikoli vlivem přímé souvislosti mezi X, Y, ale vlivem jiné náh. veličiny Z.

př: X - velikost dětské ruky Y - kvalita rukopisu

r≫0

Jak vysvětlit:

a, větší ruka drží tužku pevněji?

b, Praxe v psaní způsobuje větší ruku?

c, Je zde jiná náh. veličina? (Jaká?)

Odpověď: ANO - věk ovlivňuje jak velikost ruky, tak kvalitu psaní

K testu nezávislosti (při výběrech z N_2) používáme testové kritérium

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

Platí-li $H_0 : \varrho = 0$, pak $t \sim t(\nu)$, kde $\nu = n - 2$.

(Alternativa $H_1 : \varrho \neq 0 \Rightarrow$ test je oboustranný.)

Zamítneme-li H_0 na hladině $\alpha = 0.05$, resp. $\alpha = 0.01$, říkáme, že r je **významný**, resp. **vysoce významný**, a hodnoty x_i a y_i považujeme za závislé.

t-TESTY SHODY DVOU STŘEDNÍCH HODNOT

(x_i, y_i) náh. výběr rozsahu n z rozd. $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$, kde $\vec{\mu} = (\mu_1, \mu_2)$, $\mu_1 = E(X)$, $\mu_2 = E(Y)$.

Testujeme

$$\boxed{H_0 : \mu_1 = \mu_2}$$

• ***t*-test pro závislé výběry** - tzv. **párový *t*-test**

$z_i = y_i - x_i$ $(i = 1, \ldots, n)$ jsou realizace náh. veličiny $Z = Y - X$ se střední hodnotou $\mu = \mu_2 - \mu_1$
Pak $H_0 : \mu = 0$ a testové kritérium je

$$t = \frac{\bar{z}}{s}\sqrt{n}$$

kde $\bar{z}$ je aritm. průměr hodnot $z_1, \ldots, z_n$,
s je jejich výběr. směrodatná odchylka.

Při platnosti H_0 je $t \sim t(\nu)$, kde $\nu = n - 1$.

Musíme si rozmyslet, jestli testujeme závislé nebo nezávislé výběry.

při závislé výběry
Výsledky 1. a 2. testu u 10 náhodně vybraných studentů (max 12 bodů).

Test č. 1 / x_i	10	9	6	7	6	9	10	7	12	10
Test č. 2 / y_i	7	6	7	5	6	8	10	9	10	8

Předpokládáme, že výběr je realizací $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$
Na hladině významnosti $\alpha = 0{,}05$ testujte:
a, hypotézu, že výsledky obou testů jsou vzájemně závislé
b, –||– , že výsledky obou testů jsou vzájemně závislé

Řešení: a, $H_0: \rho = 0$ x $H_1: \rho \neq 0$

$$t = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{m-2} \qquad m = 10$$

spočteme: $\bar{x} = 8{,}6$, $s_x = 2{,}01$

$\bar{y} = 7,6$, $s_y = 1,71$

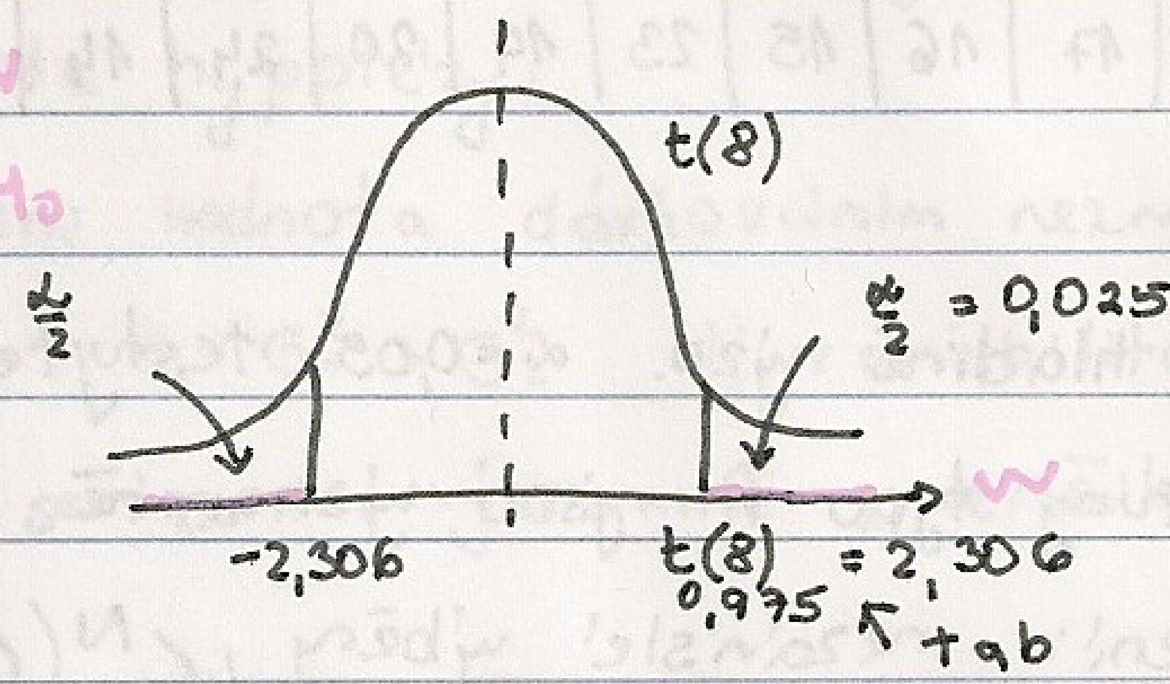
$$r = \frac{\frac{1}{10}\, 681 - 8,6 \cdot 7,6}{2,01 \cdot 1,71} \doteq 0,80 \qquad \sum_{i=1}^{10} x_i y_i = 681$$

$$t = \frac{0,8}{\sqrt{1-0,8^2}} \sqrt{10-2} = \frac{4}{3} \cdot \sqrt{8} = 3,771 \in W$$

zamítáme H_0

$H_0 \to$ $t \sim t(8)$

$\nu = 10-2 = 8$



$$W = \{t \,;\, |t| \geq 2,306\}$$

b) $H_0: \mu_1 \leq \mu_2 \quad \times \quad H_1: \mu_1 > \mu_2$

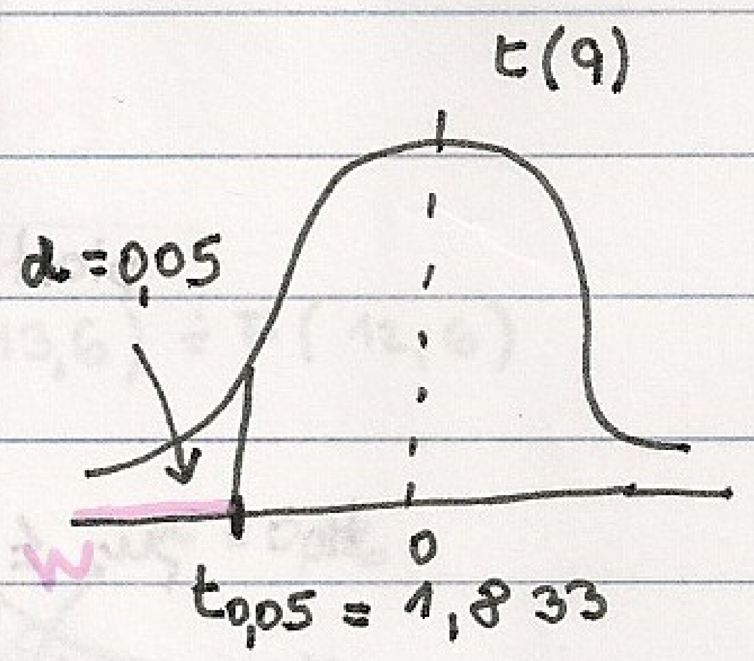
použijeme párový t-test , $Z = Y - X$

stejné jako $H_0: \mu \leq 0$ $\qquad z_i = y_i - x_i$

$\times$ $\qquad \mu = \mu_2 - \mu_1$

$H_1: \mu < 0$

$\nu = 10-1 = 9$



$z_i = y_i - x_i$										
	−3	−3	1	−2	0	−1	0	+2	−2	−2

$\bar{z} = -1,0$ $\qquad s_z = 1,35$ $\qquad t = \frac{-1,0}{1,35}\sqrt{10} = -3,279$

$\in W \Rightarrow$ zamítáme H_0

- **t-test pro nezávislé výběry**

Výběr rozsahu n_1 z rozdělení $N(\mu_1, \sigma_1^2)$:

$x_1, x_2, \ldots, x_{n_1}$, arim. průměr $\bar{x}$, výběr.rozptyl s_1^2

Výběr rozsahu n_2 z rozdělení $N(\mu_2, \sigma_2^2)$:

$y_1, y_2, \ldots, y_{n_2}$, aritm. průměr $\bar{y}$, výběr.rozptyl s_2^2

Testové kritérium :

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n_1}s_1^2 + \frac{1}{n_2}s_2^2}}$$

Platí-li $H_0: \mu_1 = \mu_2$, je $t \sim t(\nu)$, kde $\nu = min(n_1, n_2) - 1$.

struny typu A jsou významně pevnější.

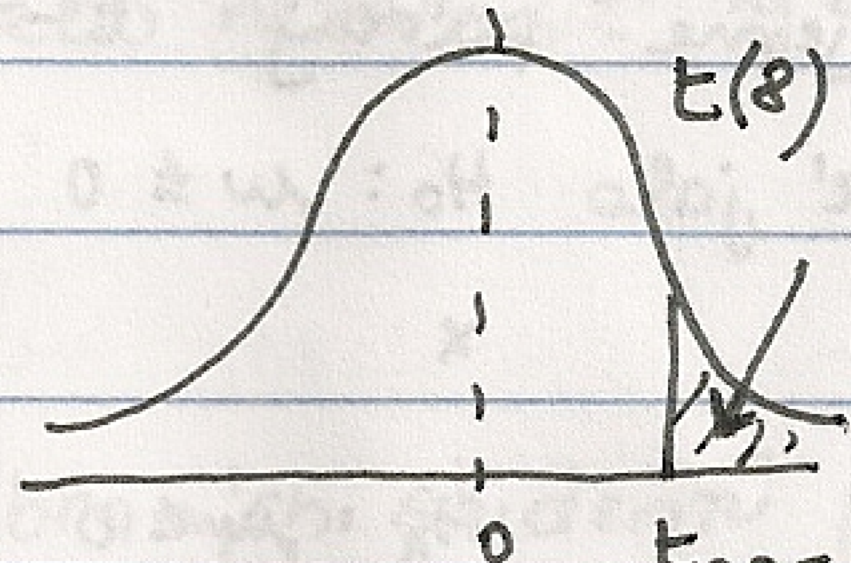
<u>Řešení</u>: nezávislé výběry, $\begin{cases} N(\mu_1, \sigma_1^2) \\ N(\mu_2, \sigma_2^2) \end{cases}$ nezávisle ⇒ dva výběry s normálním rozdělením

spočteme: $\bar{x} \doteq 21{,}4$ $\quad s_x \doteq 2{,}55 \Rightarrow s_x^2 = 6{,}5$ $\quad n_1 = 9$

$\bar{y} \doteq 17{,}8$ $\quad s_y \doteq 3{,}78 \Rightarrow s_y^2 \doteq 14{,}29$ $\quad n_2 = 12$

$$t = \frac{21{,}4 - 17{,}8}{\sqrt{\frac{6{,}5}{9} + \frac{14{,}29}{12}}} \doteq 2{,}603 \in W \Rightarrow$$

zamítáme H_0



platí-li H_0, pak $t \sim t(\vartheta)$ $\qquad \vartheta = \min(9, 12) - 1 = 8$

$W = (1{,}86\,;\, +\infty)$

H_0: $\mu_1 \leq \mu_2$ ("struny A se jeví jako pevnější")

H_1: $\mu_1 > \mu_2$

F-TEST SHODY DVOU ROZPTYLŮ

Označme dva nezávislé náhodné výběry z normálního rozdělení tak, aby $s_1^2 \geq s_2^2$.

Testové kritérium:

$$f = \frac{s_1^2}{s_2^2}$$

Platí-li $H_0 : \sigma_1^2 = \sigma_2^2$, pak $f \sim F(\nu_1, \nu_2)$, kde $\nu_1 = n_1 - 1,\ \nu_2 = n_2 - 1$.

$F(\nu_1, \nu_2)$ je tzv. **F-rozdělení** s ν_1 a ν_2 stupni volnosti

Test je pravostranný a kvantily (=kritické hodnoty testu) F-rozdělení jsou v tabulkách.

Př: Výrobní linka na 500 ml nápoje nepřesně dávkovala, proto byla seřízena. Před seřízením bylo kontrolně změřeno 14 vzorků:

$\bar{x} = 500{,}34$ [ml] $s = 3{,}24$ [ml]

Po seřízení bylo kontrolně změřeno 7 vzorků $\bar{y} = 500{,}31$ [ml] $s_y = 0{,}59$ [ml]

Na hladině významnosti $\alpha = 0{,}05$ testujte hypotézy:

a) $H_0: \mu_1 = \mu_2$ (seřízením se střední hodnota dávkováním nezměnila)

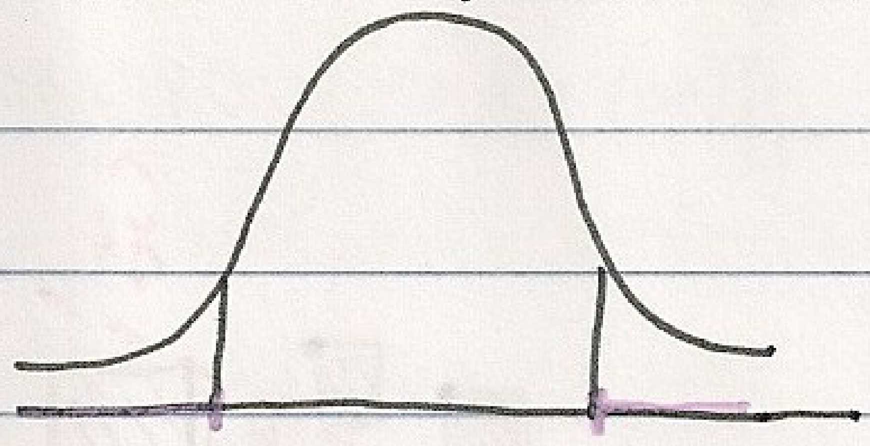
b) $H_0: \sigma_1^2 \le \sigma_2^2$ (proti alternativě $H_1: \sigma_1^2 > \sigma_2^2$ tj. variabilita dávkování se se seřízením významně zmenšila)

Řešení: nezávislé výběry

a)
$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} = \frac{500{,}34 - 500{,}31}{\sqrt{\frac{3{,}24^2}{14} + \frac{0{,}59^2}{7}}} \doteq 0{,}034 \notin W$$

=> nezamítáme H_0

$\nu = \min(7, 14) - 1 = 6$



$W = \{t;\ |t| > 2{,}2447\}$

b)
$$f = \frac{s_1^2}{s_2^2} = \frac{3{,}24^2}{0{,}59^2} = 30{,}17 \in W \Rightarrow$$

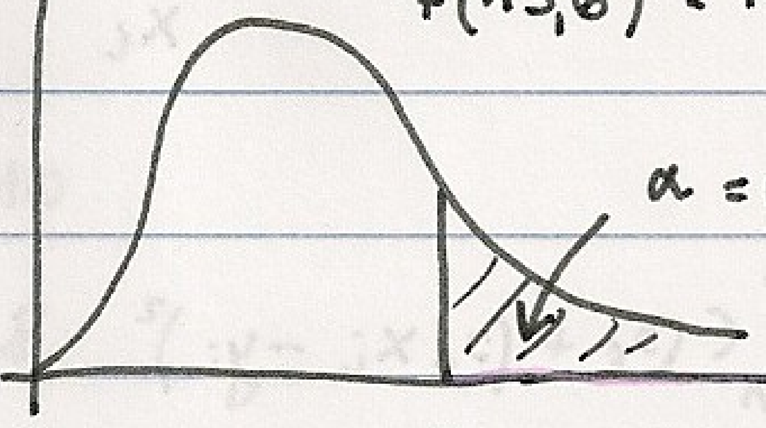
zamítáme H_0

pravostranný test:

kvantil F-rozdělení, $\nu_1 = n_1 - 1 = 13$

95%-ní $\nu_2 = n_2 - 1 = 6$

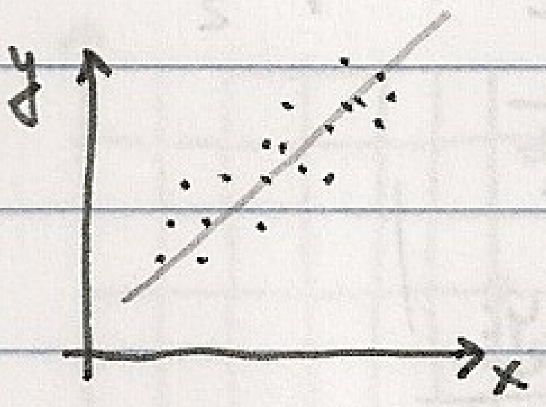
$F(13,6) \doteq F(12;6)$



$F_{0,95} \doteq 4.0$

tab

Variabilita se dávkováním zmenšila. $W = (4; +\infty)$



Nejjednodušší případ regresní křivky.

6. REGRESE, METODA NEJMENŠÍCH ČTVERCŮ

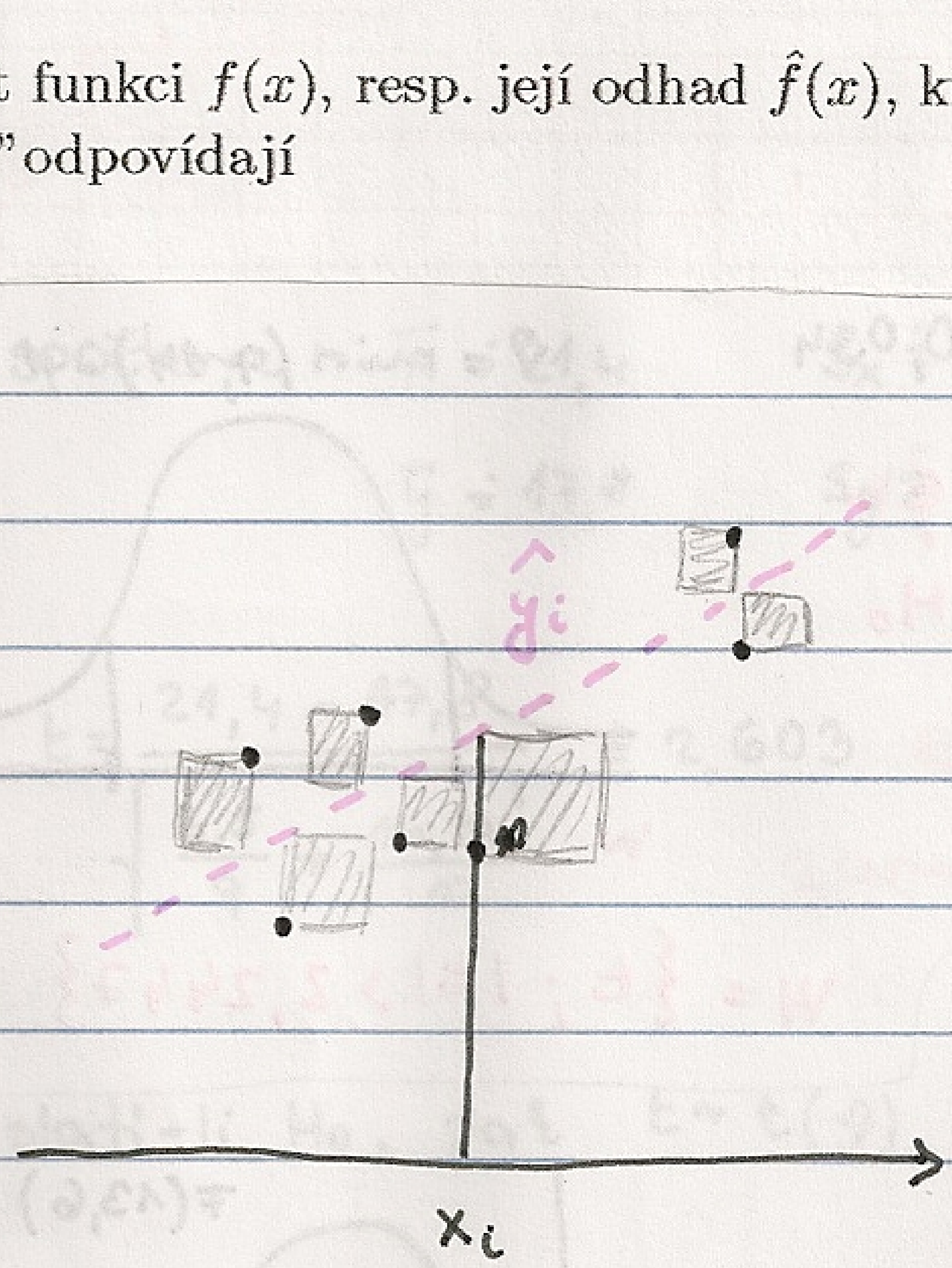
dáno: $[x_i, y_i], i = 1, 2 \dots, n$

$x_i \dots$ dány "přesně" či se zanedbatelnou chybou

$y_i \dots$ určeny "nepřesně" ($\approx$ realizace náh. veličiny,

$y_i \dots$ naměřené hodnoty **neznámé** funkce $y = f(x)$ v bodech x_i, tj. $y_i \approx f(x_i)$

cíl: určit funkci $f(x)$, resp. její odhad $\hat{f}(x)$, které tyto (s "chybou") naměřené hodnoty "nejlépe" odpovídají



$$S = \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \longrightarrow \min$$

hledáme nejmenší součet obsahů čtverců

$$S = \sum_{i=1}^{m} (\beta_0 + \beta_1 x_i - y_i)^2$$

$$\left.\begin{array}{l} \dfrac{\partial S}{\partial \beta_0} = 0 \\[2em] \dfrac{\partial S}{\partial \beta_1} = 0 \end{array}\right\} \Leftrightarrow \begin{array}{l} \sum_{i=1}^{m} 2(\beta_0 + \beta_1 x_i - y_i)^1 \cdot 1 = 0 \quad / \cdot \frac{1}{2} \\[1em] \sum_{i=1}^{m} 2(\beta_0 + \beta_1 x_i - y_i)^1 x_i = 0 \quad / \cdot \frac{1}{2} \end{array}$$

$$\boxed{\begin{array}{l} \beta_0 m + \beta_1 \sum x_i = \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 = \sum x_i y_i \end{array}}$$

soustava normálních rovnic

do matice:

$$\begin{array}{cc} \beta_0 & \beta_1 & \\ \end{array}$$
$$\left[\begin{array}{cc|c} m & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \end{array}\right]$$

9.12.

Regresní přímka

Předpokládáme, že f je má tvar

$$f(x) = b_0 + b_1 x$$

a odhady parametrů $\beta_0 \approx b_0$, $\beta_1 \approx b_1$ najdeme tzv. **metodou nejmenších čtverců** - hledáme, pro která β_0, β_1 je minimální součet čtverců

$$S = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i - y_i)^2.$$

Metodami mat. analýzy lze ověřit, že hledaná β_0, β_1 jsou řešením tzv. **soustavy normálních rovnic**

$$\beta_0 \cdot n + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

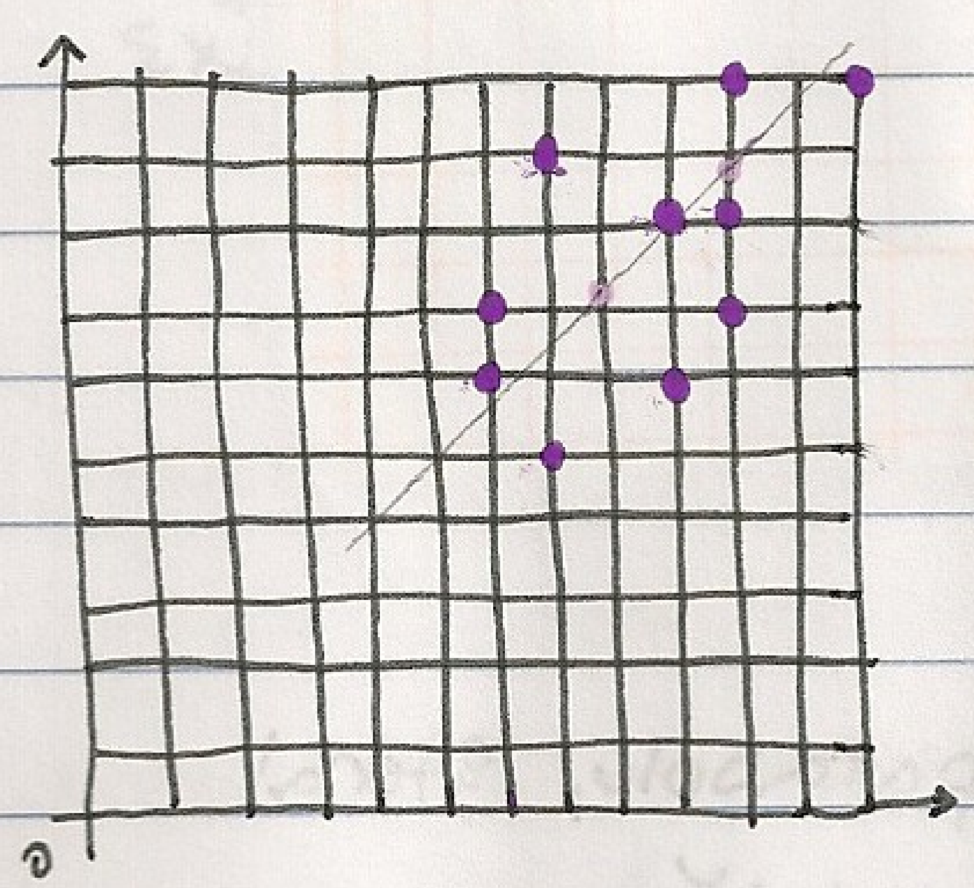
$$\begin{array}{ccc} \beta_0 & \beta_1 & \\ \end{array}$$
$$\left[\begin{array}{cc|c} m & \Sigma x_i & \Sigma y_i \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i y_i \end{array}\right]$$

Př: Pro data z příkladu z 2.12.

Test č.1	10	9	6	7	6	9	10	7	12	10
Test č.2	7	6	7	5	6	8	10	9	10	8

Bylo vypočteno $r \doteq 0,8$

(otestovali jsme, že je významný)



- Pro data z příkladu najděte odhady β_0, β_1 regresní přímky $y = b_0 + b_1 x$ metodou nejmenších čtverců.
- Odhadněte pak hodnoty $\hat{y}(8)$, $\hat{y}(10)$

$m = 10$

$\Sigma x_i = 86$ $\qquad$ $\Sigma x_i y_i = 681$

$\Sigma y_i = 76$

$\Sigma x_i^2 = 776$

soustava: $\left[\begin{array}{cc|c} 10 & 86 & 76 \\ 86 & 776 & 681 \end{array}\right]$

Cramer: $D = 364$

$D_1 = \det\begin{bmatrix} 76 & 86 \\ 681 & 776 \end{bmatrix} = 40$ $\qquad D_2 = \det\begin{bmatrix} 10 & 76 \\ 86 & 681 \end{bmatrix} = 274$

$\beta_0 = \frac{410}{364} = 1{,}13$ (kde přímka protne osu y)

$\beta_1 = \frac{274}{364} = 0{,}75$ (směrnice)

$\hat{f}(x)$: $\hat{y} = 1{,}13 + 0{,}75 \cdot x$

$\left.\begin{array}{l} \hat{y}(8) = 1{,}13 + 0{,}75 \cdot 8 = 7{,}13 \\ \hat{y}(10) = 1{,}13 + 0{,}75 \cdot 10 = 8{,}63 \end{array}\right\}$ „vyrovnané hodnoty"

Regresní parabola

Pro funkci f tvaru

$$f(x) = b_0 + b_1 x + b_2 x^2$$

odhady $\beta_0 \approx b_0$, $\beta_1 \approx b_1, \beta_2 \approx b_2$ metodou nejm. čtverců jsou ty, které minimalizují výraz

$$S = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 - y_i)^2$$

a jsou řešením soustavy

$$\beta_0 \cdot n + \beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \beta_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$\beta_0 \sum_{i=1}^{n} x_i^2 + \beta_1 \sum_{i=1}^{n} x_i^3 + \beta_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

$\left[\begin{array}{ccc|c} m & \Sigma x_i & \Sigma x_i^2 & \Sigma y_i \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i y_i \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 & \Sigma x_i^2 y_i \end{array}\right]$

- hledáme parabolu, která je bodům nejblíže př.: V bodech x_i $(i = 1, 2, \dots 8)$ byly naměřeny hodnoty y_i:

x_i	-2	-1	-1	0	0	1	1	2
y_i	11,02	6,07	5,92	2,88	2,93	2,05	1,98	3,15

Metodou nejmenších čtverců najděte odhady koeficientů

a, regresní přímky

b, regresní paraboly

a, $m = 8$

$\sum x_i = 0$

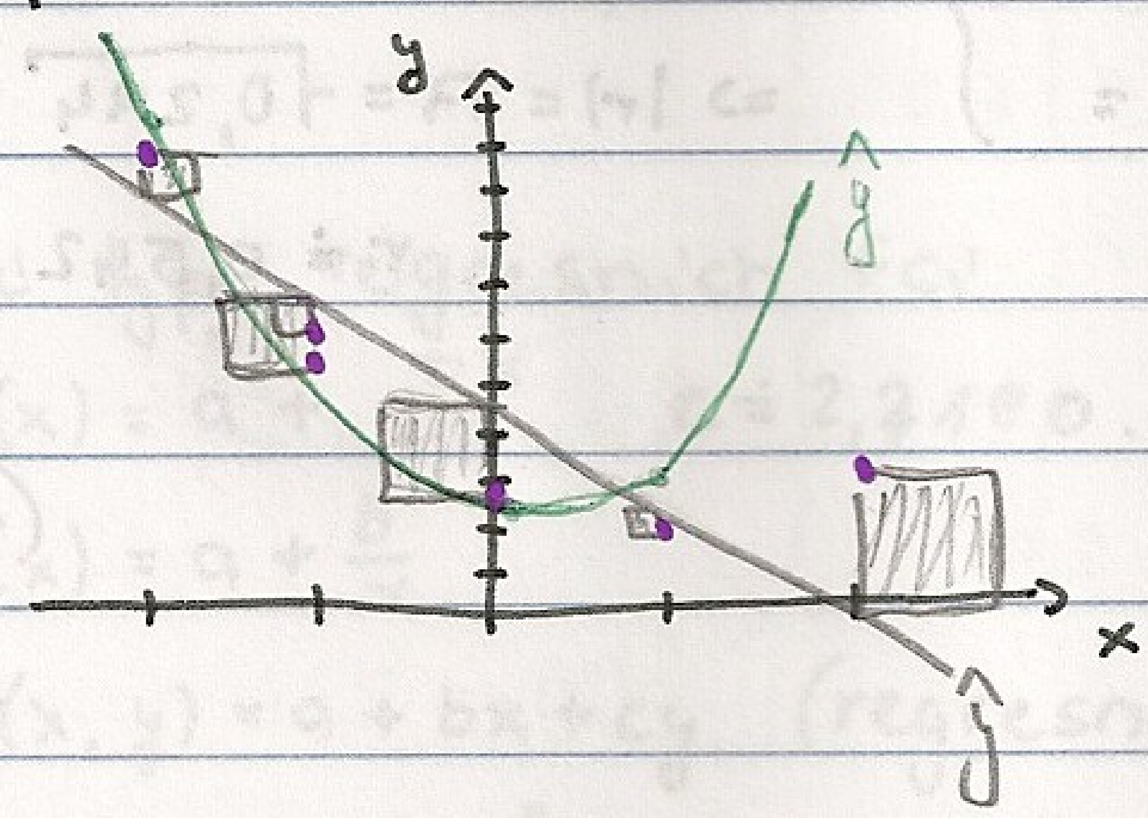
$\sum x_i^2 = 12$

$\sum y_i = 36$

$\sum x_i y_i = -23,7$

$$\left[\begin{array}{ccc|c} 8 & 0 & 12 & 36 \\ 0 & 12 & & -23,7 \\ 12 & & & \end{array}\right] \qquad \left[\begin{array}{cc|c} 8 & 0 & 36 \\ 0 & 12 & -23,7 \end{array}\right]$$

$$\hat{y} = 4,5 - 1,975\,x$$



- k určení přímky stačí dva body (průsečík s osou y a nějaký další)

b,

$$\left[\begin{array}{ccc|c} 8 & 0 & 12 & 36 \\ 0 & 12 & 0 & -23,7 \\ 12 & 0 & 36 & 72,7 \end{array}\right]$$

$\beta_0 \doteq 2,942$

$\beta_1 \doteq -1,975$

$\beta_2 \doteq 1,039$

$\sum x_i^3 = 36$

$\sum x_i^4 = 72,7$

$$\hat{y} = 2,942 - 1,975x + 1,039\,x^2$$

R.

$$R^2 = 1 - \frac{RSS}{S_T}$$

$RSS = \sum (y_i - \hat{y}_i)^2$ tzv. reziduální součet čtverců

$S_T = \sum (y_i - \bar{y})^2$ tzv. totální —"—

Pro regresní přímku: $|r| = R$

- V předchozím příkladu pro $\hat{y} = 4{,}5 - 1{,}975\,x$

$\hat{f}(-2) = 8{,}45$

$\hat{f}(-1) = 6{,}475$

$\hat{f}(0) = 4{,}5$

$\hat{f}(1) = 2{,}625$

$\hat{f}(2) = 0{,}55$

$\bar{y} = 4{,}5$

$\left.\begin{array}{l} RSS = \\ S_T = \end{array}\right\}$ $R^2 = 1 - 0{,}706 = 0{,}294 \Rightarrow$ cca 29 %

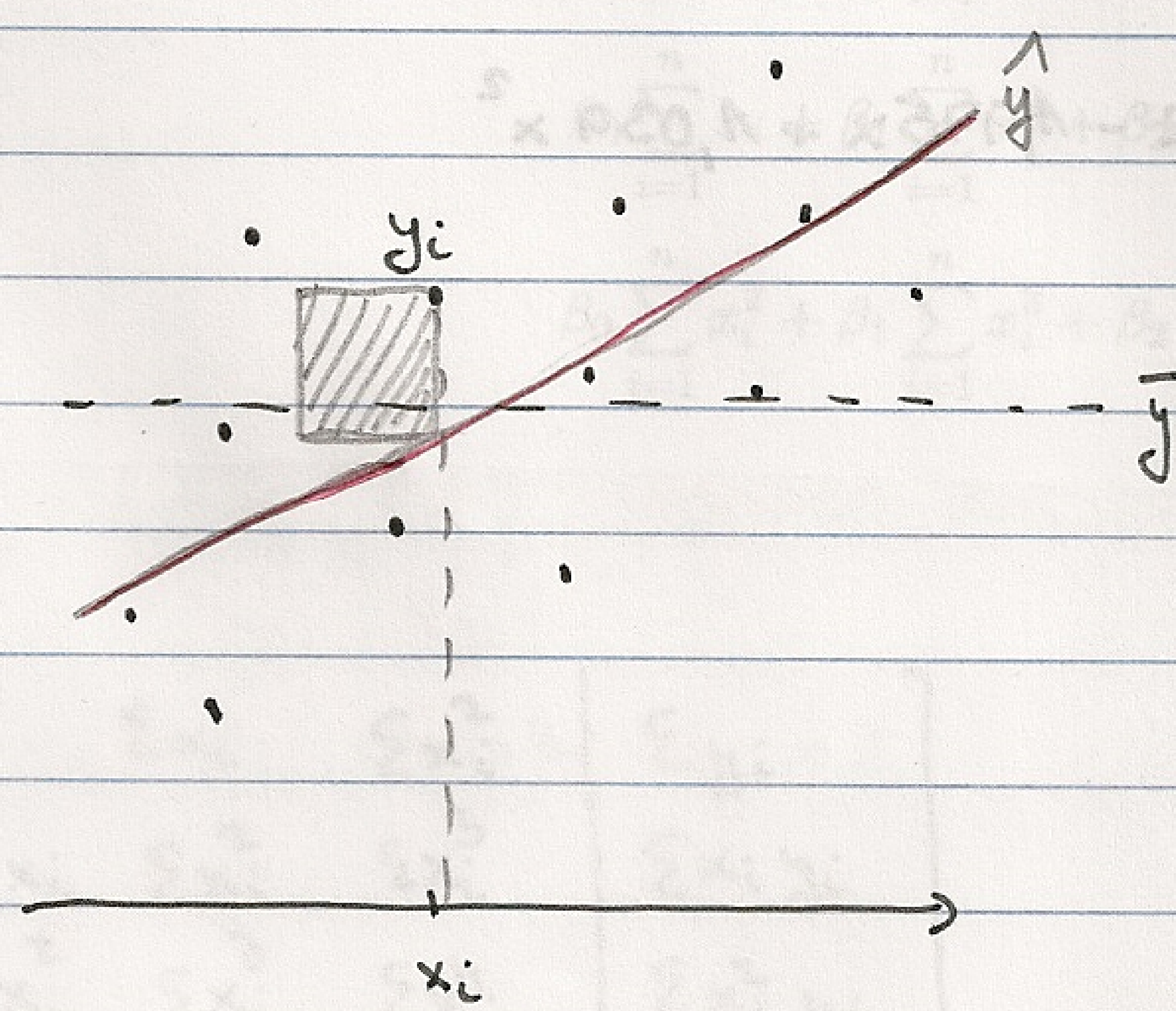
$\Rightarrow |r| = R = \sqrt{0{,}294} \doteq 0{,}542$

$r \doteq 0{,}542$

vyhovuje model reg. přímky

př: Pro příklad s výsledky testů studentů:

víme, že $r = 0{,}80$, pak $R^2 = 0{,}8^2 = 0{,}64 \Rightarrow$ cca 64 % vyhovuje model reg. přímky



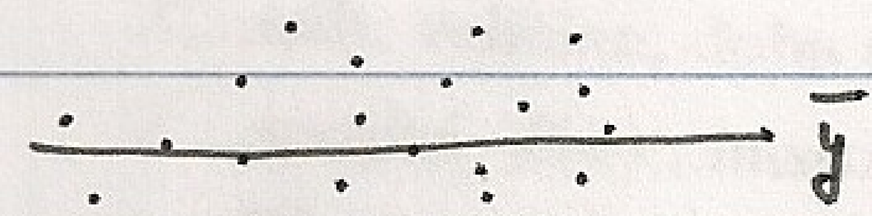
$r \to \pm 1 \Rightarrow RSS \to 0$

$\Rightarrow R^2 \to 1$

(téměř 100 %)

$r \approx 0$ $\hat{y} = \bar{y} \Rightarrow R^2 = 0$

(téměř 0%)



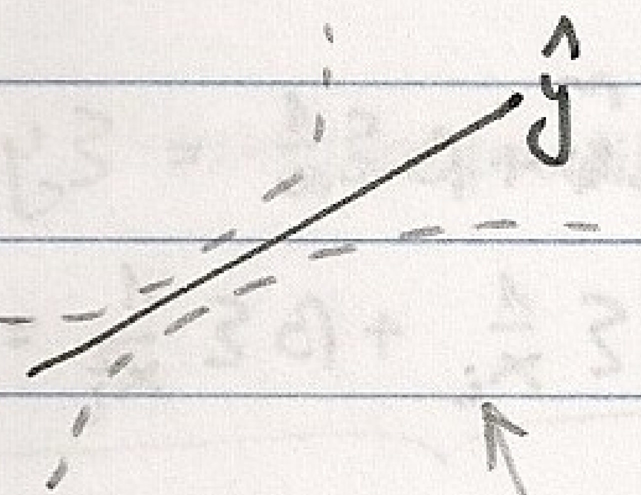
$\hat{y} = \beta_0 + \beta_1 x$ $\quad$ β_1 lze upravit na tvar

$$\beta_1 = r \cdot \frac{s_y}{s_x}$$

s_x, s_y - výběr. směrodatné odchylky

$r \longrightarrow 0 \Rightarrow \beta_1 \approx 0$ $\quad$ r - výběr. korel. koef.

$r \longrightarrow \pm 1 \Rightarrow \beta_1 \approx 0$



β_0, β_1 jsou bodové odhady $\Rightarrow$ lze konstruovat intervalové odhady

<u>pás spolehlivosti</u>: (má být symetrický)

$$\beta_0 + \beta_1 \tilde{x} + t_{m-2,\,1-\frac{\alpha}{2}} \sqrt{\frac{1}{m} + \frac{\tilde{x} - \bar{x}}{\sum x_i^2 - m\bar{x}}}$$

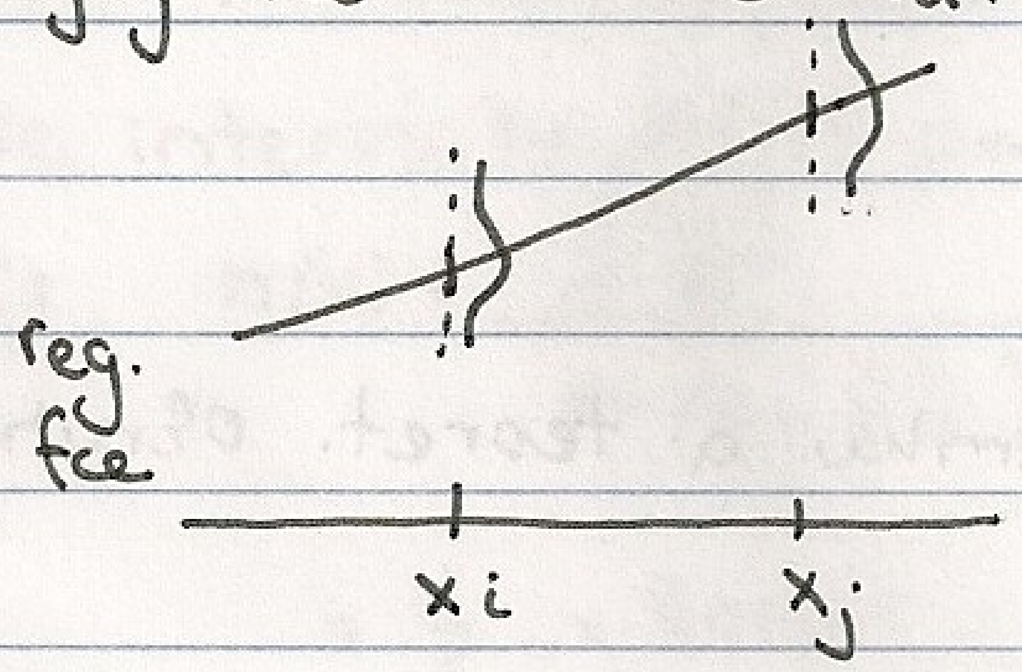
Pozn: 1, Jiné typy regresních fcí

$f(x) = a + e^{bx}$ $\quad$ $e \doteq 2{,}7180\ldots$

$f(x) = a + \frac{b}{x}$

$f(x, y) = a + bx + cy$ (regresní rovnice)

2, chyby měření "normálně" rozděleny

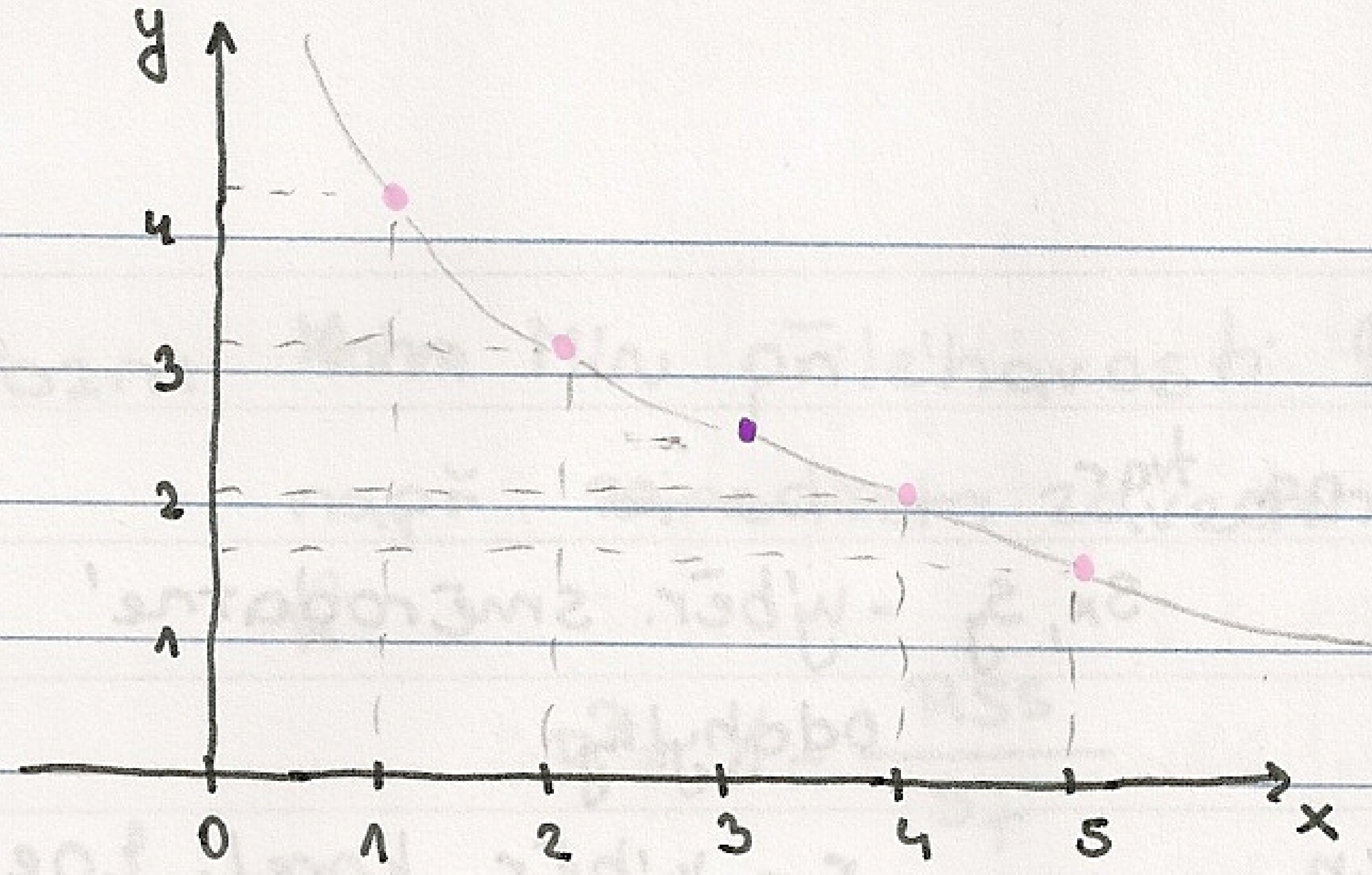


3, možnost zavést váhové fce $v(x_i) > 0$

(velký rozptyl $\Rightarrow$ malá váha)

př: Odvoďte soustavu normálních rovnic pro regresní hyperbolu $f(x): y = a + \frac{b}{x}$. Použijte ji pro naměřené hodnoty x_i, y_i a najděte odhad $\hat{f}(3)$ + graf.

x_i	1	2	4	5
y_i	4,25	3,14	2,16	1,85



Řešení:
$$S = \sum \left(\alpha + \frac{\beta}{x_i} + y_i\right)^2$$

$$\left.\begin{array}{l} \frac{\partial S}{\partial x} = 0 \\ \frac{\partial S}{\partial y} = 0 \end{array}\right\} \Leftrightarrow \left.\begin{array}{ll} 2 \cdot \sum \left(\alpha + \frac{\beta}{x_i} - y_i\right)^2 \cdot 1 = 0 & \frac{1}{2} \\ 2 \cdot \sum \left(x + \frac{y}{x} - y\right) \cdot \frac{1}{x} = 0 & \frac{1}{2} \end{array}\right\} \Leftrightarrow \underbrace{\begin{array}{l} \alpha m + \beta \sum \frac{1}{\alpha} = \sum y_i \\ \alpha \sum \frac{1}{x_i} + \beta \sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i} \end{array}}_{\text{soustava rovnic}}$$

$m = 4$

$\sum \frac{1}{x_i} = 1{,}95$

$\sum \frac{1}{x_i^2} = 1{,}3525$

$\sum y_i = 11{,}4$

$\sum \frac{y_i}{x_i} = 6{,}73$

$$\left[\begin{array}{cc|c} 4 & 1{,}95 & 11{,}4 \\ 1{,}95 & 1{,}3525 & 6{,}73 \end{array}\right]$$

Crammer:

$$\left.\begin{array}{l} \alpha \doteq 1{,}43 \\ \beta \doteq 2{,}92 \end{array}\right\} \hat{y} = 1{,}43 + \frac{2{,}92}{x}$$

$$\hat{y}(3) \doteq 2{,}40$$

KE ZKOUŠCE:

- dal na courseware ukázkovou písemku a teoret. otázky
- v ústní části dostaneme vzorce:

Intenzita poruch, Weibullovo rozdělení

X...náh. veličina, doba do poruchy (součástky, zařízení)
X...spojitá, $f(x)$...hustota, $F(x)$...distribuční fce,
(Protože $x > 0$ je čas, je $f(x) = 0$ pro $x \leq 0$.)

Definujme **spolehlivostní funkci** $R(x)$ vztahem

$$R(x) = P(X > x) = 1 - F(x)\,,$$

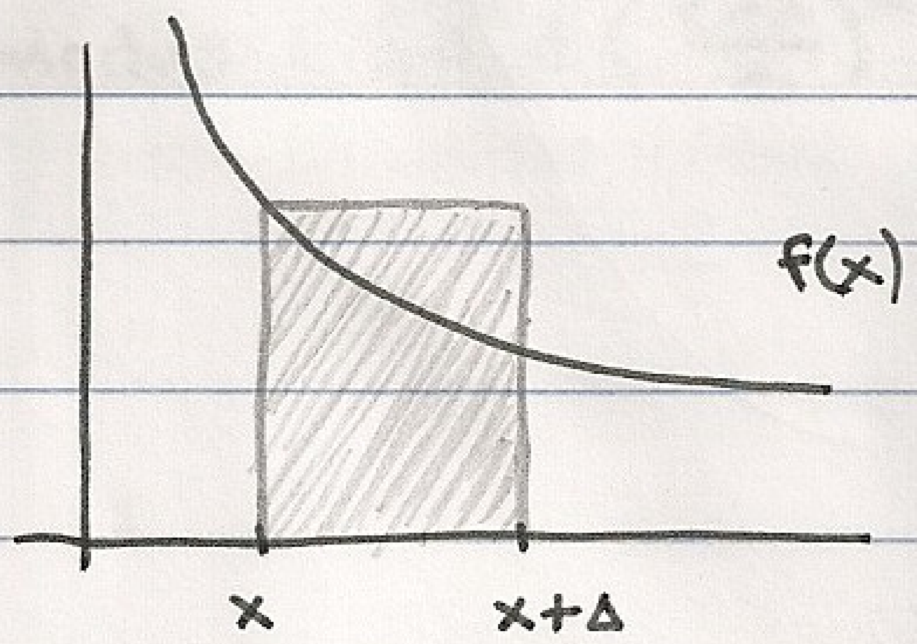
intenzitu poruch $h(x)$ vztahem

$$h(x) = \frac{f(x)}{R(x)} = \frac{f(x)}{1 - F(x)}\,.$$

Výpočtem pro libovolné $x_0 > 0$ dostaneme:

$$P(X < x_0 + \Delta \,|\, X > x_0) = \frac{P(x_0 < X < x_0 + \Delta)}{P(X > x_0)} \approx$$

$$\approx \frac{f(x_0)\,\Delta}{1 - F(x_0)} = h(x_0)\,\Delta$$

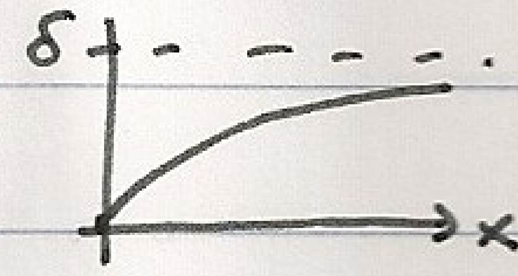


Př: Nechť $X \sim Exp(\delta)$, $\delta > 0$

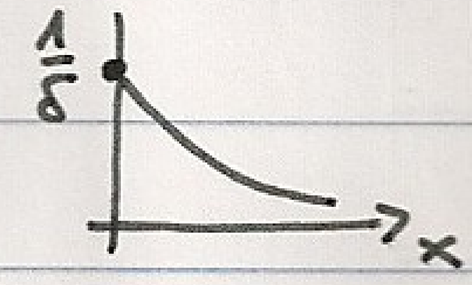
Určete intenzitu $h(x)$ pro X.

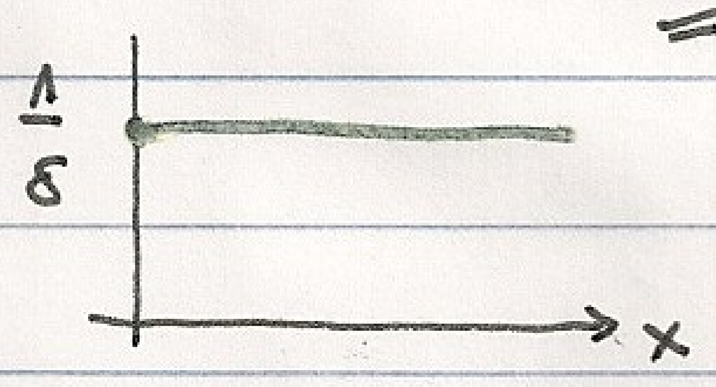
Řešení: $h(x) = \dfrac{\frac{1}{\delta} e^{-\frac{x}{\delta}}}{1 - (1 - e^{-\frac{x}{\delta}})} = \underline{\underline{\dfrac{1}{\delta}}} = \text{konst}$

$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\frac{x}{\delta}} & x > 0 \end{cases}$



$F^{-1}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{\delta} \cdot e^{-\frac{x}{\delta}} & \end{cases}$





Intenzita poruchy je stále konstantní.

př: Řekneme, že náh. spoj. veličina X má rozdělení bez paměti, platí-li, že

$$P(X > x_0 + \Delta \mid X > \Delta) = P(X > x_0) \quad \forall x_0, \Delta > 0$$

$\Rightarrow$ zapomene odpracovanou dobu

např: $X \sim Exp(\delta)$ je rozdělením bez paměti

$$P(X > x_0 + \Delta \mid X > x_0) = \frac{P(X > x_0 + \Delta)}{P(X > x_0)} = \frac{1 - F(x_0 + \Delta)}{1 - F(x_0)} =$$

$$= \frac{1 - (1 - e^{-\frac{x_0 + \Delta}{\delta}})}{1 - (1 - e^{-\frac{x_0}{\delta}})} = 1 - (1 - e^{-\frac{x_0}{\delta}}) = P(X > 0) \quad \forall x_0, \Delta > 0$$

11. přednáška

K popisu životnosti se často hodí použít tzv. Weibullovo rozdělení, které je zobecněním exponenciálního rozdělení.

WEIBULLOVO rozdělení s parametry $c > 0$, $\delta > 0$:

X má *distribuční funkci*

$$F(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0] \\ 1 - e^{-(\frac{x}{\delta})^c} & \text{pro } x \in (0, +\infty) \end{cases}$$

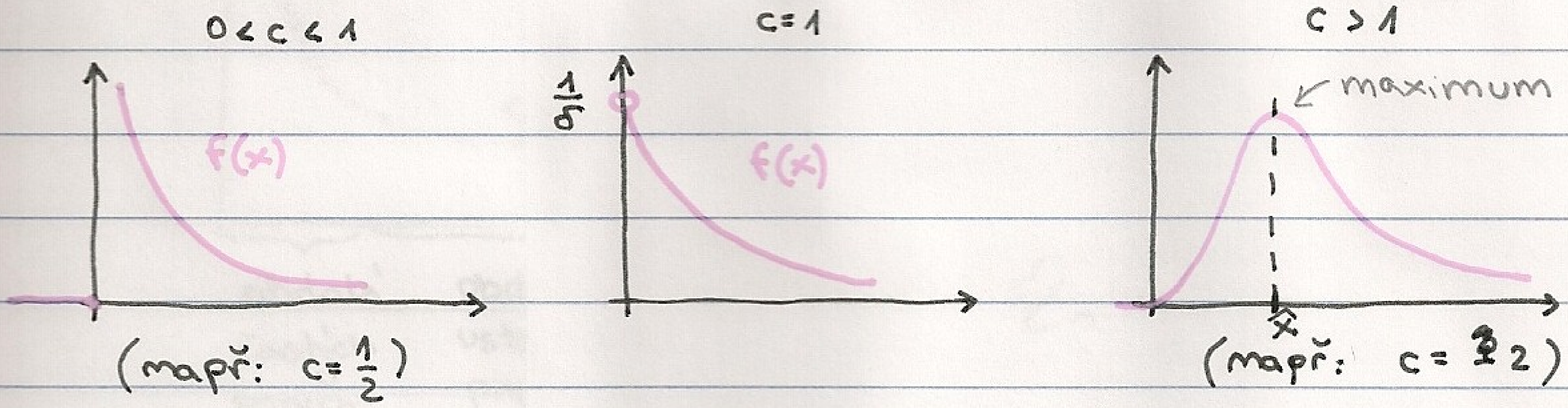
Píšeme: $X \sim W(c, \delta)$.

[$W(1, \delta)$ je zřejmě $Exp(\delta)$.]

Hustota ppsti:

$$f(x) = F'(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0] \\ \frac{c}{\delta^c}\, x^{c-1}\, e^{-(\frac{x}{\delta})^c} & \text{pro } x \in (0, +\infty) \end{cases}$$

Grafy hustot:



- zkoumáme body podezřelé z extrému (tam, kde je derivace nulová)

→ počítáme druhou nebo třetí derivaci

$$f'(x) = \frac{c}{\delta^c} e^{-\left(\frac{x}{\delta}\right)^c} \left[(c-1)x^{c-2} + x^{c-1}\left(c\left(\frac{x}{\delta}\right)^{c-1} \cdot \frac{1}{\delta}\right)\right] = \underbrace{\frac{c}{\delta^c} e^{-\left(\frac{x}{\delta}\right)^c}}_{>0} \underbrace{x^{c-2}}_{>0} \left[c-1 + x^c \cdot \frac{c}{\delta^c}\right]$$

$f'(x) = 0 \iff [\,\dots\,] = 0$

modus : $\hat{x} = \delta\left(\frac{c-1}{c}\right)^{\frac{1}{c}}$ (extrém fce)

např: $\left.\begin{array}{l} c = 2 \\ \delta = 2 \end{array}\right\} \Rightarrow \hat{x} = 2\left(\frac{1}{2}\right)^{\frac{1}{2}} = 2 \cdot \sqrt{\frac{1}{2}} = \sqrt{2}$

Kvantily $W(c, \delta)$:

$p \in (0,1)$: $F(x_p) = p$

$1 - e^{-\left(\frac{x}{\delta}\right)^c} = p$

$x_p = \delta\left[-\ln(1-p)\right]^{\frac{1}{c}}$

např: medián $x_{0,5}$ pro $c = 2$, $\delta = 2$

$$x_{0,5} = 2 \cdot \left[-\ln(1-0{,}5)\right]^{\frac{1}{2}} = 2 \cdot \sqrt{\ln 2}$$

$\boxed{\hat{x} \neq x_{0,5}}$!

Je-li $X \sim W(c, \delta)$, pak intenzita poruch (viz (2)):

$$h(x) = \frac{f(x)}{1-F(x)} = \frac{c}{\delta}\left(\frac{x}{\delta}\right)^{c-1}$$

Pro $c = 1$ [tj. je-li $X \sim Exp(\delta)$] dostaneme :

$h(x) = \frac{1}{\delta}$ (= **konstantní** funkce) pro $x > 0$.

Pro $0 < c < 1$ je $h(x)$ **klesající** funkce pro $x > 0$.

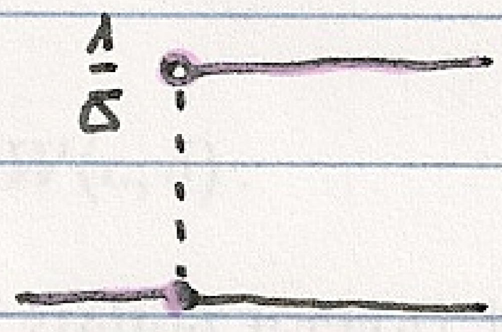
Pro $c > 1$ je $h(x)$ **rostoucí** funkce pro $x > 0$.

V praxi často:
nejprve $0 < c < 1$ (etapa častých poruch),
pak $c \approx 1$ (etapa ustáleného provozu),
nakonec $c > 1$ (etapa stárnutí).

Grafem $h(x)$ je pak tzv. **vanová křivka**.

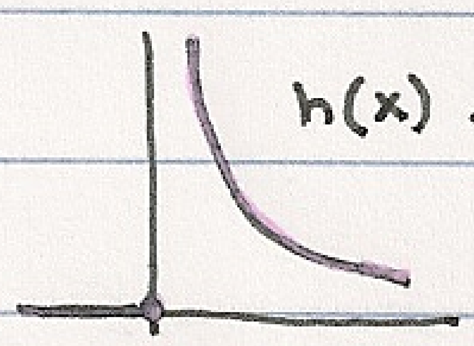
$$h(x) = \frac{\frac{c}{x^{\delta}} x^{c-1} \cdot \cancel{e^{-\left(\frac{x}{\delta}\right)^c}}}{1-(1-\cancel{e^{-\left(\frac{x}{\delta}\right)^c}})} = \frac{c}{\delta}\left(\frac{x}{\delta}\right)^{c-1}$$

1, $c = 1$



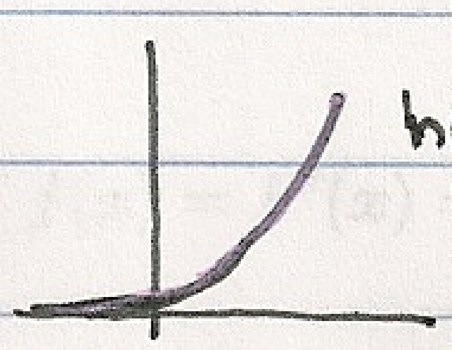
$h(x)$ ... konst

2, $0 < c < 1$

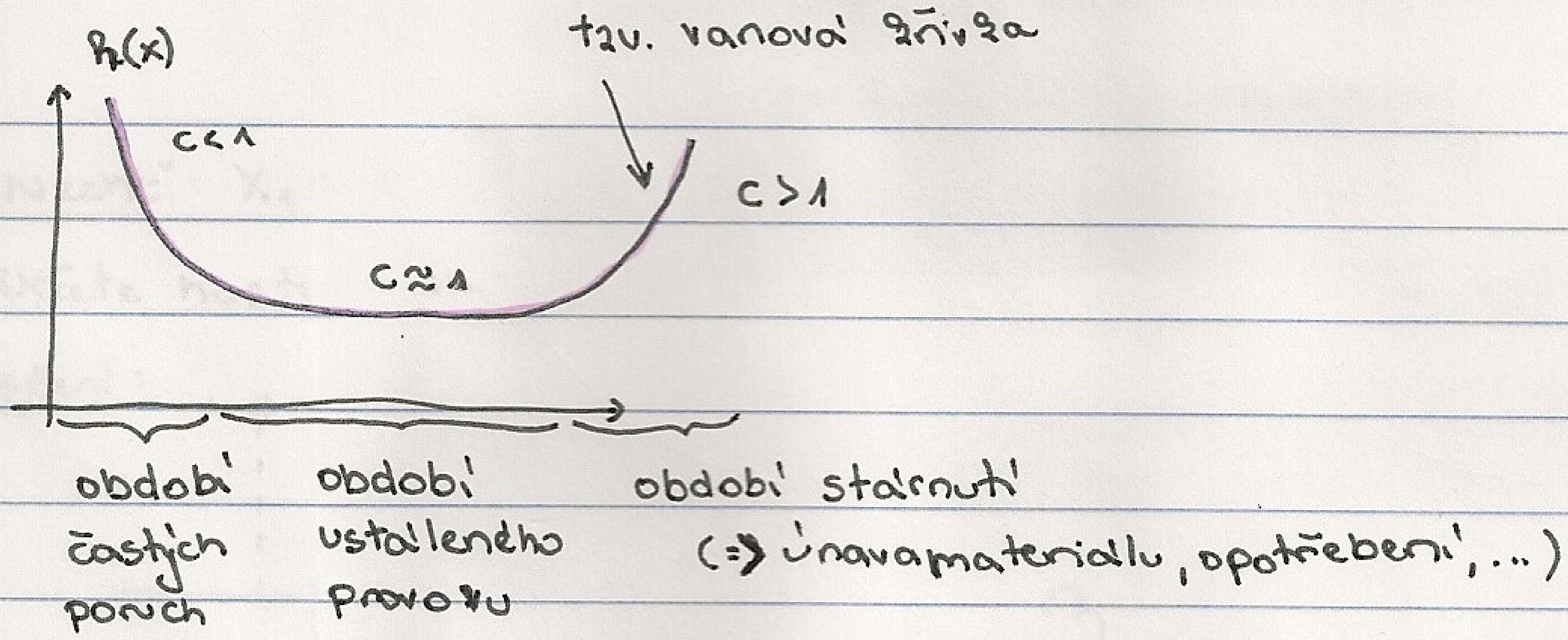


$h(x)$ ... kles

3, $c > 1$



$h(x)$ ... rost



př: ○○○○○○○ „řetěz" - pevnost závisí na pevnosti nejslabšího článku

$X_i \sim W(c, \delta)$ nezávislé $i = 1, \dots m$

$X = \min(X_1, X_2, \dots X_m)$ s distrib. fcí $F(x)$

určíme rozdělení veličiny X:

$$P(X > x) = P(X_1 > x, X_2 > x, \dots, X_m > x) \underset{\text{nezávislé}}{=} \prod_{i=1}^{m} P(X_i > x)$$ odtud

$$F(x) = P(X \le x) = 1 - P(X > x) = 1 - \prod_{j=1}^{m}(1 - F_j(x)) = 1 - \prod_{j=1}^{m} e^{-\left(\frac{x}{\delta}\right)^c} =$$

$$= 1 - e^{-\sum_{i=1}^{m}\left(\frac{x}{\delta}\right)^c} = 1 - e^{-m\left(\frac{x}{\delta}\right)^c} = 1 - e^{-\left(m^{\frac{1}{c}} \cdot \frac{x}{\delta}\right)^c}$$ což je

distribuční fce rozdělení $W\left(c, \frac{\delta}{m^{\frac{1}{c}}}\right)$ (Weibullovo)

- minimum nezávislých veličin, které mají Weibullovo rozdělení má také Weibullovo rozdělení, pouze s jiným parametrem

spec.: $c = 1$ $\quad X_i \sim Exp(\delta)$ nezávislé $\Rightarrow X \sim Exp\left(\frac{\delta}{m}\right)$

$\uparrow \min(X_1, \dots X_m)$

Součet náhodných veličin, gama rozdělení

X_1 ... spojitá s hustotou ppsti $f_1(x_1)$ a distr. fcí $F_1(x_1)$

X_2 ... spojitá s hustotou ppsti $f_2(x_2)$ a distr. fcí $F_2(x_2)$

$\vec{X} = (X_1, X_2)$ ... se sdruženou hustotou ppsti $f(x_1, x_2)$

Předpokládejme, že X_1, X_2 jsou **nezávislé**.
[Pak $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$].

Nechť $\mathbf{Y = X_1 + X_2}$.

Pak Y je náh. veličinou, označme:
$g(y)$ ... hustota Y,
$G(y)$ ... distrib. fce Y

$g(y) = ?$
$G(y) = ?$

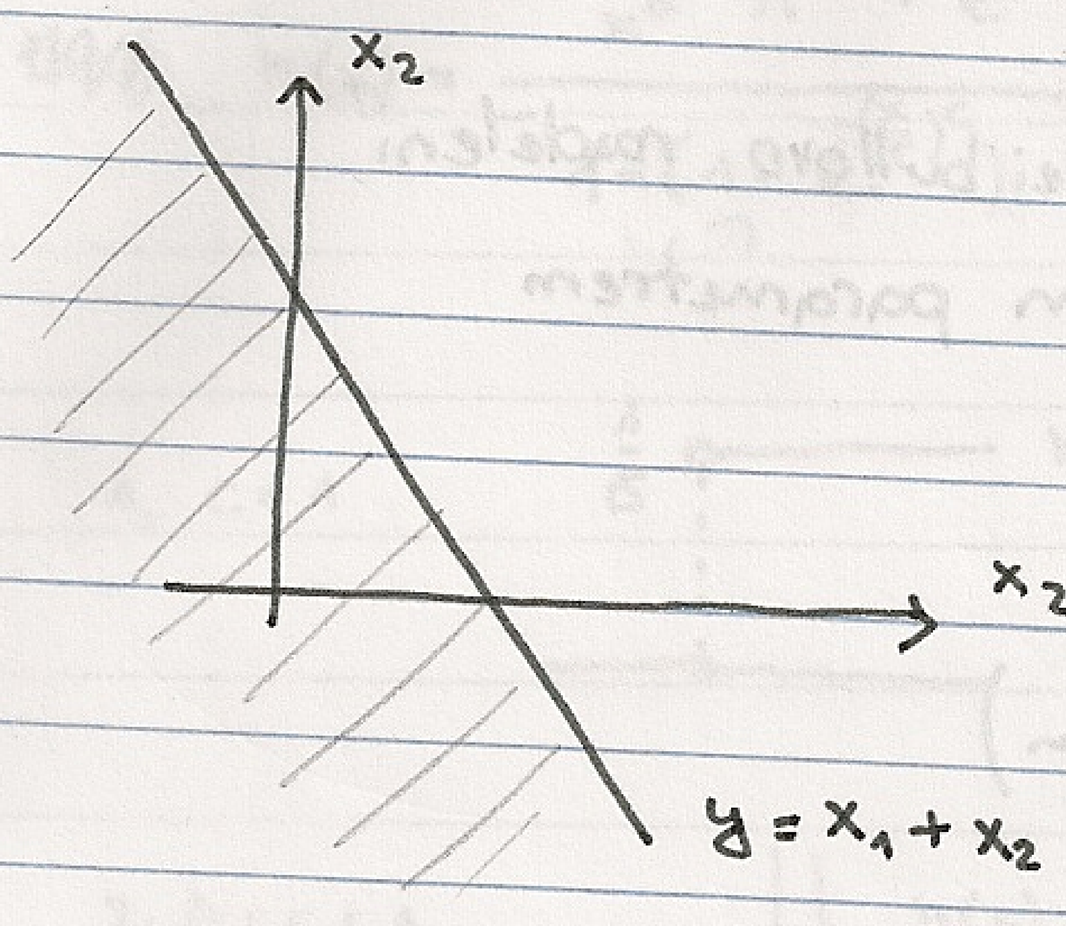
Přímým výpočtem $G(y)$ a její derivací dle y dostaneme

$$g(y) = \int_{-\infty}^{+\infty} f_1(y - x_2) \cdot f_2(x_2)\, dx$$

Výraz vpravo je tzv. KONVOLUCE hustot f_1, f_2, píšeme $(\mathbf{f_1 * f_2})(\mathbf{y})$.
Platí $(f_1 * f_2)(y) = (f_2 * f_1)(y)$ ("symetrie").

$$G(y) = P(Y \le y) = P(X_1 + X_2 \le y) = \iint\limits_{\substack{\mathbb{R}^2 \\ x_1 + x_2 \le y}} f(x_1, x_2)\, dx_1\, dx_2 =$$

$x_1 + x_2 \le y \Rightarrow x_1 \le y - x_2$



$$\int_{-\infty}^{+\infty} \Big(\underbrace{\int_{-\infty}^{y - x_2} f_1(x_1)\, dx_1}_{F_1(y - x_2)} \Big) f_2(x_2)\, dx_2 =$$

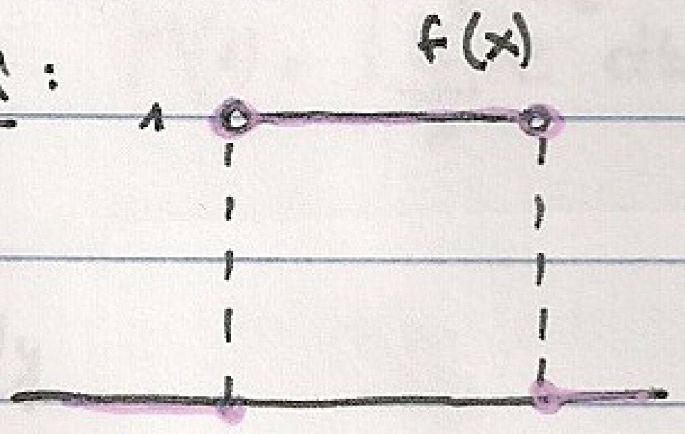
$$g(y) = \frac{dG}{dy} = \int_{-\infty}^{\infty} f_1(y - x_2) \cdot f_2(x_2)\, dx_2$$

záměna $\int .. \frac{d}{dy}$

Př: Nechť $X_1 \sim R(0,1)$, $X_2 \sim R(0,1)$ nezávislé.

Určete hustotu rozdělení $Y = X_1 + X_2$ a načrtněte její graf.

Řešení:

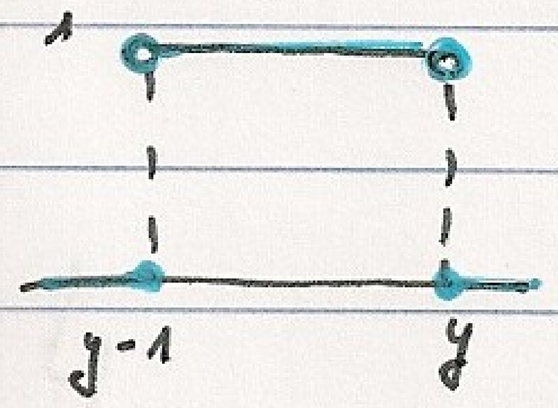


X_1, X_2

$$f(x) = \begin{cases} 1 & \dots\ 0 < x < 1 \\ 0 & x \notin (0,1) \end{cases}$$

$$f(y-x) = \begin{cases} 1 & \dots\ 0 < y - x < 1 \\ 0 & \text{jinde} \end{cases} \Rightarrow -y < -x < -y+1 \qquad y > x > y-1$$

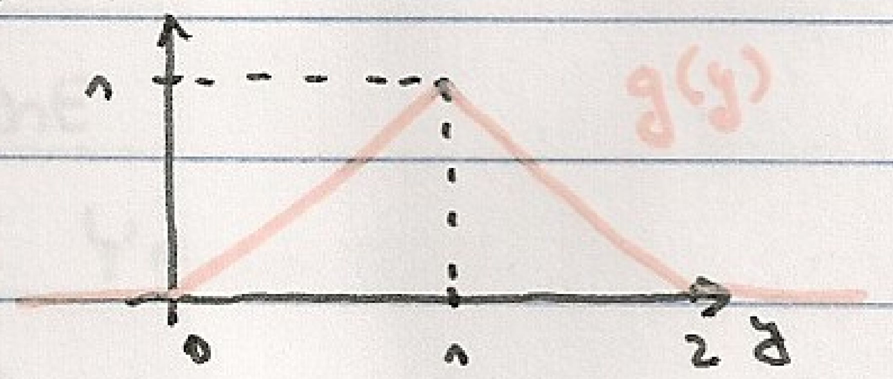
$$g(y) = \int_{-\infty}^{\infty} f(y-x) \cdot f(x)\, dx$$

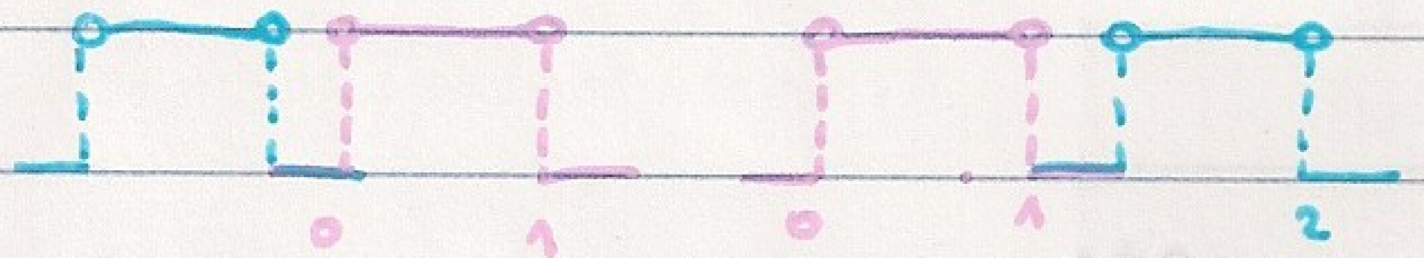


$$= \begin{cases} \text{pro } y \in \langle 0,1\rangle: \displaystyle\int_0^y 1\, dx = y \\ \text{pro } y \in \langle 1,2\rangle: \displaystyle\int_{y-1}^{1} 1\, dx = 2-y \end{cases}$$

$$f(y-x)\cdot f(x) = \begin{cases} 1 & \text{pro } x \in (\max(0, y-1); \min(1,y)) \\ 0 & \dots \text{pro } y \leq 0 \text{ nebo } y \geq 2 \end{cases}$$

Graf:





pozn: čím více ~~...~~ nezávislých veličin nasčítáme, tím více se bude $g(y)$ podobat Gaussově křivce

Př: Nechť $X_i \sim Exp(\delta)$, pro $i = 1,2$, nezávislé.

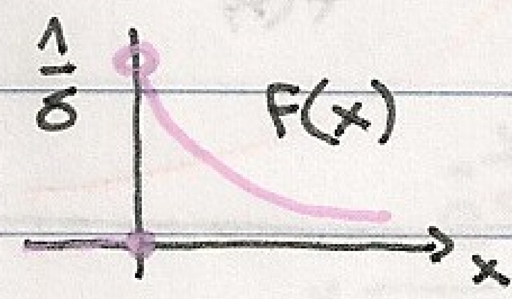
Určete hustotu $g(y)$ rozdělení $Y = X_1 + X_2$

Řešení:

$$g(y) = \int_{-\infty}^{\infty} f(y-x)\, f(x)\, dx =$$

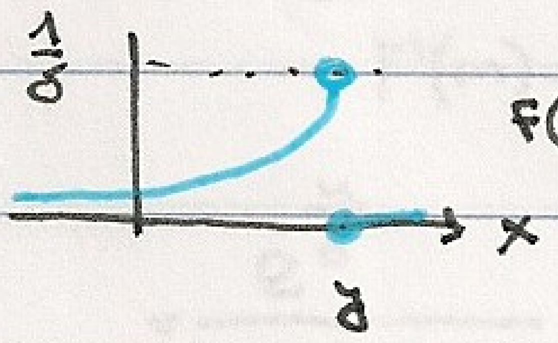
$$= \int_0^y \frac{1}{\delta} e^{-\frac{y-x}{\delta}} \cdot \frac{1}{\delta} e^{-\frac{x}{\delta}}\, dx =$$

$$= \frac{1}{\delta^2} \int_0^y e^{-\frac{y}{\delta}}\, dx = \frac{y \cdot e^{-\frac{y}{\delta}}}{\delta^2} \ \dots \text{pro } y > 0$$



$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{\delta} e^{-\frac{x}{\delta}} & x > 0 \ \dots\ y - x > 0 \end{cases}$$

$x < y$

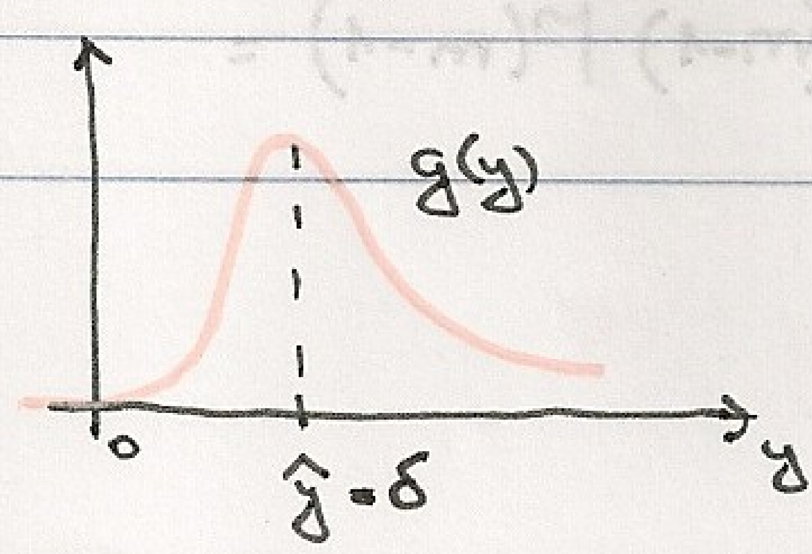


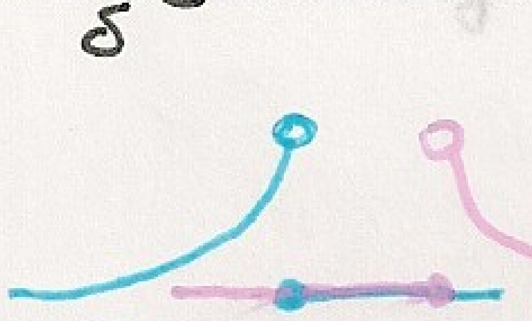
• růžová je pevná, modrou mohu hýbat

$$f(y-x) = \begin{cases} 0 & \dots\ x \leq 0 \\ \frac{1}{\delta} e^{-\frac{y-x}{\delta}} & \dots\ y - x > 0 \end{cases}$$

$x < y$

graf:





=)

$$g'(y) = \underbrace{\frac{1}{\delta^2}\cdot e^{-\frac{y}{\delta}}}_{0}\underbrace{\left(1-\frac{y}{\delta}\right)}_{0} \Rightarrow \hat{y} = \delta$$

Součet n nezávislých náhodných veličin s rozdělením $Exp(\delta)$ má **gama rozdělení** s parametry m, δ.

GAMA rozdělení s parametry $m > 0$, $\delta > 0$:

X má *hustotu ppsti*

$$f(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0] \\ \frac{x^{m-1}}{\Gamma(m)\,\delta^m}\, e^{-\left(\frac{x}{\delta}\right)} & \text{pro } x \in (0, +\infty) \end{cases}$$

Píšeme: $X \sim \Gamma(m, \delta)$.

Poznámky:

1) **Gama funkce** $\Gamma(m)$ je definována takto:

$$\Gamma(m) = \int_0^{+\infty} t^{m-1} \cdot e^{-t}\, dt.$$

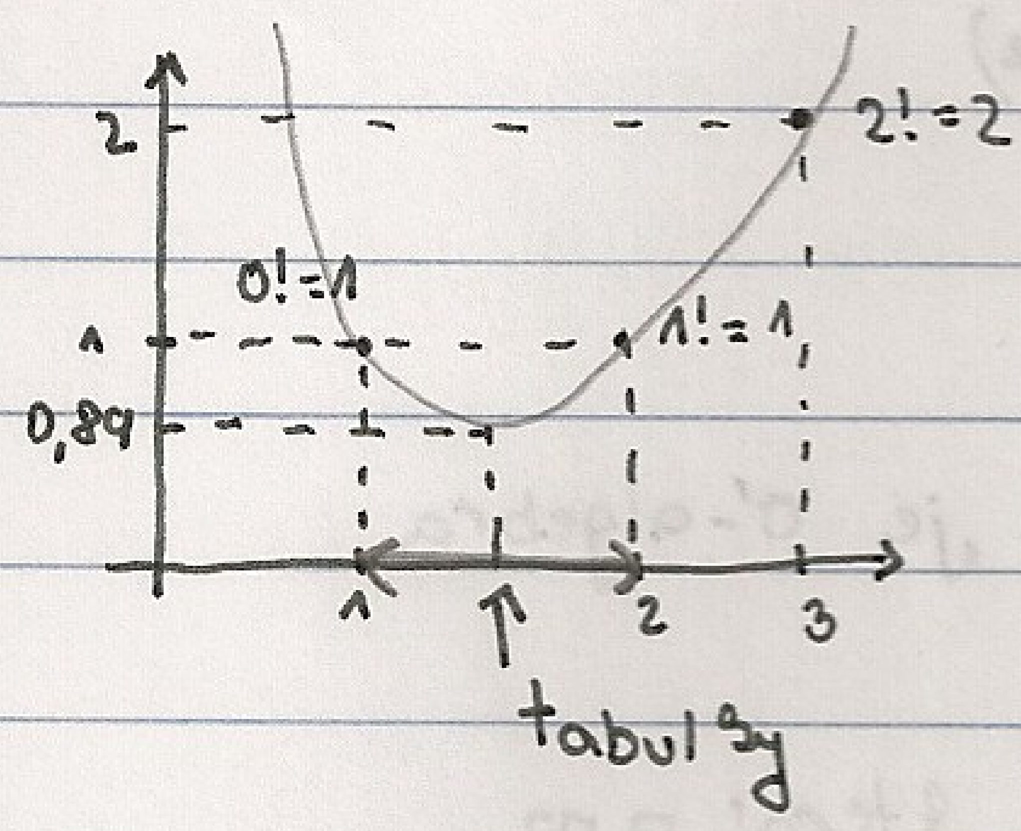
2) Volbou $m = \frac{\nu}{2}$, $\delta = 2$, kde $\nu \in N$, dostaneme χ^2-**rozdělení** (s ν stupni volnosti)

$= (m-1)(m-2)\,\Gamma(m-2) = \dots = (m-1)(m-2)\dots 3\cdot 2\cdot \Gamma(1)$

přitom $\Gamma(1) = \int_0^{\infty} \underbrace{t^0}_{1} e^{-t}\,dt = \left[\frac{e^{-t}}{-1}\right]_0^{\infty} = 0-(-1) = 1$

lze odvodit: $\Gamma(m) = m\,\Gamma(m-1)$, $m>0$

a tedy pro $m = n \in \mathbb{N}$ je $\Gamma(n) = (m-1)!$ $\quad \Gamma(\frac{1}{2}) = \sqrt{2}$



Pozn: Rozdělení z předchozího příkladu (zřejmě) $Y \sim \Gamma(2, \delta)$

Př: Nechť $X_i \sim Exp(\delta)$, $i = 1, \dots m$, nezávislé, $Y = \sum_{i=1}^{m} X_i$
$(m \geq 2)$

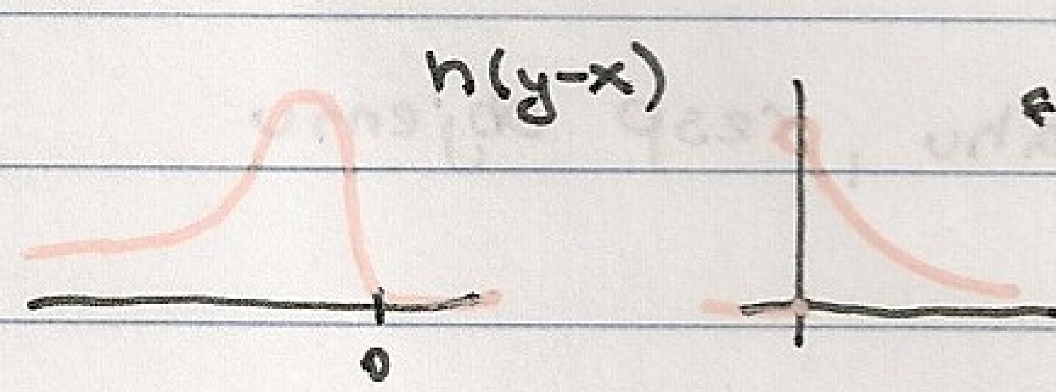
Pak $Y \sim \Gamma(m, \delta)$

<u>Důkaz:</u> pro $m=2$ platí - viz předchozí příklad (matematickou indukcí)

Nechť platí pro $\underline{m}$ .., dokážeme platnost pro $m+1$

$$Y = \sum_{i=1}^{m+1} X_i = \underbrace{\sum_{i=1}^{m} X_i}_{\sim \Gamma(m,\delta)} + \underbrace{X_{m+1}}_{\sim Exp(\delta)}$$

$$h(x) = \frac{x^{m-1} \cdot e^{-\frac{x}{\delta}}}{\Gamma(n) \cdot \delta^n} \qquad f(x) = \frac{1}{\delta} e^{-\frac{x}{\delta}} \quad .. \; x>0$$



$$g(y) = (h * f)(y) = \int_{-\infty}^{\infty} h(y-x) f(x)\,dx = \int_0^y \frac{(y-x)^{m-1} \cdot e^{-\frac{y-x}{\delta}}}{\Gamma(m)\,\delta^m} \cdot \frac{1}{\delta} e^{-\frac{x}{\delta}}\,dx =$$

$$= \frac{e^{-\frac{y}{\delta}}}{\Gamma(m) \cdot \delta^{m+1}} \int_0^y (y-x)^{m-1}\,dx =$$

$$= \frac{e^{-\frac{y}{\delta}}}{\Gamma(m)\,\delta^{m+1}} \left[\frac{(y-x)^m}{-m}\right]_{x=0}^{x=y} = \frac{y^m \cdot e^{-\frac{y}{\delta}}}{\Gamma(m) \cdot \delta^{m+1} \cdot m}$$

... což je hustota $\Gamma(m+1, \delta)$

$\Gamma(m+1)$

Axiomatická (kolmogorovova) definice ppsti.

Definice: Nechť Ω je libovolná neprázdná množina.

Pak neprázdný systém $\mathcal{A}$ podmnožin množiny Ω se nazývá σ-algebra, jestliže platí:

$$A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A} \quad (1)$$

$$A_i \in \mathcal{A},\ i=1,2,\dots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \quad (2)$$

Definice: (kolmogorov)

Nechť Ω je neprázdná množina, nechť $\mathcal{A}$ je σ-algebra náhodných jevů definovaných na Ω.

Pak reálnou funkci $P(A)$ definovanou na $\mathcal{A}$, která pro $A \in \mathcal{A}$, $A_1, A_2, \dots \in \mathcal{A}$, $A_i \cap A_j \neq 0$ pro $i \neq j$ splňuje:

$$P(\Omega) = 1 \quad (A1)$$

$$P(A) \geq 0 \quad (A2)$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (A3)$$

nazveme pravděpodobností. Trojici $(\Omega, \mathcal{A}, P)$ označujeme jako pravděpodobnostní prostor.

Geometrická ppst

Nechť $\Omega \subset \mathbb{R}^1$, resp. $\Omega \subset \mathbb{R}^2$, resp $\Omega \subset \mathbb{R}^3$,