Jak modelovat výsledky náhodných pokusů?

Martina Litschmannová

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

Opakování: Základní pojmy z teorie pravděpodobnosti

- **Co je to náhodný pokus?**

  Děj, jehož výsledek není předem jednoznačně určen podmínkami, za nichž probíhá.

- **Co je to náhodný jev?**

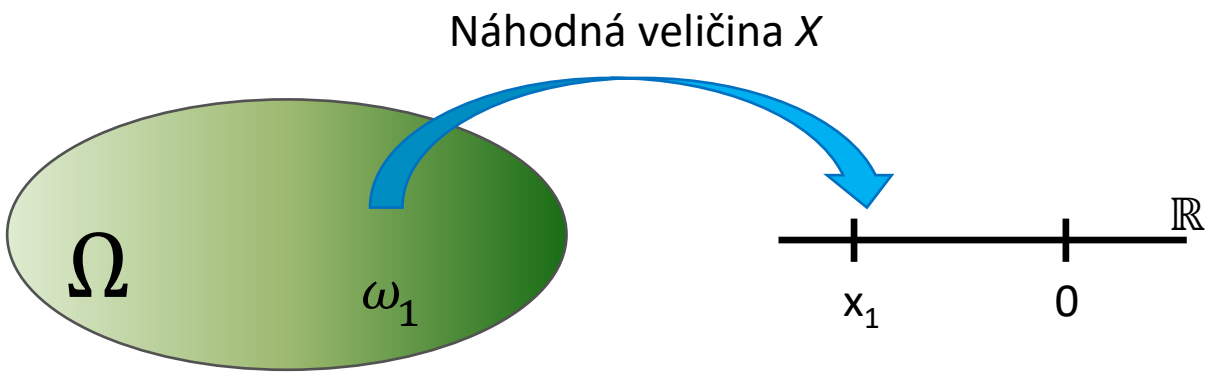
  Tvrzení o výsledku náhodného pokusu, přičemž o pravdivosti tohoto tvrzení lze po ukončení pokusu rozhodnout.

Jak modelovat výsledky náhodných pokusů?

Definice

Náhodná veličina je funkce, která zobrazuje elementární jevy $\omega \in \Omega$ na reálná čísla.

- NV obvykle značíme velkými písmeny (X, Y, Z).
- NV přiřadí každému elementárnímu jevu ω reálné číslo (převádí elementární jevy (abstraktními objekty) na čísla).
- Hodnota $X(\omega)$ náhodné veličiny X závisí na tom, který elementární jev ω nastal.
- Víme-li, který elementární jev ω nastal, známe hodnotu $X(\omega)$ náhodné veličiny X.



Zápis jevů pomocí náhodných veličin

Zápisem $(X = a)$ rozumíme jev složený ze všech elementárních jevů $\omega \in \Omega$, pro které $X(\omega) = a$,

$$\text{tj. } (X = a) = \{\omega \in \Omega : X(\omega) = a\}.$$

Podobně:

- $(X < a) = \{\omega \in \Omega : X(\omega) < a\}$,
- $(X > a) = \{\omega \in \Omega : X(\omega) > a\}$,
- $(a < X < b) = \{\omega \in \Omega : a < X(\omega) < b\}$, …

Náhodná veličina (NV)

Náhodná veličina – **číselné** vyjádření výsledku náhodného pokusu

$\boldsymbol{X}$ **… výsledek hodu kostkou**

Náhodný pokus: **Hod kostkou**

Náhodný jev: **Padne sudé číslo**. $(X \in \{2; 4; 6\})$

Náhodná veličina (NV)

Náhodná veličina – **číselné** vyjádření výsledku náhodného pokusu

X **… počet dívek mezi 1 000 náhodně vybranými dětmi**

Náhodný pokus: **Náhodný výběr 1 000 dětí a zjištění počtu dívek mezi nimi.**

Náhodný jev: **Mezi 1 000 náhodně vybranými dětmi bude více než 500 dívek**.

$(X > 500)$

Náhodná veličina (NV)

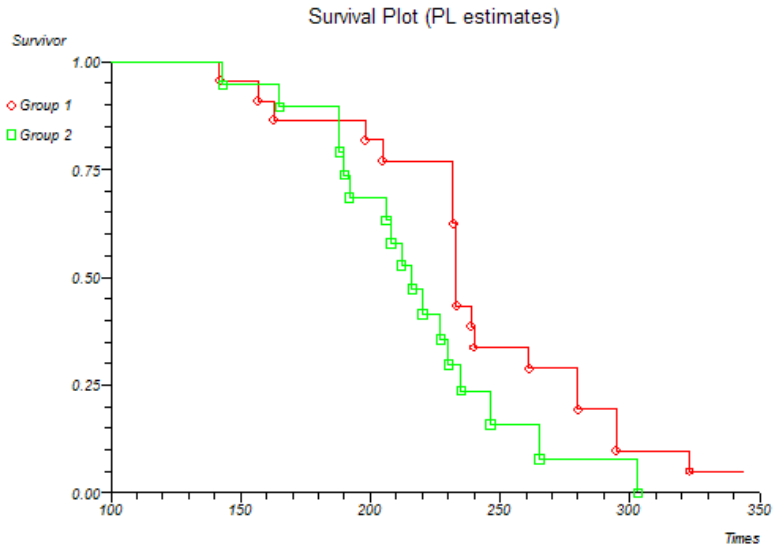
Náhodná veličina – **číselné** vyjádření výsledku náhodného pokusu

X **… doba do remise onemocnění**

Náhodný pokus: **Pozorování doby do remise onemocnění**

Náhodný jev: **Doba do remise nepřekročí 60 měsíců.**

$(X \leq 60)$



Náhodná veličina (NV)

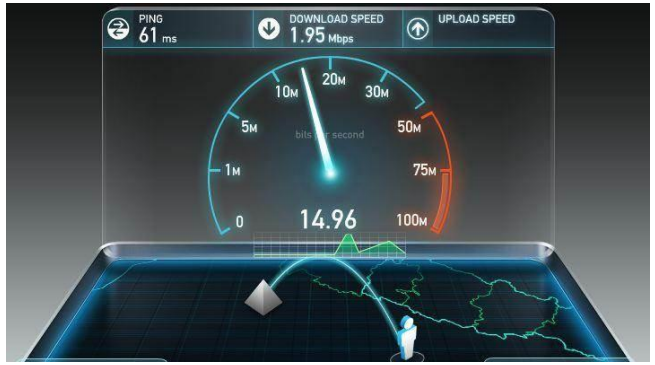
Náhodná veličina – **číselné** vyjádření výsledku náhodného pokusu

X **… rychlost připojení k internetu (Mb/s)**

Náhodný pokus: **Měření rychlosti připojení k internetu (download)**

Náhodný jev: **Rychlost připojení k internetu je vyšší než 20 Mb/s.**

$(X > 20)$



Náhodná veličina (NV)

Náhodná veličina – **číselné** vyjádření výsledku náhodného pokusu

POZOR!

Číselně vyjádření lze použít i u veličin, které nejsou kvantitativní ze své podstaty (např. pohlaví – muž (0), žena (1)).

Jak popsat chování náhodné veličiny?

Rozdělení pravděpodobnosti náhodné veličiny - předpis, který jednoznačně určuje všechny pravděpodobnosti typu

$$P(X \in M), \text{ kde } M \subset \mathbb{R}.$$

$$(\text{tj. } P(X = a), P(X < a), P(X > a), P(a < X < b), \ldots, \text{ kde } a, b \in \mathbb{R})$$

Jak zadat rozdělení pravděpodobnosti?

- funkci zadanou analyticky,
- výčtem hodnot NV a příslušných pravděpodobností,
- graficky.

Jak popsat chování náhodné veličiny?

Rozdělení pravděpodobnosti náhodné veličiny - předpis, který jednoznačně určuje všechny pravděpodobnosti typu

$$P(X \in M), \text{ kde } M \subset \mathbb{R}.$$

$$(\text{tj. } P(X = a), P(X < a), P(X > a), P(a < X < b), \ldots, \text{ kde } a, b \in \mathbb{R})$$

Jak zadat rozdělení pravděpodobnosti?

- distribuční funkcí,
- pravděpodobnostní funkcí (diskrétní NV),
- hustotou pravděpodobnosti (spojitá NV).

Distribuční funkce $F(x)$

Distribuční funkce $F(t)$ je pravděpodobnost, že náhodná veličina X bude **menší než** dané reálné číslo t.

$$\boxed{F(t) = P(X < t)}$$

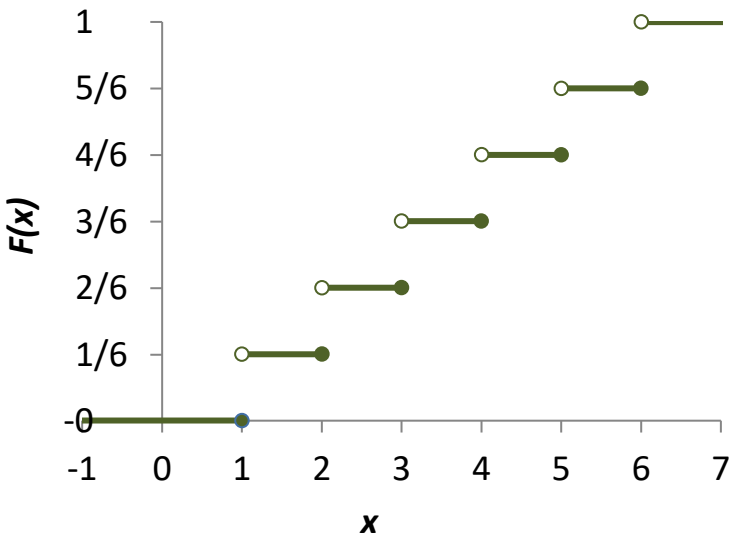
- $F(t)$ udává pravděpodobnost, že náhodná veličina nabývá hodnot menších než t.
- Distribuční funkce **jednoznačně určuje** rozdělení NV, tj. známe-li distribuční funkci, umíme určit pravděpodobnost $P(X \in M)$ pro libovolnou $M \subset \mathbb{R}$.
- Distribuční funkci náhodné veličiny X někdy značíme $F_X(t)$.

Někteří autoři definuji $F(t) = P(X \leq t)$!!!

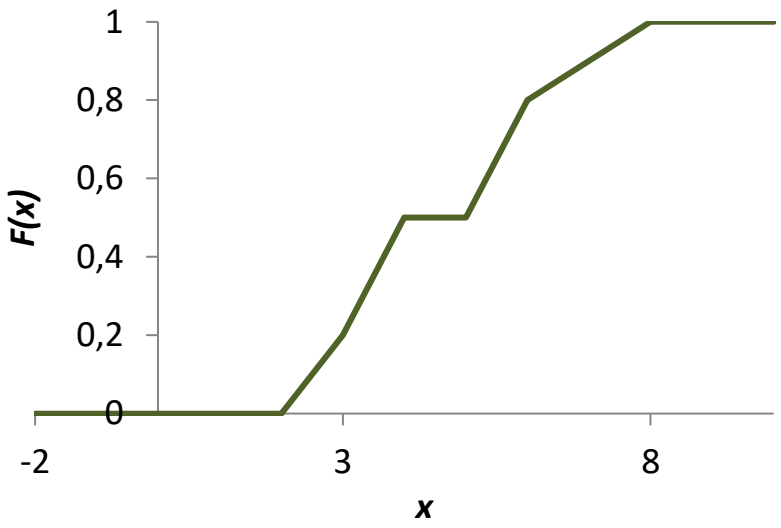
Distribuční funkce $F(x)$

Distribuční funkce $F(t)$ je pravděpodobnost, že náhodná veličina X bude **menší než** dané reálné číslo t.

$$F(t) = P(X < t)$$



Ukázka distribuční funkce diskrétní náhodné veličiny



Ukázka distribuční funkce spojité náhodné veličiny

Základní typy náhodných veličin

Diskrétní NV – mohou nabývat spočetně mnoha hodnot

Příklady: počet průjezdů automobilů Klimkovickým tunelem mezi 13:00 – 14:00, počet kybernetických útoků na určitý web během dne, počet dní nemocenské, počet zákazníků v lékárně během jednoho dne, …

Spojité NV – mohou nabývat všech hodnot na nějakém intervalu (mají spojitou distribuční funkci)

Příklady: doba do remise onemocnění, výška, váha, BMI, IQ, doba mezi dvěma následujícími kybernetickými útoky na web, chyba měření, …

Základní typy náhodných veličin

Diskrétní náhodná veličina (dále DNV)

- může nabývat spočetně mnoha hodnot
- DNV X s distribuční funkcí $F_X(t)$ je charakterizována pravděpodobnostní funkcí $P(X = x_i)$, tj. funkcí pro níž platí:

$$F_X(t) = \sum_{x_i<t} P(X = x_i) = \sum_{x_i<t} P(x_i)$$

Spojitá náhodná veličina (dále SNV)

- může nabývat všech hodnot na nějakém intervalu (má spojitou distribuční funkci)
- SNV X s distribuční funkcí $F_X(t)$ je charakterizována hustotou pravděpodobností $f(x)$, tj. funkcí pro níž platí:

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

Diskrétní náhodná veličina

Definice

Náhodná veličina X má diskrétní rozdělení pravděpodobnosti (zkráceně: „je diskrétní“), právě když nabývá spočetně mnoha hodnot.

DNV lze charakterizovat: pravděpodobnostní funkcí, distribuční funkcí

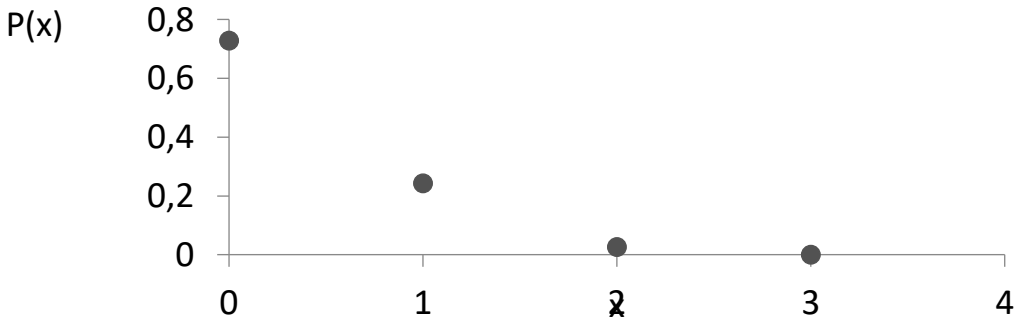
Pravděpodobnostní funkce

- $P(x_i) \geq 0$
- $\sum_i P(X = x_i) = 1$

- Lze zadat:
  - ✓ předpisem, $\forall x \in \{0; 1; 2; 3\}: P(X = x) = \binom{3}{x} \cdot 0{,}1^x \cdot 0{,}9^{3-x}$
  - ✓ tabulkou,

x	0	1	2	3
$P(x)$	0,729	0,243	0,027	0,001

  - ✓ grafem.



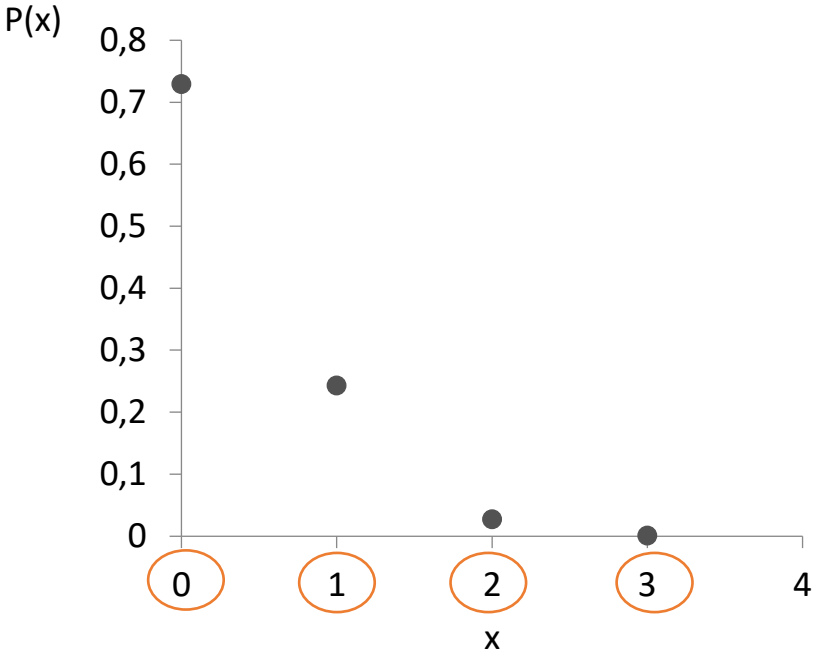
Příklad 1

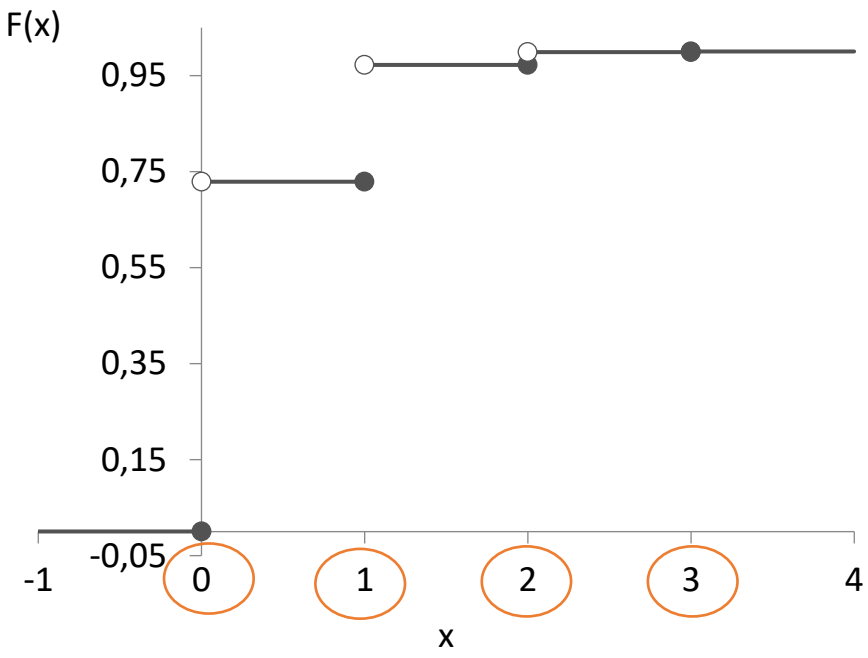
Při ověřování kvality výroby jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličino X. Z dlouhodobého pozorování jsou známy údaje uvedené v následující tabulce.

Počet vadných výrobků	Pravděpodobnost
0	0,25
1	0,50
2	0,25

Určete pravděpodobnostní a distribuční funkci počtu vadných výrobků v testovaném vzorku.

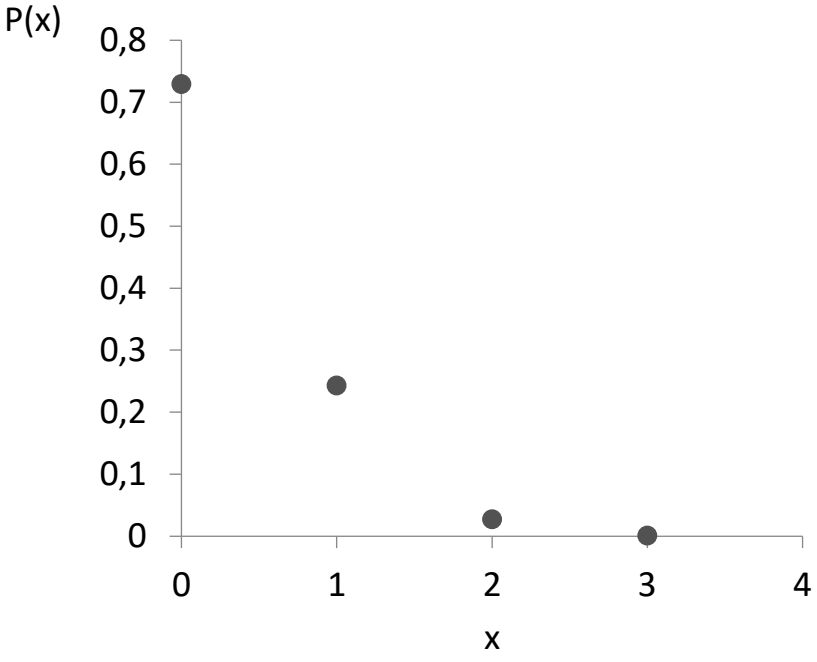
Vztah mezi pravděpodobnostní a distribuční funkci

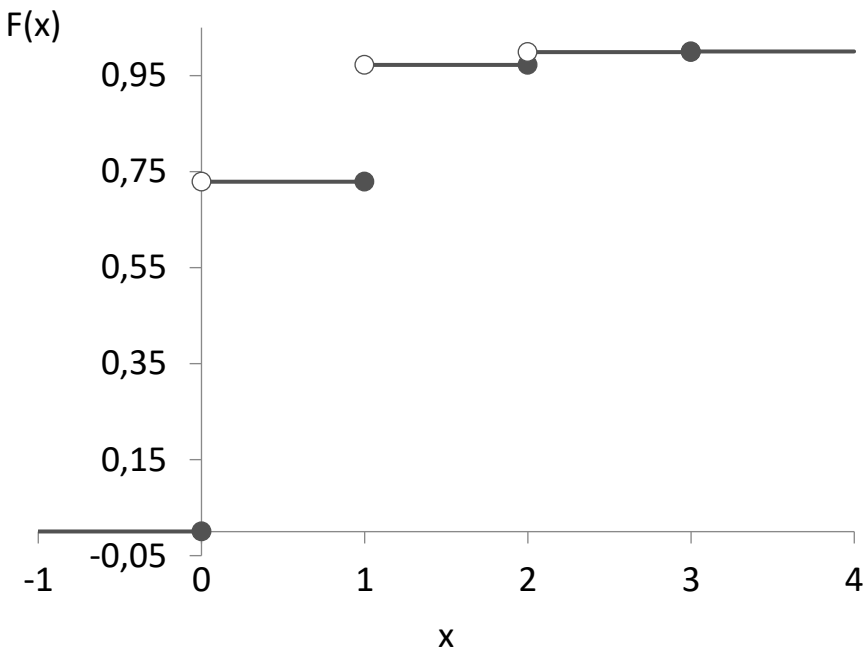




- Body nespojitosti distribuční f-ce jsou body, v nichž je pravděpodobnostní f-ce nenulová.

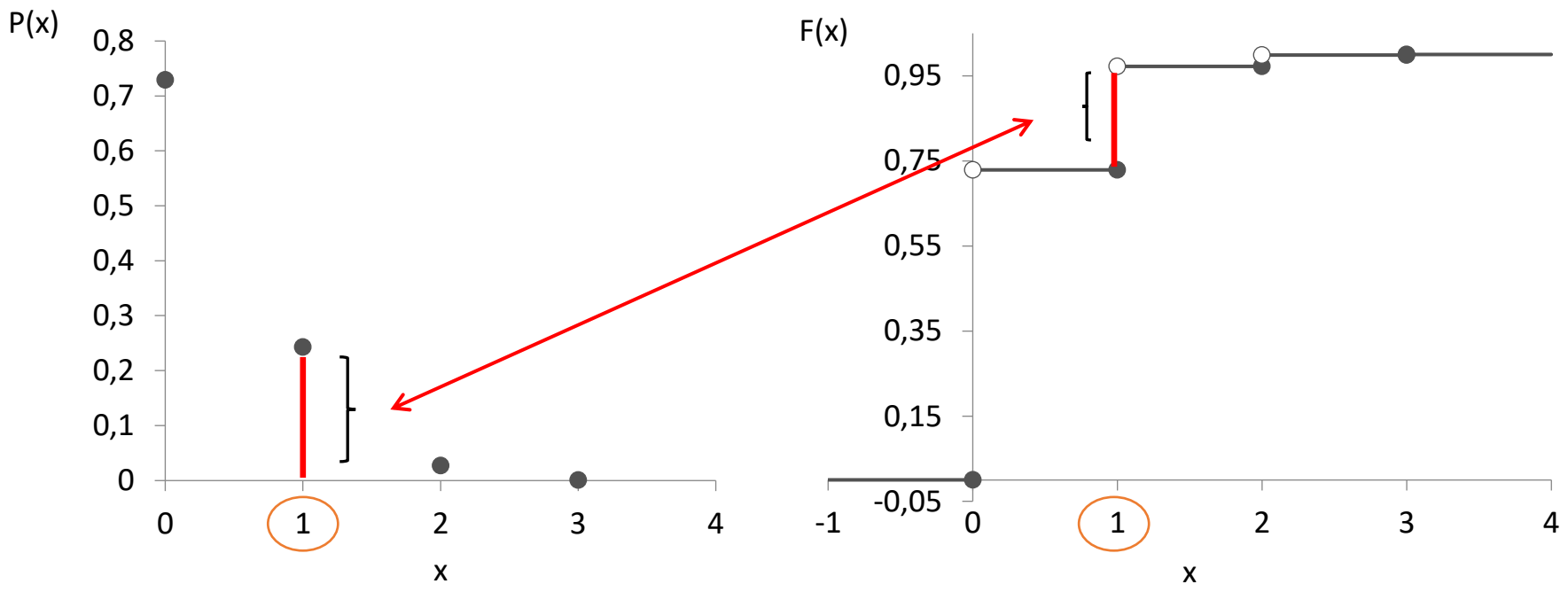
Vztah mezi pravděpodobnostní a distribuční funkci





- Body nespojitosti distribuční f-ce jsou body, v nichž je pravděpodobnostní f-ce nenulová.
- $P(X = a) = \lim_{x \to a+} F(x) - F(a)$

Vztah mezi pravděpodobnostní a distribuční funkci



- Body nespojitosti distribuční f-ce jsou body, v nichž je pravděpodobnostní f-ce nenulová.
- $P(X = a) = \lim_{x \to a+} F(x) - F(a)$, tj. velikost „skoku“ distribuční funkce v bodech nespojitosti je rovna příslušným hodnotám pravděpodobnostní funkce.

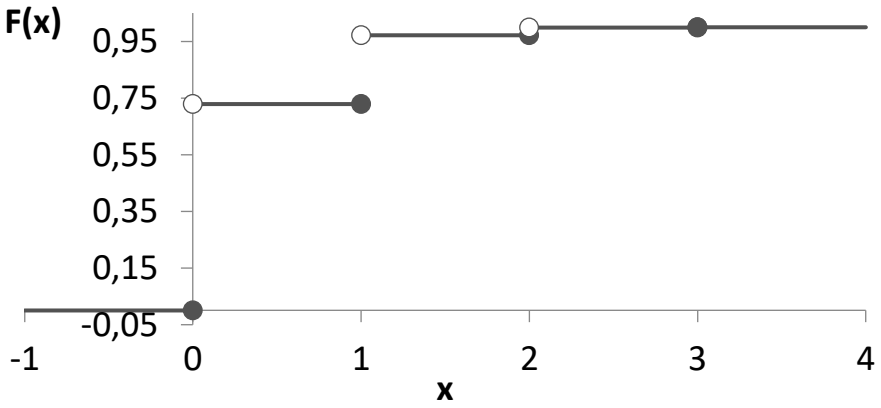
Distribuční funkce

- $F(t) = \sum_{x_i < t} P(x_i)$

- Lze zadat:
  - ✓ předpisem,
  - ✓ tabulkou,
  - ✓ grafem.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 0{,}729 & 0 < x \leq 1 \\ 0{,}972 & 1 < x \leq 2 \\ 0{,}999 & 2 < x \leq 3 \\ 1 & 3 < x \leq \infty \end{cases}$$

x	$F(x)$
$(-\infty; 0\rangle$	0
$(0; 1\rangle$	0,729
$(1; 2\rangle$	0,972
$(2; 3\rangle$	0,999
$(3; \infty)$	1



Distribuční funkce $F(x)$

Distribuční funkce $F(t)$ je pravděpodobnost, že náhodná veličina X bude **menší než** dané reálné číslo t.

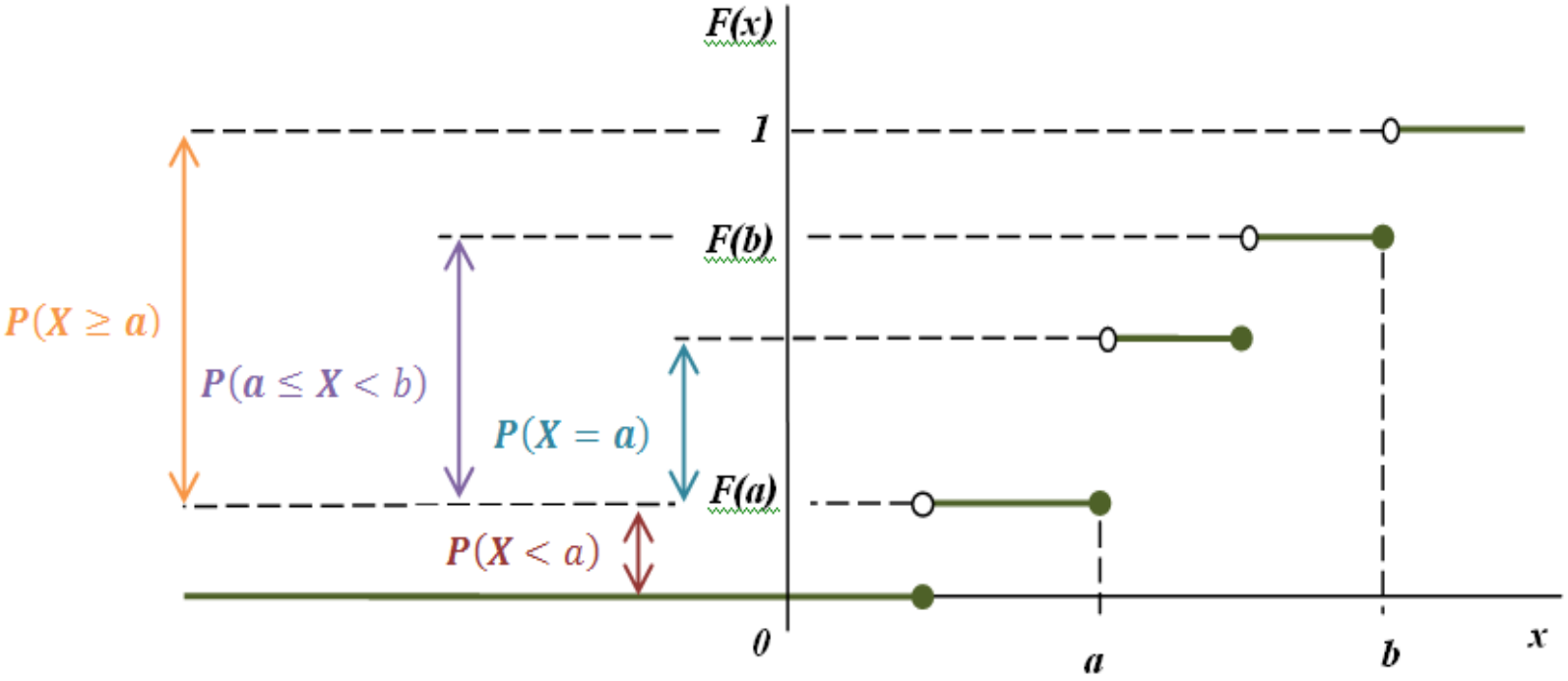
$$\boxed{F(t) = P(X < t)}$$

Vlastnosti distribuční funkce:

- $0 \leq F(t) \leq 1$,
- je neklesající,
- je zleva spojitá,
- má nejvýše spočetně mnoho bodů nespojitosti,
- $\lim\limits_{t \to -\infty} F(t) = 0$ („začíná“ v 0),
- $\lim\limits_{t \to \infty} F(t) = 1$ („končí“ v 1).

Někteří autoři definuji $F(t) = P(X \leq t)$!!!

Vztah mezi pravděpodobnostní a distribuční funkci



$$P(X = a) = \lim_{x \to a+} F(x) - F(a)$$

Vztah mezi pravděpodobnostní a distribuční funkci

- $P(X < a) = F(a) = \sum_{x_i < a} P(x_i)$
- $P(X \geq a) = 1 - P(X < a) = 1 - F(a) = \sum_{x_i \geq a} P(x_i)$
- $P(a \leq X < b) = P(X < b) - P(X < a) = F(a) - F(b) = \sum_{a \leq x_i < b} P(x_i)$
- $P(X = a) = \lim_{x \to a+} F(x) - F(a)$

Příklad 2

Při ověřování kvality výroby jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličino X, která má níže uvedenou distribuční funkci.

$$F(x) = \begin{cases} 0{,}00, x \in (-\infty; 0\rangle \\ 0{,}25, x \in (0; 1\rangle \\ 0{,}75, x \in (1; 2\rangle \\ 1{,}00, x \in (2; \infty). \end{cases}$$

Určete pravděpodobnostní funkci počtu vadných výrobků v testovaném vzorku.

Spojitá náhodná veličina

- $P(X < a) = F(a) = \sum_{x_i < a} P(x_i)$
- $P(X \geq a) = 1 - P(X < a) = 1 - F(a) = \sum_{x_i \geq a} P(x_i)$
- $P(a \leq X < b) = P(X < b) - P(X < a) = F(a) - F(b) = \sum_{a \leq x_i < b} P(x_i)$
- $P(X = a) = \lim_{x \to a+} F(x) - F(a)$

Spojitá náhodná veličina

Definice

Náhodná veličina X má spojité rozdělení pravděpodobnosti (zkráceně: „je spojitá“) právě když má spojitou distribuční funkci.

SNV lze charakterizovat: hustotou pravděpodobnosti, distribuční funkcí

Proč se pro popis SNV
nepoužívá
pravděpodobnostní funkce?

Spojitá náhodná veličina

Definice

Náhodná veličina X má spojité rozdělení pravděpodobnosti (zkráceně: „je spojitá“) právě když má spojitou distribuční funkci.

SNV lze charakterizovat: hustotou pravděpodobnosti, distribuční funkcí

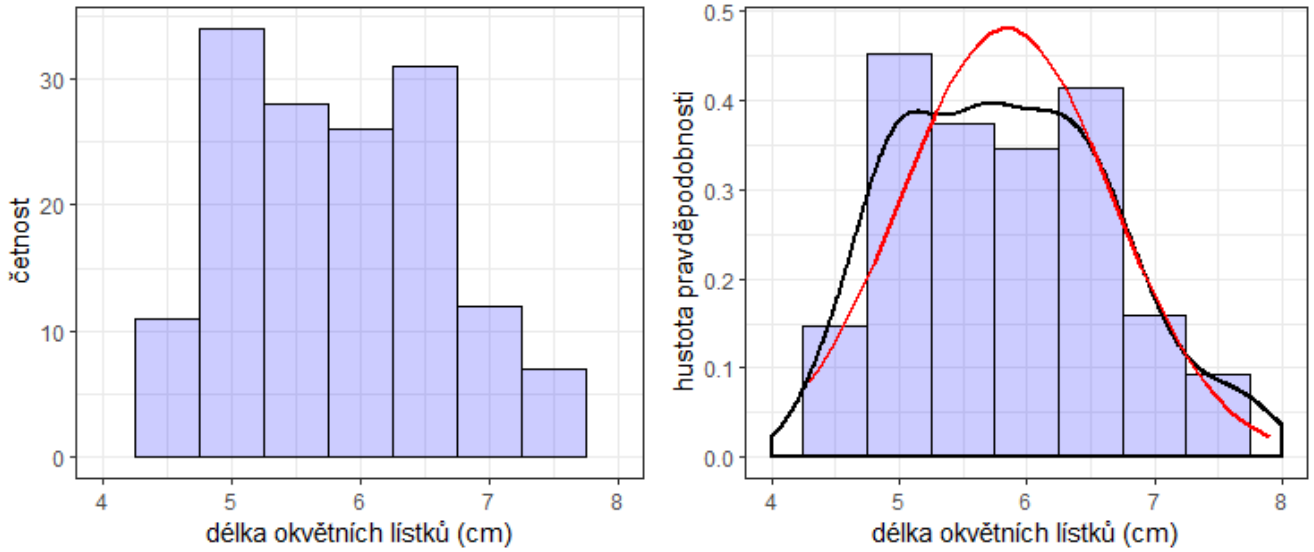
SNV má spojitou distribuční funkci a proto:
$\forall x \in \mathbb{R}: P(x) = 0$

Hustota pravděpodobnosti $f(x)$

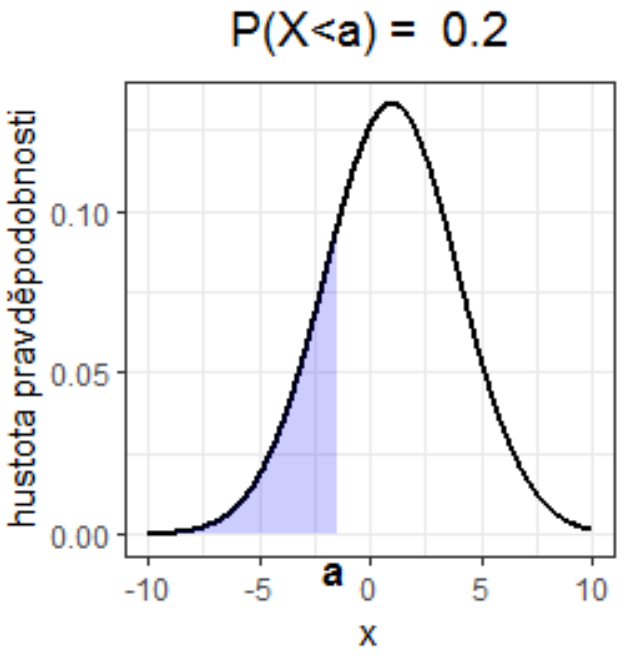
- Hustota pravděpodobnosti je funkce popisující rozdělení spojité náhodné veličiny.

$$F(t) = \int_{-\infty}^{t} f(x)\,dx \quad \Rightarrow \quad f(x) = \frac{dF}{dx}$$

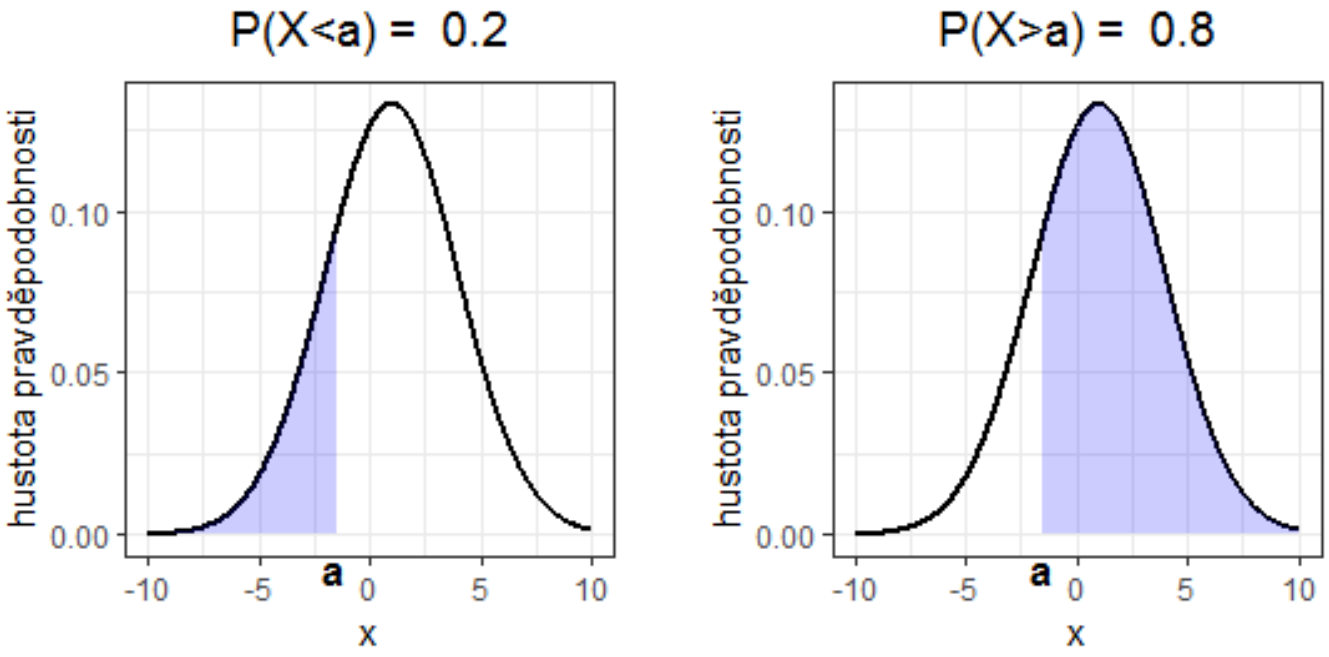
- **Vlastnosti hustoty pravděpodobnosti**
  - $f(x)$ je reálná nezáporná funkce,
  - $\int_{-\infty}^{\infty} f(x) = 1$
    (plocha pod křivkou hustoty je 1),
  - $\lim_{x \to -\infty} f(x) = 0$ („začíná v 0“),
  - $\lim_{x \to \infty} f(x) = 0$ („končí v 0“)



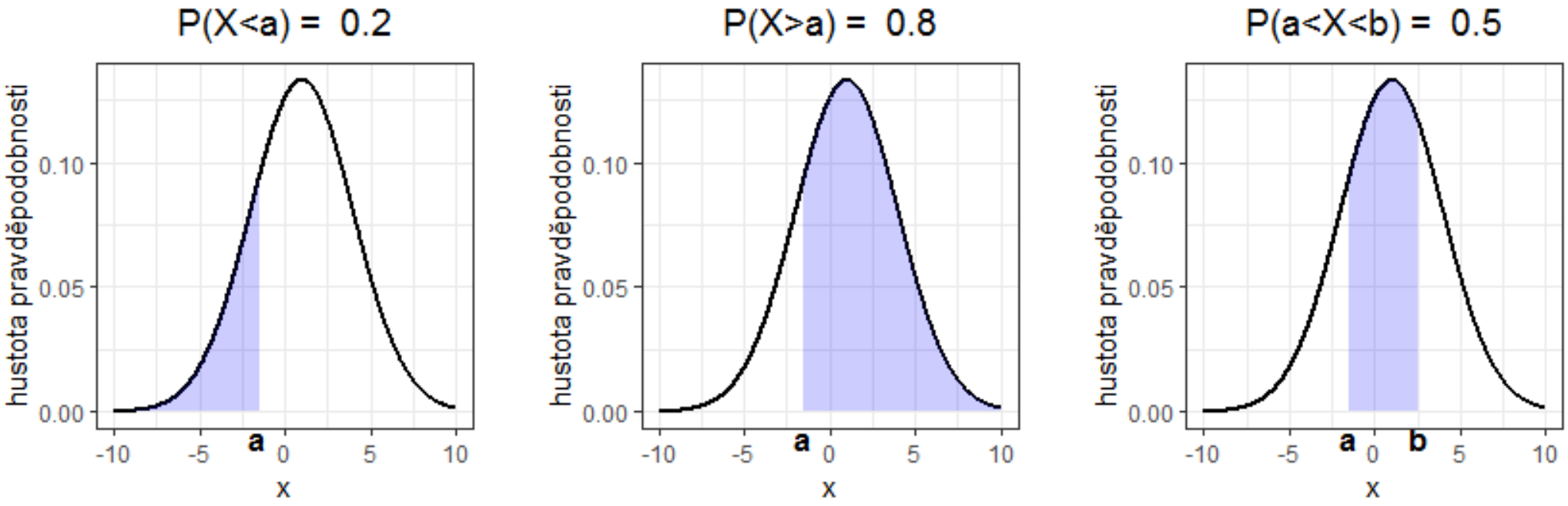
Vztah mezi pravděpodobností a hustotou p-stí



Vztah mezi pravděpodobností a hustotou p-stí

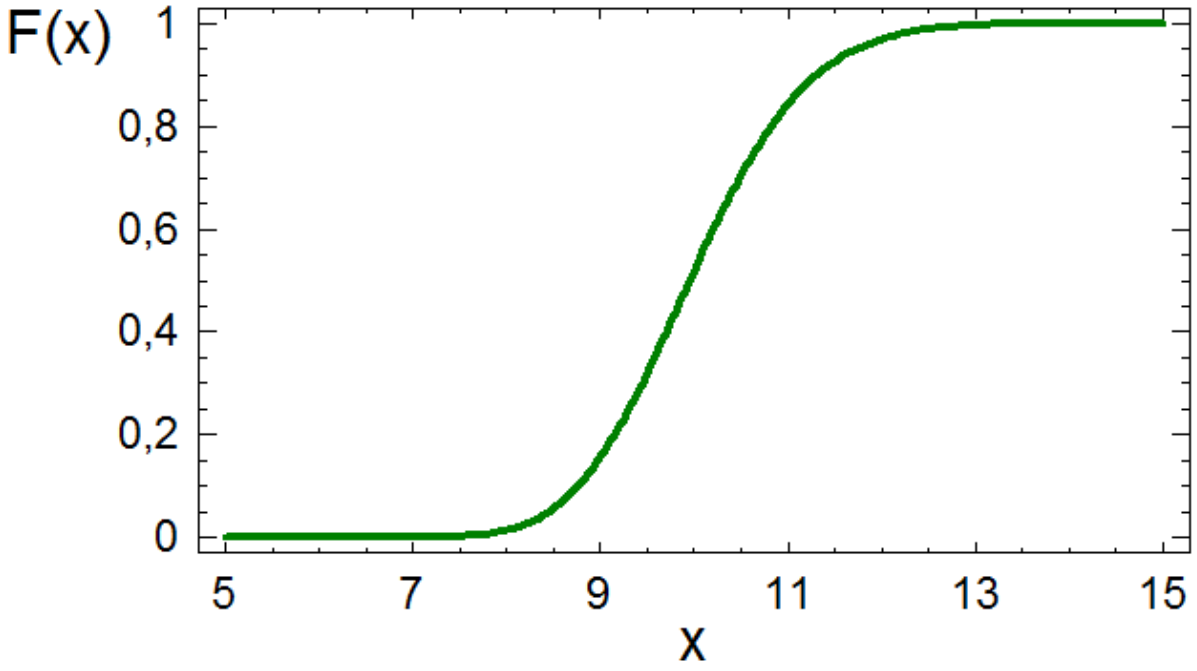


Vztah mezi pravděpodobností a hustotou p-stí

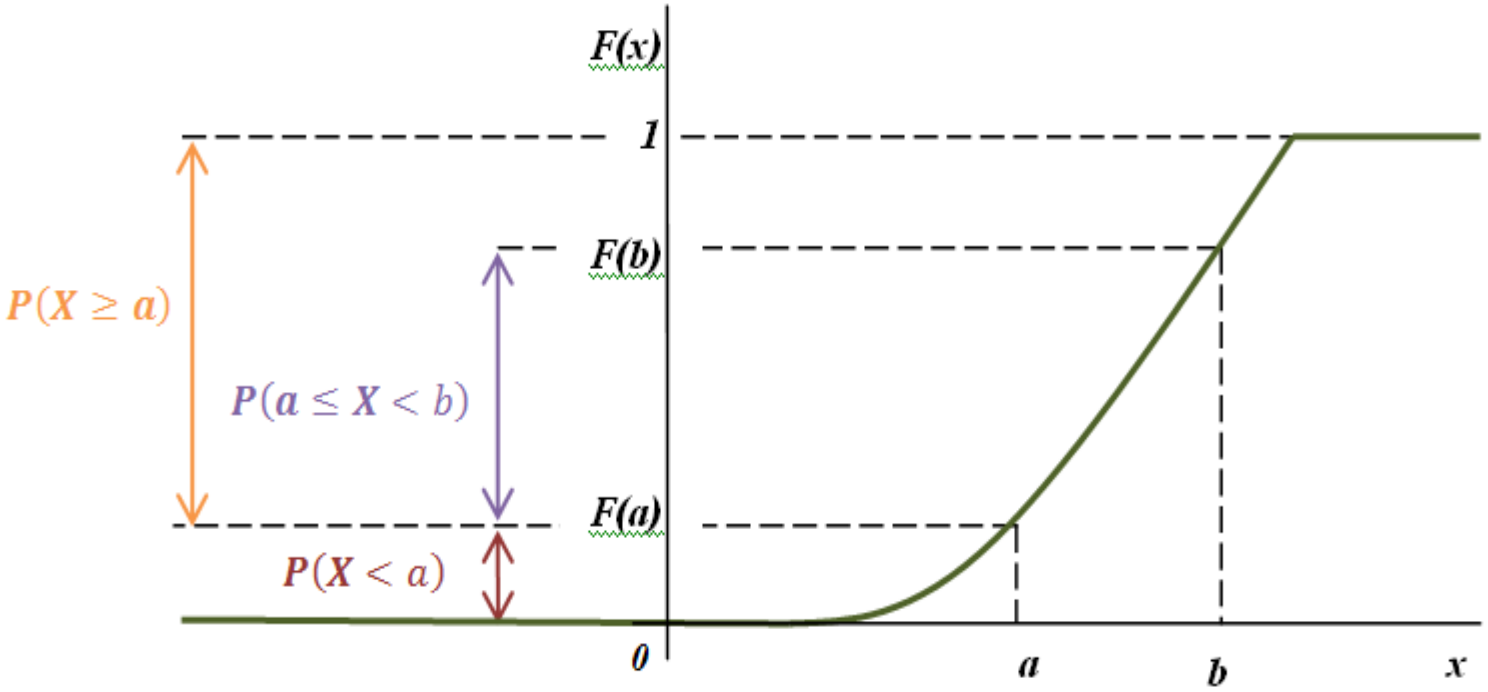


Distribuční funkce $F(x)$

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



Vztah mezi pravděpodobností a distribuční funkcí



$$P(X = a) = 0$$

Vztah mezi pravděpodobností, hustotou p-stí a distribuční funkcí

- $P(X < a) = F(a) = \int_{-\infty}^{a} f(x)\,dx$
- $P(X \geq a) = 1 - P(X < a) = 1 - F(a) = 1 - \int_{-\infty}^{a} f(x)\,dx = \int_{a}^{\infty} f(x)\,dx$
- $P(a \leq X < b) = P(X < b) - P(X < a) = F(a) - F(b) = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx = \int_{a}^{b} f(x)\,dx$
- $P(X = a) = \lim_{x \to a+} F(x) - F(a) = 0$

Je-li náhodná veličina X spojitá,
pak lze ve výše uvedených nerovnostech zaměňovat symboly ostré a neostré nerovnosti,
tj. $P(X < a) = P(X \leq a)$ apod.

Příklad 3

Tramvaje jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd tramvaje.

Rozdělení dané náhodné veličiny je dáno hustotou pravděpodobností, která je v intervalu 0-10 minut konstantní, mimo tento interval je nulová, tj.

$$f(x) = \begin{cases} c, & x \in \langle 0; 10 \rangle \\ 0, & x \notin \langle 0; 10 \rangle \end{cases}$$

a) Určete konstantu c.

b) Určete distribuční funkci $F(x)$.

c) Určete pravděpodobnost, že cestující bude čekat nejvýše 5 minut, alespoň 3 minuty, právě 7 minut

K čemu potřebujeme popisovat chování NV?

- NV převádějí abstraktní základní prostor Ω na čísla (s čísly se lépe pracuje).

- NV slouží jako model pro naše empirická pozorování (data).

  - V teorii pravděpodobnosti s nimi pracujeme teoreticky – jejich rozdělení považujeme za dané a zkoumáme jejich vlastnosti.

  - Ve statistice se snažíme cosi usoudit o jejich neznámém rozdělení na základě konkrétních realizací.

Proč potřebujeme číselné charakteristiky NV?

- Distribuční funkce (pravděpodobnostní funkce, hustota pravděpodobnosti) popisují rozdělení NV jednoznačně, do všech podrobností.

- Někdy nás zajímá pouze některý aspekt NV, který se dá popsat jedním číslem:
  - očekávaná hodnota NV,
  - variabilita možných hodnot,
  - hodnota, pod níž leží pouze určité množství hodnot NV,
  - šikmost rozdělení,
  - koncentrace hodnot NV kolem očekávané hodnoty (špičatost rozdělení).

Číselné charakteristiky NV

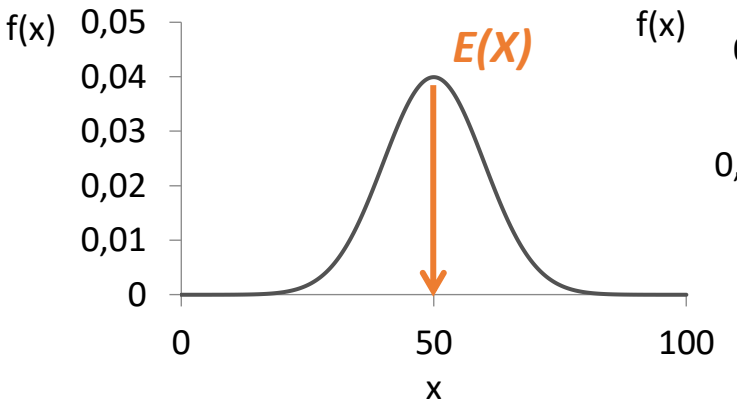
- Obecný moment r-tého řádu (značí se μ_r nebo $E(X^r)$ pro $r = 1, 2, \ldots$)
  - pro diskrétní NV: $\mu_r = \sum_{(i)} x_i^r \cdot P(x_i)$
  - pro spojitou NV: $\mu_r = \int_{-\infty}^{\infty} x^r \cdot f(x)\, dx$

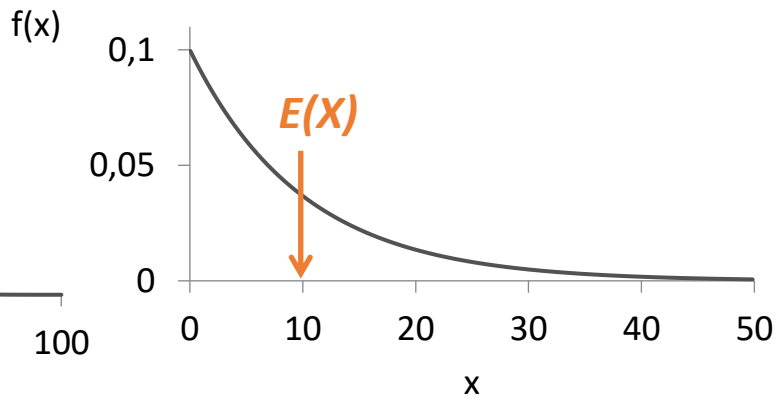
  - Střední hodnota (angl. expected value, mean; značí se $E(X)$nebo μ)
    - pro diskrétní NV: $\mathrm{E}(X) = \mu = \sum_{(i)} x_i \cdot P(x_i)$
    - pro spojitou NV: $\mathrm{E}(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

Význam střední hodnoty

Střední hodnotu $E(X)$ náhodné veličiny X lze chápat jako:

- průměrnou (očekávanou) hodnotu NV X, kolem níž hodnoty NV kolísají,
- míru polohy, populační průměr,
- vážený průměr všech možných hodnot $\left(\mathrm{E}(X) = \sum_{(i)} x_i \cdot P(x_i)\right)$,
- „těžiště“ možných hodnot.





Vlastnosti střední hodnoty

- $\forall a, b \in \mathbb{R}$: $E(aX + b) = aE(X) + b$,

- $E\left(\sum_i^n X_i\right) = \sum_i^n E(X_i)$,

  tj. střední hodnota součtu NV je rovna součtu jejich středních hodnot (vždy),

- $X_1, \cdots, X_n$ nezávislé $\Rightarrow$ $E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$,

  tj. jsou-li NV nezávislé, pak střední hodnota součinu NV je rovna součinu jejich středních hodnot (obecně to neplatí).

Kvantily

- **p-kvantil x_p** (také 100p%-ní kvantil) je číslo, pro které platí:

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

$$\Rightarrow F(x_p) = p \;\Rightarrow x_p = F^{-1}(p)$$

(tj. kvantilová funkce $F^{-1}(p)$ je funkcí inverzní k distribuční funkci $F(x_p)$)

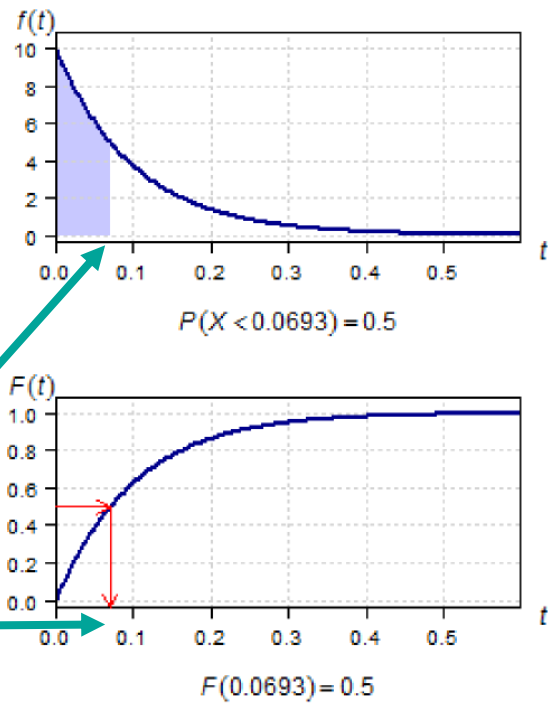
- Kvantily obvykle určujeme pouze pro SNV.

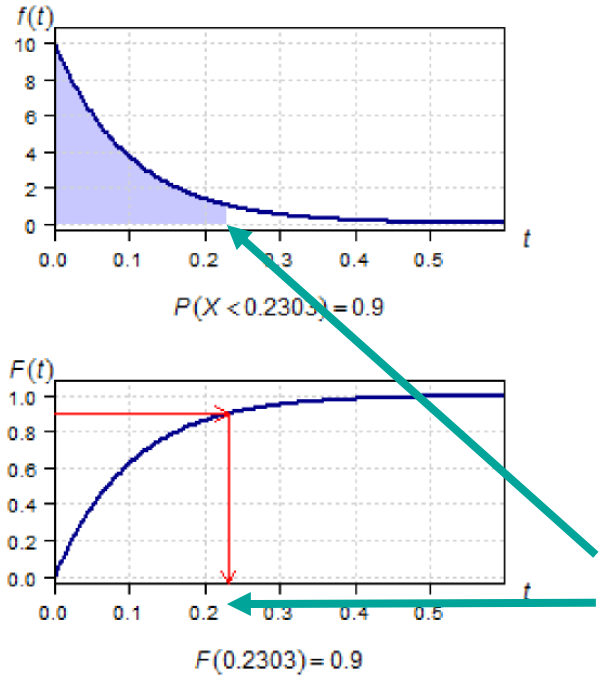
Kvantily

- **p-kvantil x_p** (také 100p%-ní kvantil) je číslo, pro které platí:

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

$$\Rightarrow F(x_p) = p \;\Rightarrow x_p = F^{-1}(p)$$





$x_{0,5}$
50% kvantil

$x_{0,9}$
90% kvantil

Význačné kvantily

Kvartily

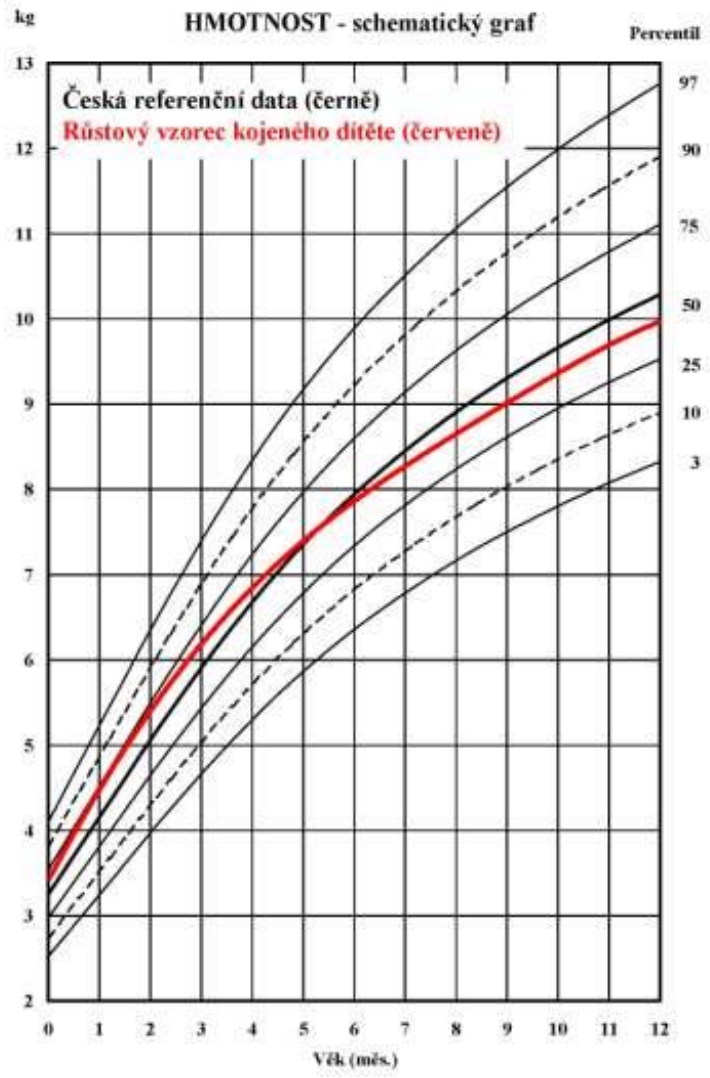
- Dolní kvartil $x_{0,25}$
- Medián $x_{0,5}$
- Horní kvartil $x_{0,75}$

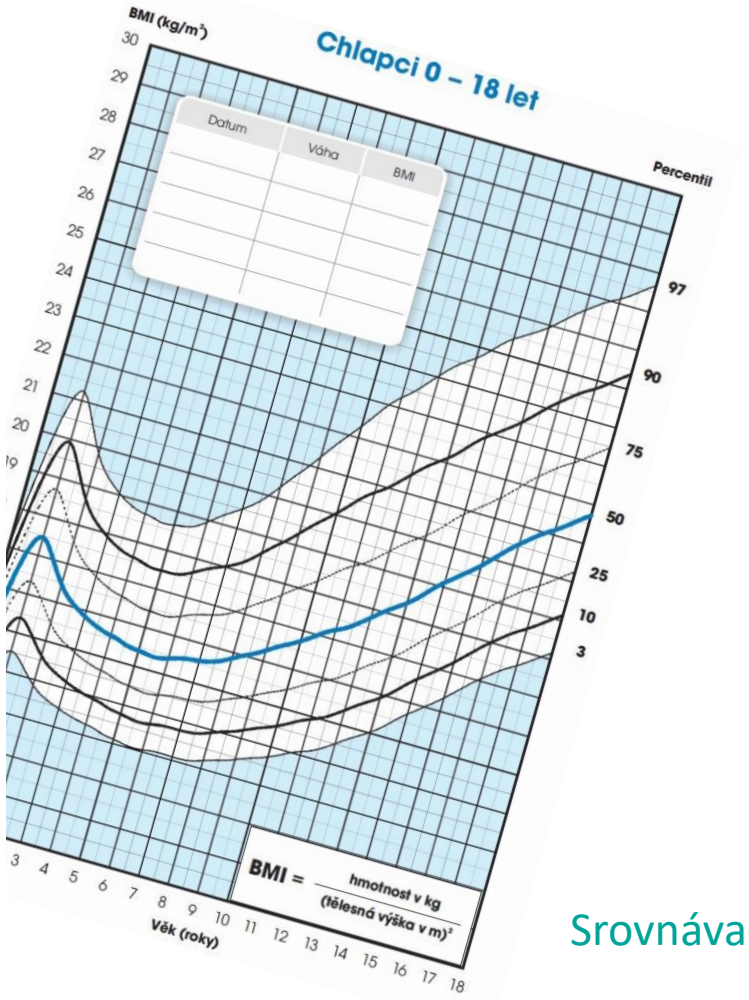
Decily – $x_{0,1}; x_{0,2}; \dots ; x_{0,9}$

Percentily – $x_{0,01}; x_{0,02}; \dots; x_{0,03}$

Minimum x_{min} a Maximum x_{max}

Kde se setkáme s kvantily?





Srovnávací testy

Modus

Modus $\hat{x}$ – typická hodnota náhodné veličiny

pro diskrétní NV: $\forall x_i$: $P(X = \hat{x}) \geq P(X = x_i)$

(tzn. modus je taková hodnota DNV, v níž $P(x_i)$ nabývá svého maxima)

pro spojitou NV: $\forall x$: $f(\hat{x}) \geq f(x)$

(tzn. modus je taková hodnota SNV, v níž $f(x)$ nabývá svého maxima)

- Modus není těmito podmínkami určen jednoznačně, tzn. náhodná veličina může mít několik modů (např. výsledek hodu kostkou).
- Má-li NV právě jeden modus, mluvíme o unimodálním rozdělení NV.
- Má-li NV unimodální symetrické rozdělení, pak $E(X) = x_{0,5} = \hat{x}$.

Příklad 4

Při ověřování kvality výroby jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličino X. Z dlouhodobého pozorování jsou známy údaje uvedené v následující tabulce.

Počet vadných výrobků	Pravděpodobnost
0	0,25
1	0,50
2	0,25

Určete střední hodnotu a modus počtu vadných výrobků v testovaném vzorku.

Příklad 5

Tramvaje jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd tramvaje.

Rozdělení dané náhodné veličiny je dáno hustotou pravděpodobností, která je v intervalu 0-10 minut konstantní, mimo tento interval je nulová, tj.

$$f(x) = \begin{cases} 0{,}1, & x \in \langle 0; 10 \rangle \\ 0, & x \notin \langle 0; 10 \rangle \end{cases}$$

Určete střední hodnotu náhodné veličiny X.

Číselné charakteristiky NV

- Centrální moment r-tého řádu (značí se μ'_r nebo $E\left(\left(X - E(X)\right)^r\right)$ pro $r = 1, 2, \ldots$)
  - pro diskrétní NV: $\mu'_r = \sum_{(i)}\left(x_i - E(X)\right)^r \cdot P(x_i)$
  - pro spojitou NV: $\mu'_r = \int_{-\infty}^{\infty}\left(x - E(X)\right)^r \cdot f(x)\, dx$

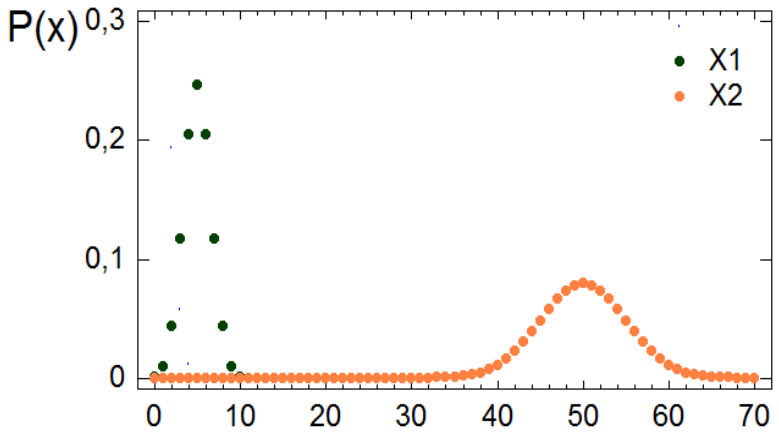
  - Rozptyl (angl. dispersion, resp. variance, značí se $\mu_2{}'$ nebo $D(X)$ nebo σ^2)
    - pro diskrétní NV: $D(X) = \sigma^2 = \sum_{(i)}(x_i - EX)^2 \cdot P(x_i)$
    - pro spojitou NV: $D(X) = \sigma^2 = \int_{-\infty}^{\infty}(x - EX)^2 \cdot f(x)\, dx$

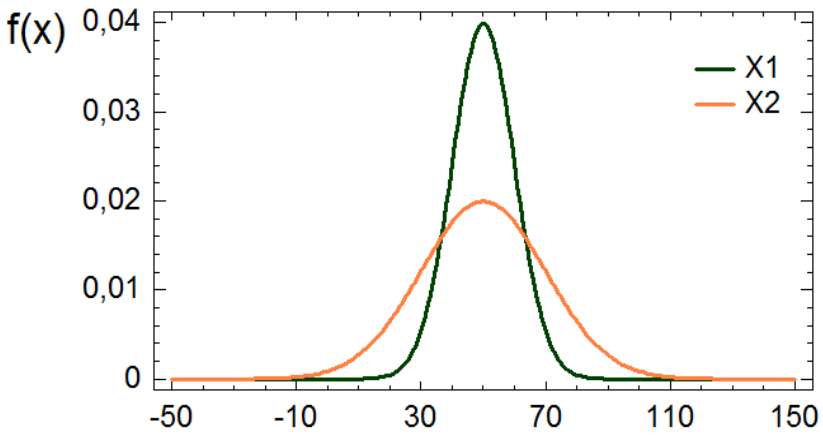
Výpočetní vztah pro rozptyl:

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

Význam rozptylu

- Míra variability dat kolem střední hodnoty.
- Střední kvadratická odchylka od střední hodnoty $\left(D(X) = E(X - E(X))^2\right)$.
- Malý rozptyl ≈ hodnoty NV se s vysokou pravděpodobností objevují blízko $E(X)$.
- Velký rozptyl ≈ hodnoty NV se často objevují ve velké vzdálenosti od $E(X)$.





$$\boldsymbol{D(X1) < D(X2)}$$

Jednotka rozptylu je kvadrátem jednotky náhodné veličiny!!!

Vlastnosti rozptylu

- $\forall a, b \in \mathbb{R}$: $\quad D(aX + b) = a^2 D(X)$,

- $X_1, \cdots, X_n$ nezávislé $\Rightarrow D\left(\sum_i^n X_i\right) = \sum_i^n D(X_i)$,

  tj. jsou-li NV nezávislé, pak střední hodnota součtu NV je rovna součtu jejich středních hodnot (obecně to neplatí).

Směrodatná odchylka

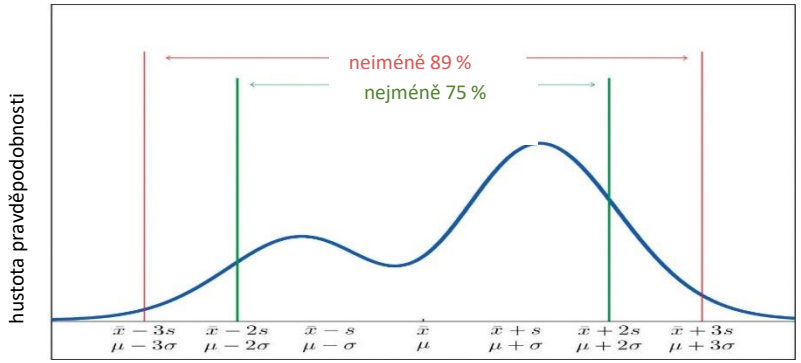
Směrodatná odchylka σ

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

Jakou představu o náhodné veličině X si lze udělat
na základě její stř. hodnoty μ a sm. odchylky σ?

$$\forall k > 0: P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}$$ (Čebyševova nerovnost)

$\boldsymbol{k}$	$P(\mu - k\sigma < X < \mu + k\sigma)$
1	>0
2	>0,75
3	>0,89



Směrodatná odchylka neumožňuje srovnávat variabilitu náhodných veličin měřených v různých jednotkách!

Variační koeficient

Variační koeficient $\gamma(X)$ definujme pouze pro nezáporné náhodné veličiny.

$$\boxed{\gamma = \frac{\sigma}{\mu}, \text{ resp. } \frac{\sigma}{\mu} \cdot 100\ (\%)}$$

- Variační koeficient – směrodatná odchylka v procentech střední hodnoty
- Čím nižší variační koeficient, tím homogennější soubor.
- $\gamma(X) > \ 50\%$ $\approx$ silně rozptýlený soubor.

Příklad 6

Při ověřování kvality výroby jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličino X. Z dlouhodobého pozorování jsou známy údaje uvedené v následující tabulce.

Počet vadných výrobků	Pravděpodobnost
0	0,25
1	0,50
2	0,25

Určete rozptyl, směrodatnou odchylku a variační koeficient počtu vadných výrobků v testovaném vzorku.

Příklad 7

Tramvaje jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd tramvaje.

Rozdělení dané náhodné veličiny je dáno hustotou pravděpodobností, která je v intervalu 0-10 minut konstantní, mimo tento interval je nulová, tj.

$$f(x) = \begin{cases} 0{,}1, & x \in \langle 0; 10 \rangle \\ 0, & x \notin \langle 0; 10 \rangle \end{cases}$$

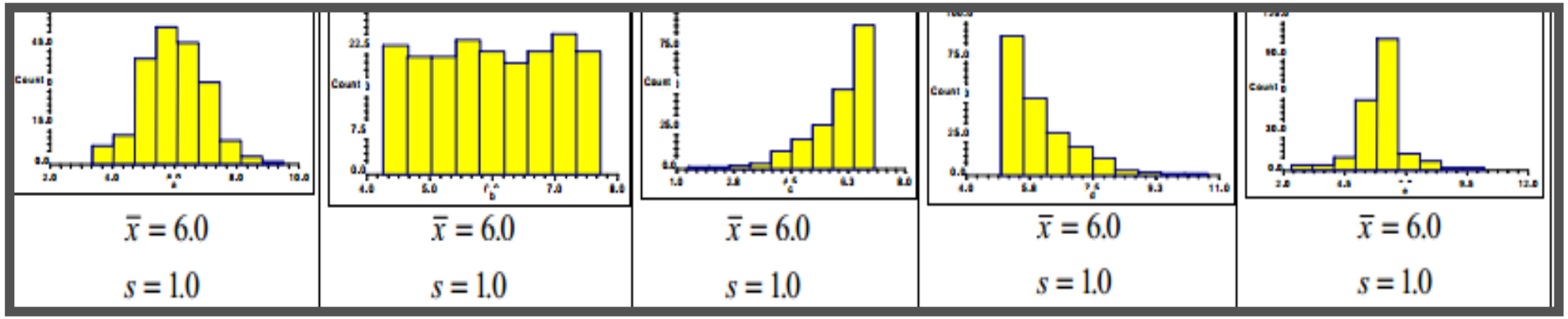
Určete rozptyl, směrodatnou odchylku a variační koeficient náhodné veličiny X.

Příklad 8

Modelujeme výšku chlapců ve věku 3,5 – 4 roky. Vysvětlete:

1) 2% kvantil modelované náhodné veličiny je 93 cm.
2) Střední hodnota modelované NV je 102 cm, směrodatná odchylka je 4,5 cm. V jakém rozpětí lze očekávat výšku chlapců ve věku 3,5 – 4 roky? Pro interpretaci využijte Čebyševovu nerovnost.
3) Střední hodnota modelované NV je 102 cm, směrodatná odchylka je 4,5 cm. Posuďte variabilitu modelované NV. (Není příliš vysoká? Pro posouzení použijte variační koeficient.)
4) Víme, že pro distribuční funkci modelované NV platí: $F(111) = 0{,}98$. Co jsme se dozvěděli?

Jsou míry polohy a míry variability dostatečné pro posouzení rozdělení sledovaných veličin?



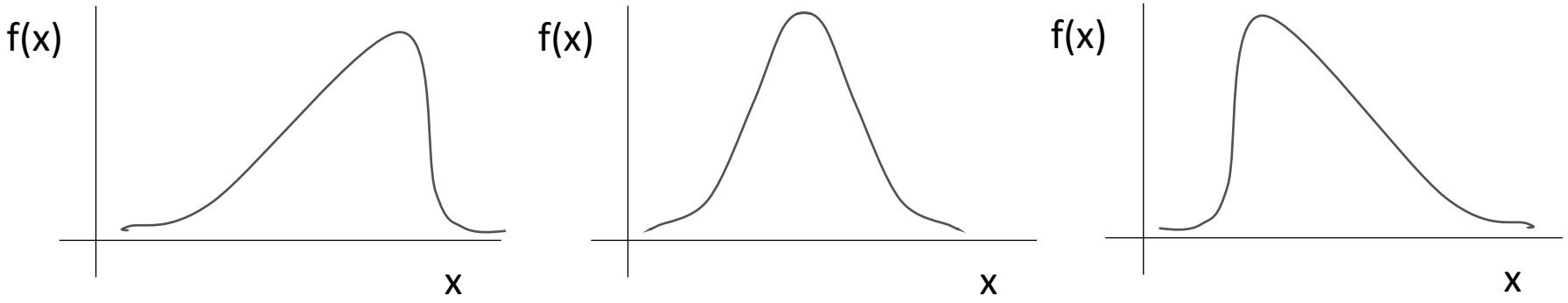
Zdroj: TVRDÍK, J.: Základy matematické statistiky, Ostravská univerzita, 2008

Všech pět ukázek má stejné charakteristiky polohy i variability (průměry i směrodatné odchylky jsou shodné).
Přesto na první pohled vidíme, že tvary rozdělení dat jsou různé.

Šikmost

Šikmost α_3 - míra symetrie rozdělení NV

$$\boxed{\alpha_3 = \frac{\mu_3{}'}{\sigma^3} = \frac{E(X - E(X))^3}{\sigma^3}}$$



$\alpha_3 < 0$
negativně zešikmené rozdělení

$\alpha_3 = 0$
symetrické rozdělení

$\alpha_3 > 0$
pozitivně zešikmené rozdělení

$E(X) < x_{0,5} < \hat{x}$

$\hat{x} = x_{0,5} = E(X)$

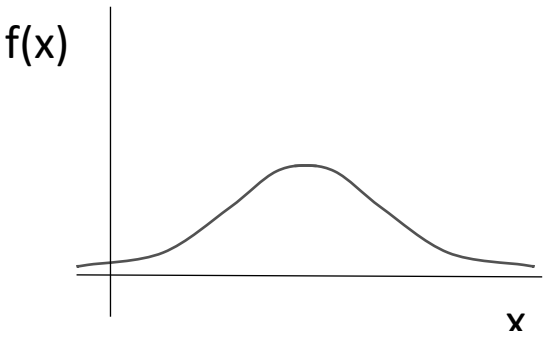
$\hat{x} < x_{0,5} < E(X)$

obvykle

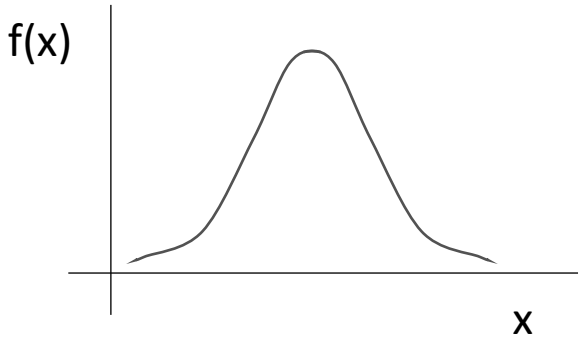
Špičatost

Špičatost α_4 - míra koncentrace kolem průměru

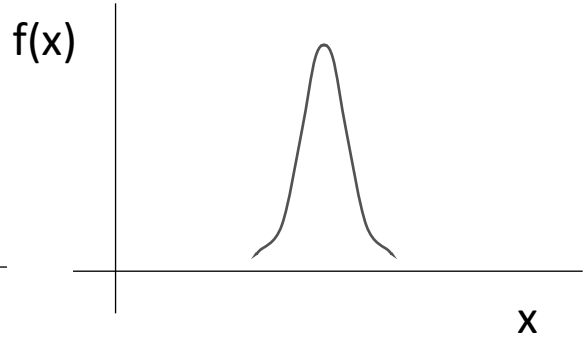
$$\boxed{\alpha_4 = \frac{\mu_4{}'}{\sigma^4} = \frac{E(X - E(X))^4}{\sigma^4}}$$



$\alpha_4 < 3$
špičatost menší
než u norm. rozdělení
(plošší rozdělení)



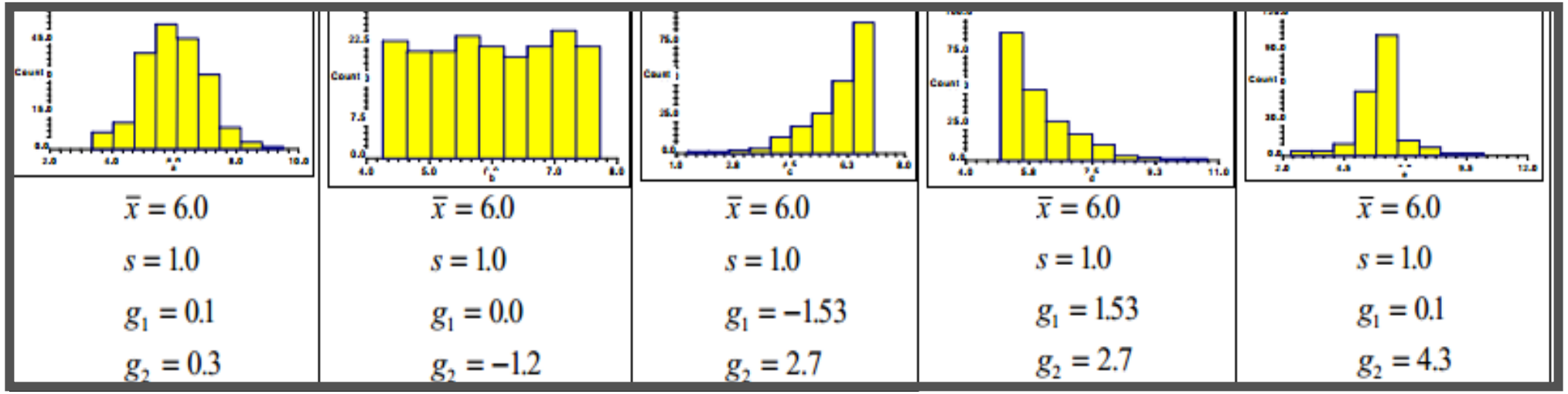
$\alpha_4 = 3$
špičatost odpovídající
normálnímu rozdělení



$\alpha_4 > 3$
špičatost větší
než u norm. rozdělení
(špičatější rozdělení)

- standardizovaná špičatost $\alpha_4 - 3$

Jsou míry polohy a míry variability dostatečné pro posouzení rozdělení sledovaných veličin?



Zdroj: TVRDÍK, J.: Základy matematické statistiky, Ostravská univerzita, 2008

Všech pět ukázek má stejné charakteristiky polohy i variability (průměry i směrodatné odchylky jsou shodné).
Přesto na první pohled vidíme, že tvary rozdělení dat jsou různé.
K číselnému vyjádření těchto rozdílů nám slouží další charakteristiky - šikmost (g_1, angl. skewness) a špičatost (g_2, angl. kurtosis).

Transformace náhodné veličiny

- Transformací NV *X* rozumíme aplikaci prosté reálné funkce *g* tak, že vznikne nová náhodná veličina
$$Y = g(X).$$

- NV *Y* může nabývat jiných hodnot než NV *X*,

  může mít také jiné rozdělení pravděpodobnosti, které chceme najít,

  tj. hledáme pravděpodobnostní f-ci, resp. hustotu pravděpodobnosti NV *Y*.

Transformace náhodné veličiny

$$Y = g(X)$$

$$\forall y \in \mathbb{R}: \quad F_Y(y) = P(Y < y) = P(g(X) < y)$$

Pro diskrétní NV:

$$P_Y(y) = P(Y = y) = P(g(X) = y) = P\big(X \in g^{-1}(y)\big) = \sum_{x \in g^{-1}(y)} P_X(x)$$

Pro spojitou NV:

- Je-li g rostoucí f-ce: $\forall y \in \mathbb{R}: \quad F_Y(y) = P\big(X < g^{-1}(y)\big) = F_X(g^{-1}(y))$
- Je-li g klesající f-ce: $\forall y \in \mathbb{R}: \quad F_Y(y) = P\big(X > g^{-1}(y)\big) = 1 - F_X(g^{-1}(y))$

- Je-li g spojitě diferencovatelná f-ce: $f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$

Děkuji za pozornost!

martina.litschmannova@vsb.cz

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY