

Šance

- Šance (odds) $O(A) = 1:4$ odpovídá $P(A) = 1/5$

$$O(A) = \frac{P(A)}{P(\neg A)} = \frac{P(A)}{1 - P(A)}$$

$$P(A) = \frac{O(A)}{1 + O(A)}$$

Podíl šancí

- *OR* (odds ratio)
- Podíl šance, že se vyskytne jev *A* za podmínky jevu *B* k šanci, že se jev *A* vyskytne, když *B*

neplatí

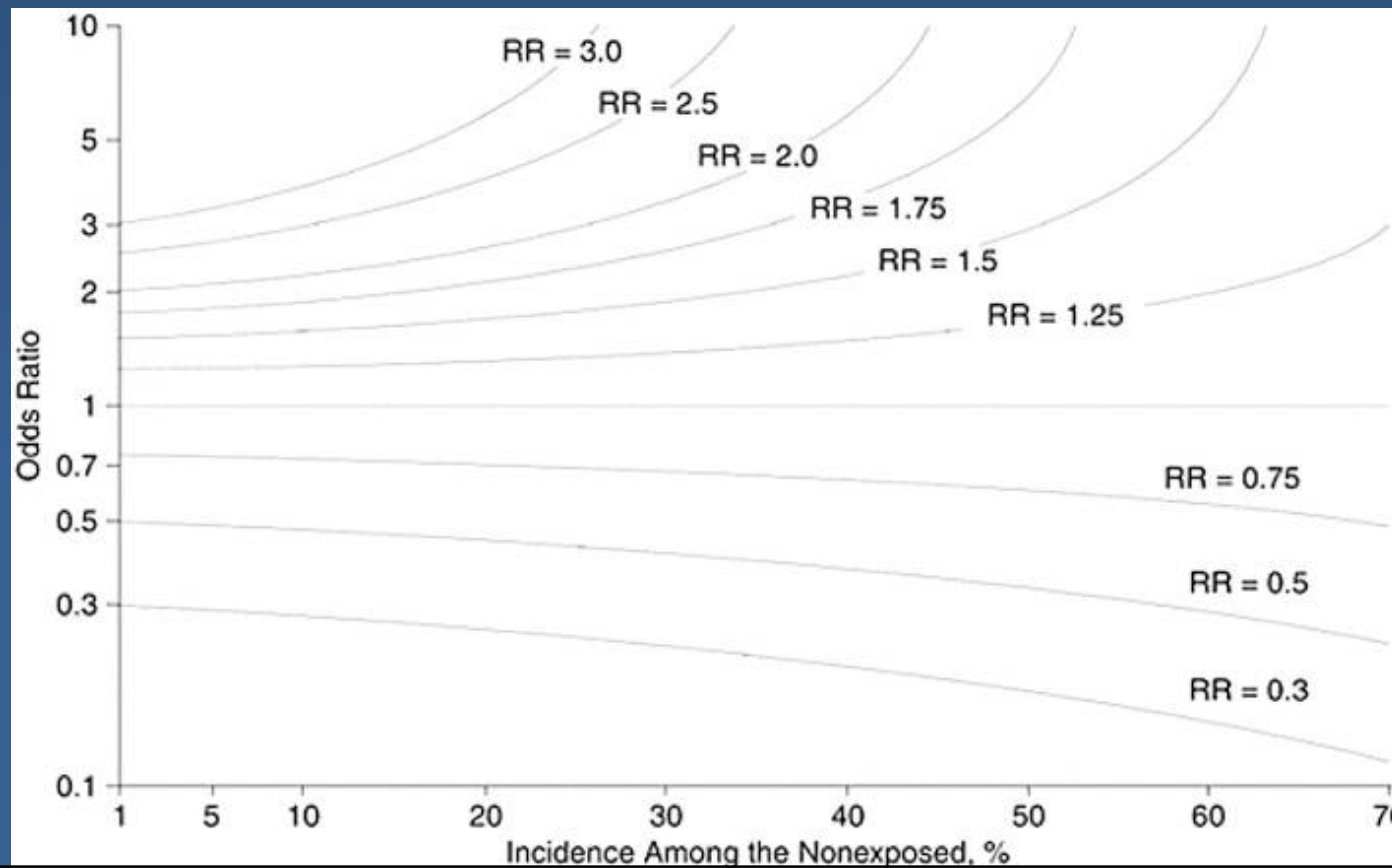
$$OR = \frac{O(A|B)}{O(A|\neg B)}$$

- Šance na výskyt rakoviny plic (jev *A*) u kuřáků (jev *B*) je 5/4, u nekuřáků (jev $\neg B$) 1/8

$$OR = \frac{\frac{5}{4}}{\frac{1}{8}} = \frac{40}{4} = 10$$

Relativní riziko (RR) vs. podíl šancí (OR)

RR	OR
Snadná interpretace rizik (% událostí)	Méně intuitivní interpretace
Určitá matematická omezení	Výhodné matematické vlastnosti
Závislé na bazálním riziku	Nezávislé na bazálním riziku (srovnání studií s různým bazálním rizikem)



Pro malé bazální riziko (< 10%) jsou RR a OR v podstatě srovnatelné

Výpočet

Srovnání výskytu událostí mezi rameny A a B



S událostí
(výskyt
onemocnění)



Bez události
(zdraví)

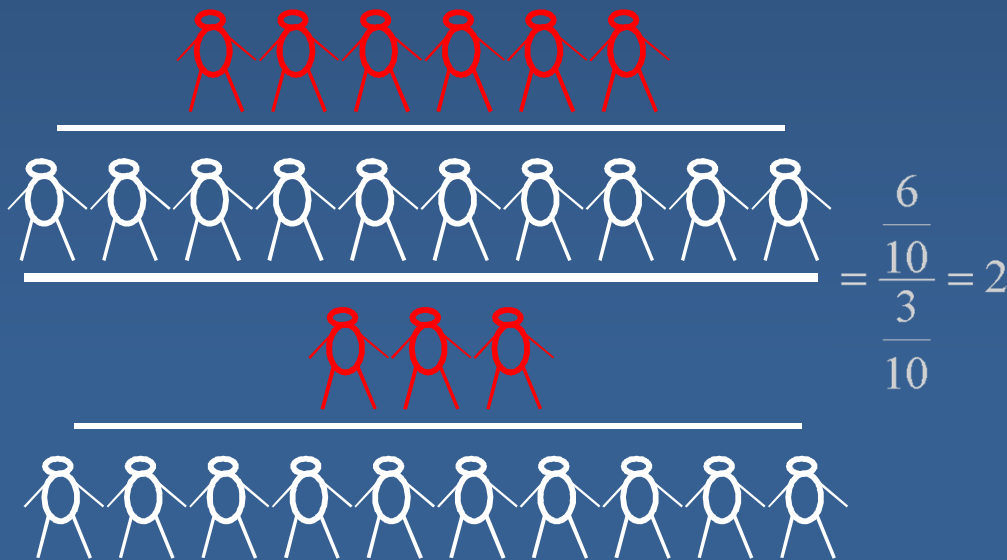
Rameno A



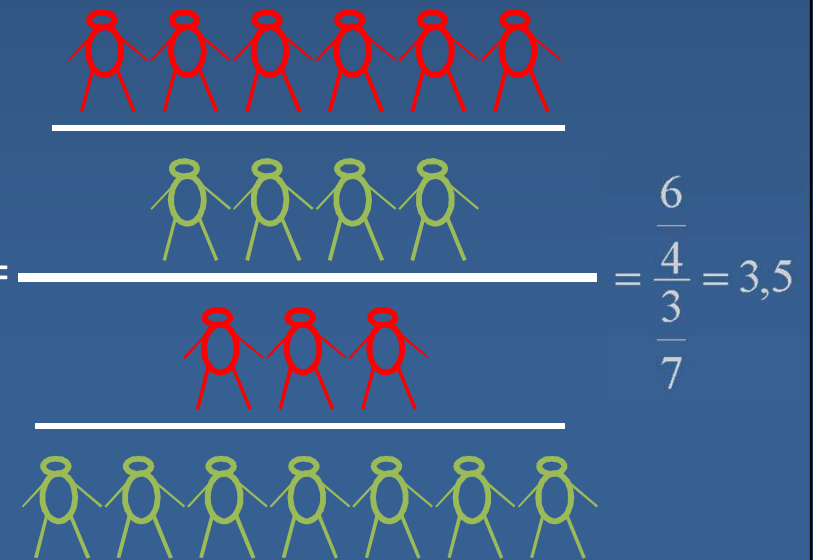
Rameno B



RR =



OR =



Věrohodnostní poměr

- LR (likelihood ratio)
- Podíl psti, že se vyskytne jev A za podmínky jevu B k psti, že se jev A vyskytne, když B neplatí

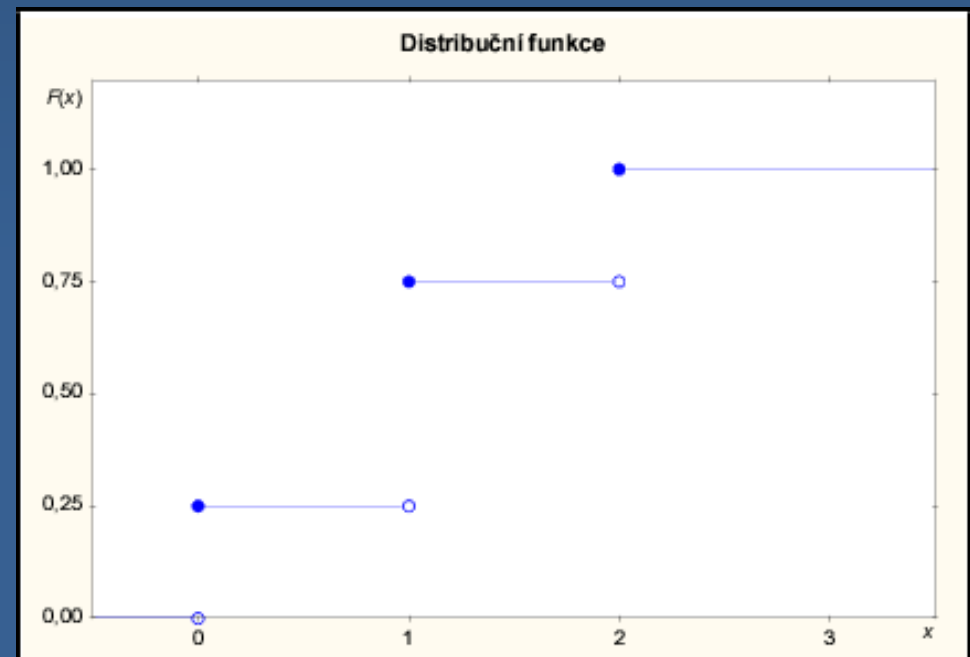
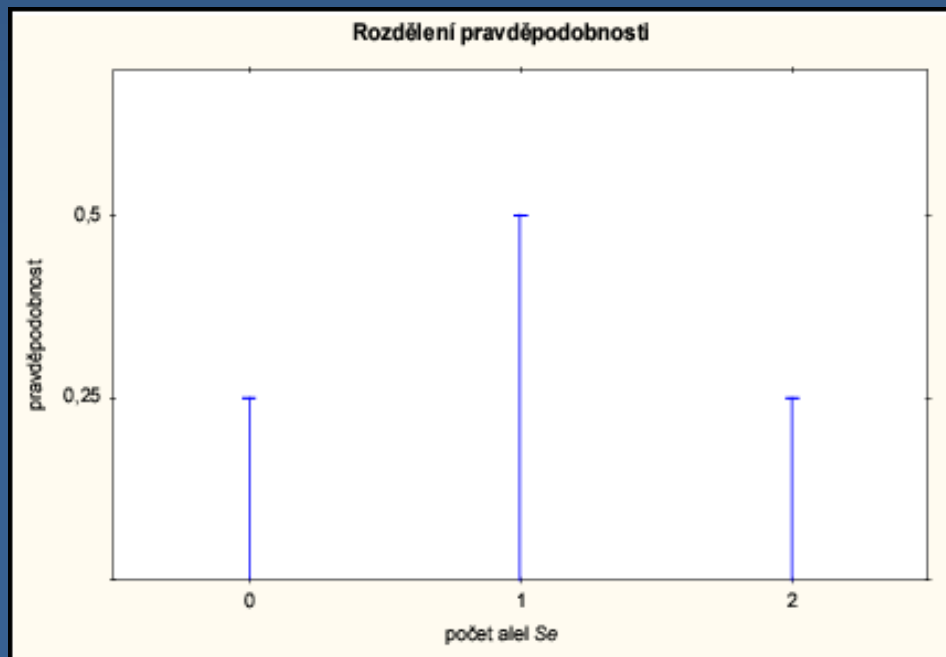
$$LR = \frac{P(A|B)}{P(A|\neg B)}$$

6. Náhodná veličina

- Realizace náhodného jevu
- Jak často určité hodnoty náhodné veličiny nastávají, je popsáno rozdělením psti

Diskrétní náhodná veličina

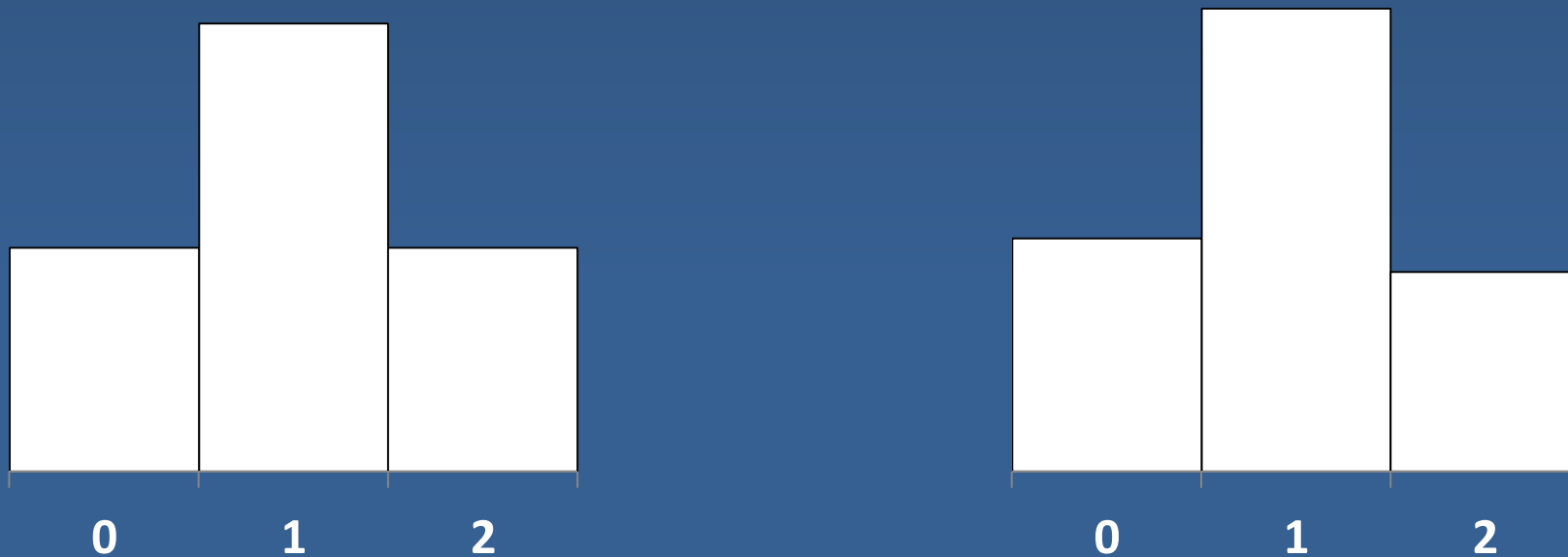
- Může nebývat jen určitých hodnot x_i , každé hodnotě x_i je přiřazena pst $P(X = x_i) > 0$
- Součet $P(X = x_i)$ je roven 1
- Psti $P(X = x_i)$ charakterizují diskrétní rozdělení psti
- Distribuční fce $F(x) = P(X \leq x)$, pro $-\infty < x < +\infty$



Model -> realita

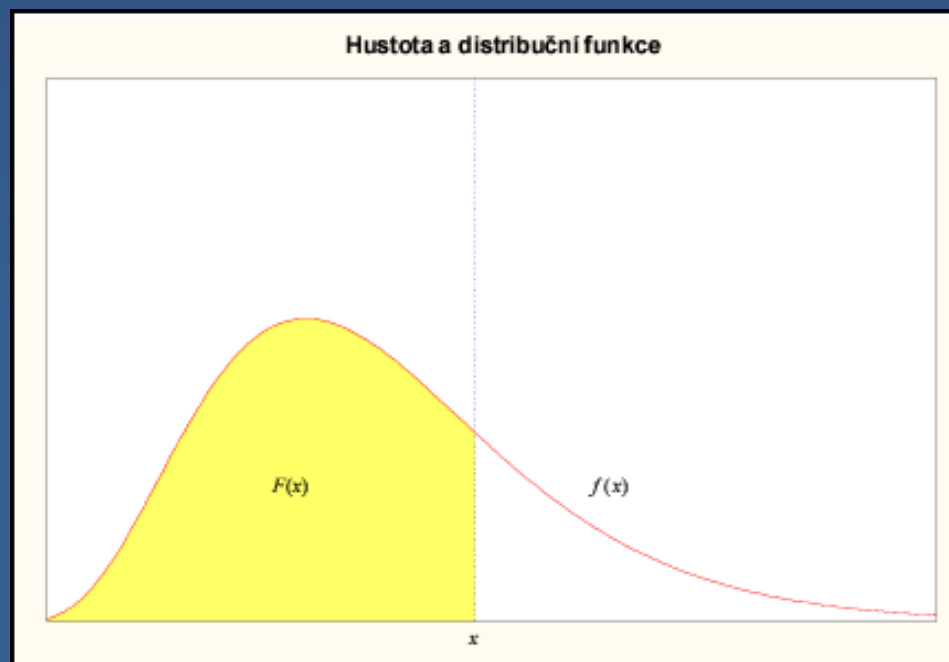
- 1000xopakovaný hod 2 mincemi

$p(x)$	x	n_x	n_x/n
0,25	0	260	0,260
0,5	1	517	0,517
0,25	2	223	0,223



Spojité náhodná veličina

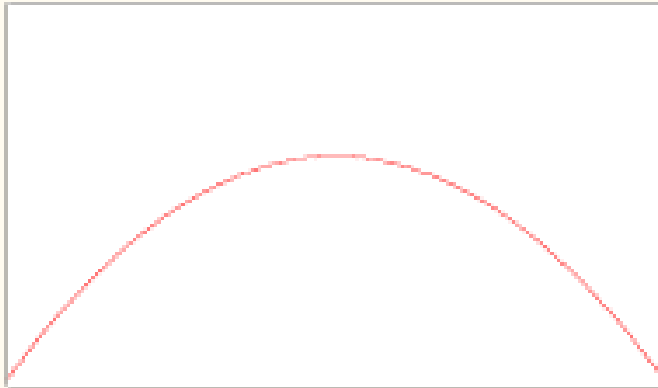
- X nabývá hodnot x z určitého intervalu
- Reálná nezáporná fce $f(x)$ – hustota (popisuje pstní rozdělení)
- Distribuční funkce $F(x)$ je plocha pod hustotou



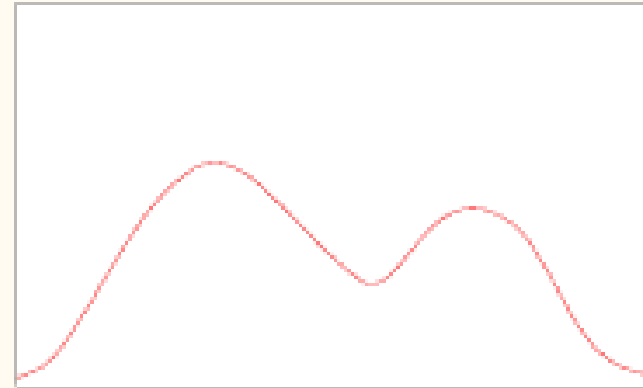
Tvar rozložení

Symetrické jednovrcholové

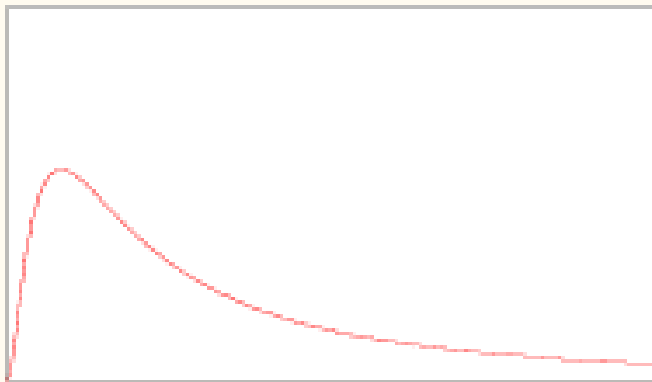
Dvouvrcholové



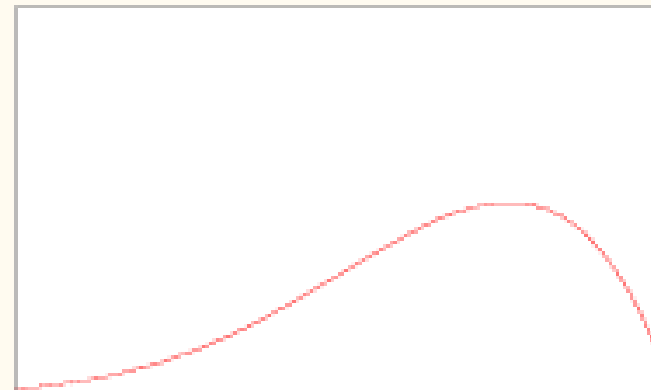
a)



b)



c)



d)

Pravostranně asymetrické

Levostranně asymetrické

Alternativní rozložení

- Jev může nabývat jednoho ze dvou stavů – 0 nebo 1

$$P(X = 1) = p, P(X = 0) = 1 - p$$

- $X \sim A(p)$

Binomické rozložení

- Diskrétní
- n nezávislých pokusů
- Úspěch x neúspěch
- Pst úspěchu π , pst neúspěchu $1 - \pi$ a jsou v každém pokusu stejné
- Celkový počet úspěchů X v n nezávislých pokusech
- Nabývá celočíselných hodnot od 0 do n

Binomické rozložení

- Pst, že v n nezávislých pokusech nastane právě k úspěchů

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \text{ pro } k = 0, 1, 2, \dots, n$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

počet k -členných kombinací z

n objektů

$$n! = n(n-1)(n-2)\dots 1$$

- $X \sim Bi(n, \pi)$

Binomické rozložení - příklad

- s jakou pravděpodobností neudělá 12 z 50 stejně připravených studentů zkoušku, když je pst neúspěchu 0,2
- $bi(50,0,2)$

$$P(X = 12) = \binom{50}{12} \cdot 0,2^{12} \cdot 0,8^{38} = 0,103$$

Poissonovo rozložení

- Řídké jevy

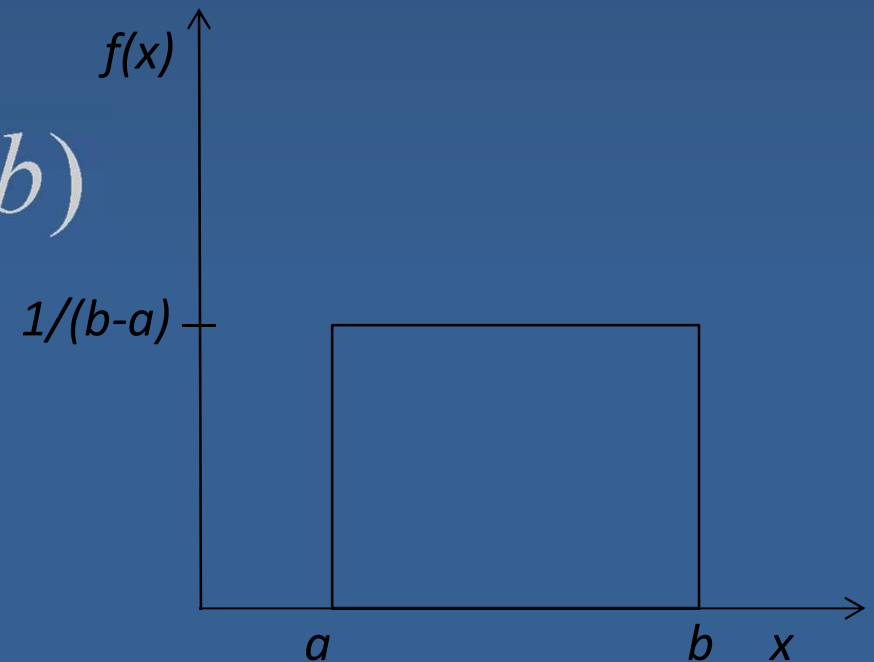
$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

- Kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu...
- $X \sim Po(\lambda), \lambda > 0$

Rovnoměrné rozložení

- Na intervalu (a,b) , kde $-\infty < a < b < \infty$
- Ve všech bodech daného intervalu má konstantní hustotu pravděpodobnosti

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a,b) \\ 0 & \text{pro } x \notin (a,b) \end{cases}$$



Normální rozložení

- Tělesná výška, diastolický krevní tlak, vitální kapacita plic,...
- Gaussovo rozdělení

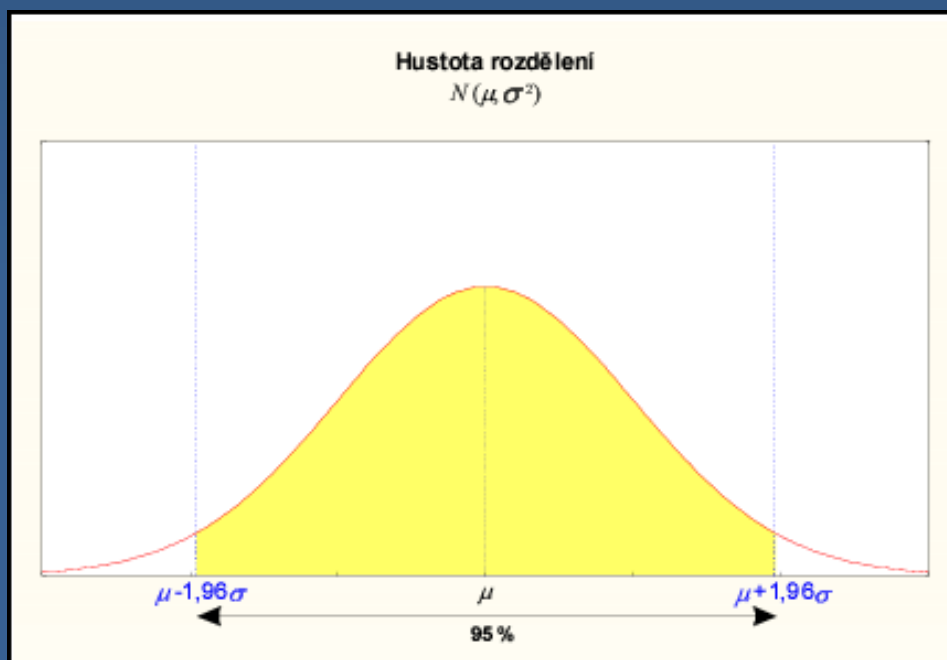
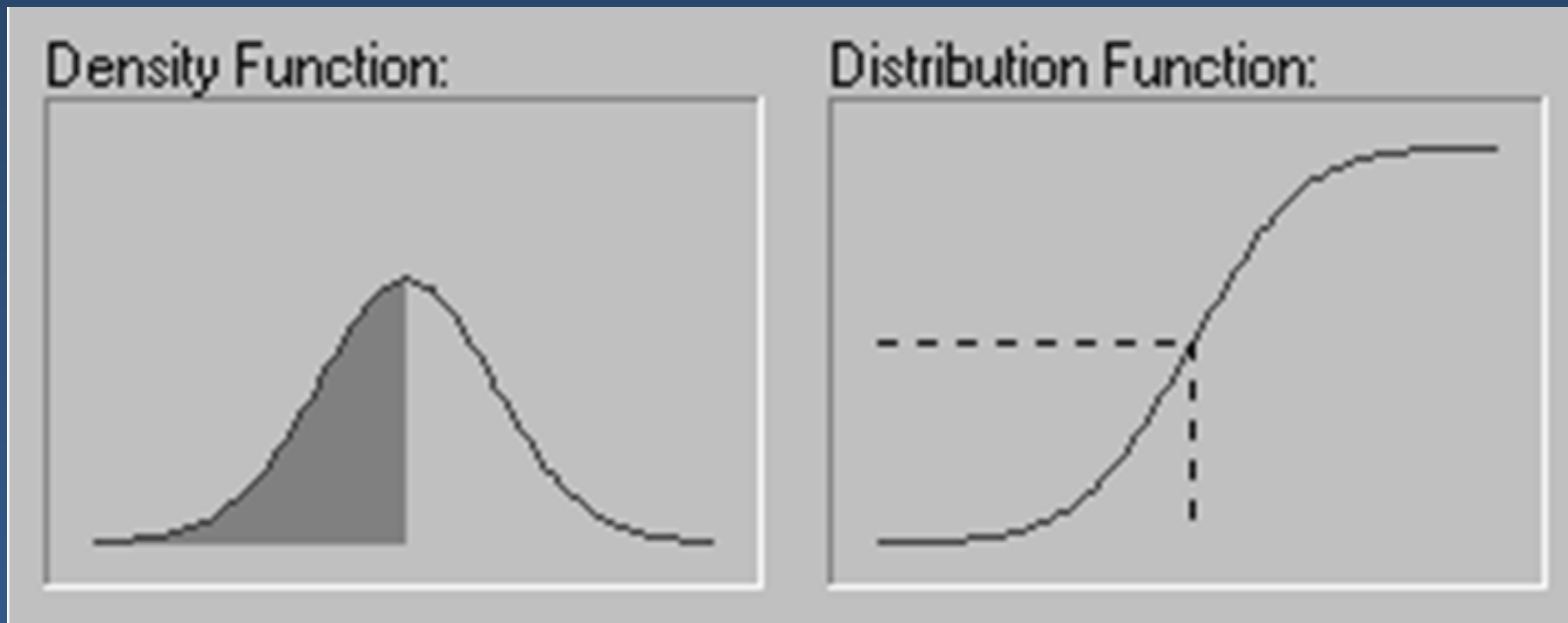
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\pi = 3,14$ a $e = 2,72$ (matematické konstanty)

$\mu, \sigma > 0$ parametry určující polohu křivky na ose x a její „roztažení“ podél osy x

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

Normální rozložení



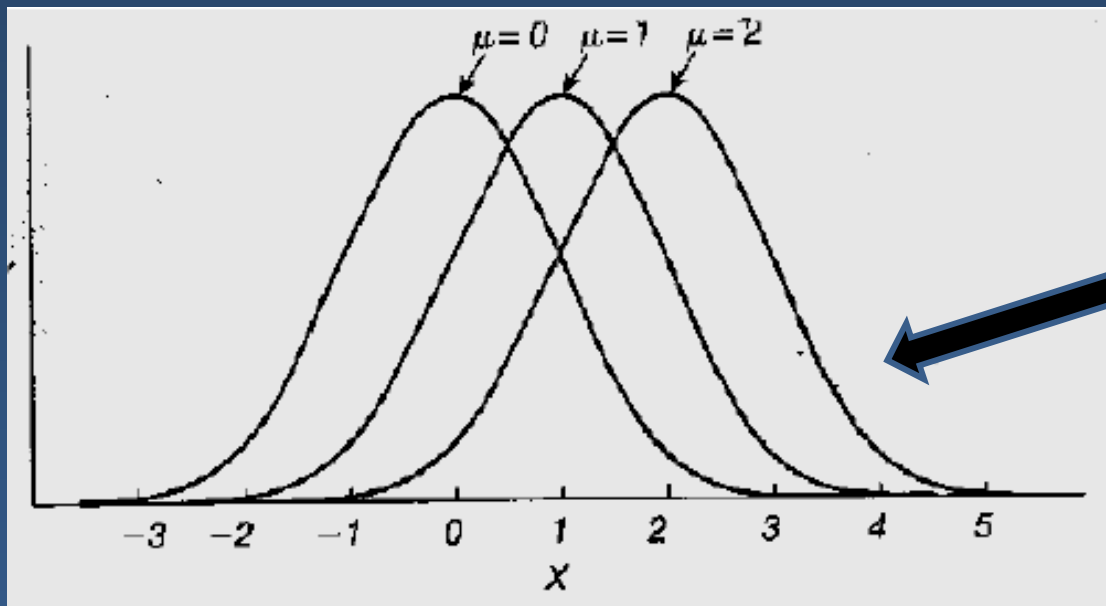
Pst, že náhodná veličina nabude hodnot z určitého intervalu = plocha pod hustotou nad tímto intervalem

68,27% leží mezi $\mu \pm 1\sigma$

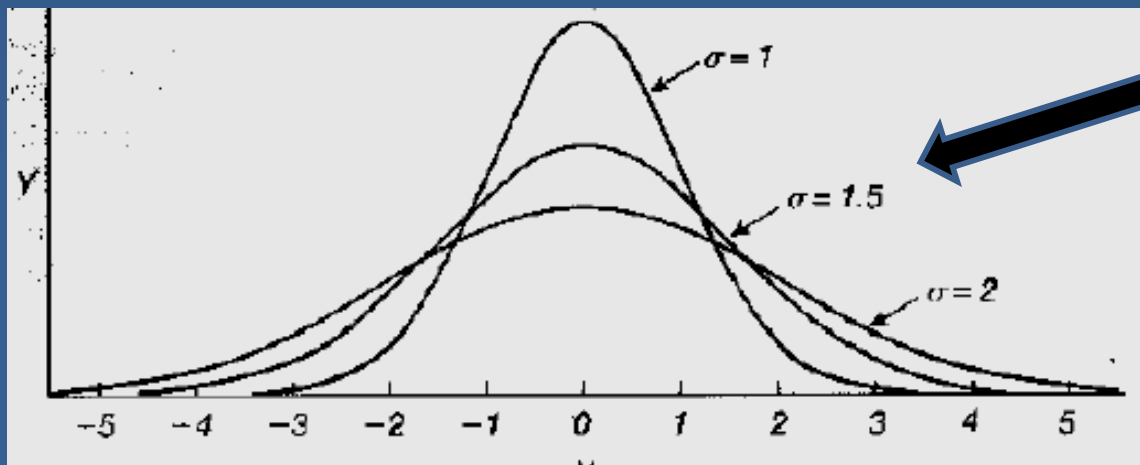
95% leží mezi $\mu \pm 1,96\sigma$

99% leží mezi $\mu \pm 2,576\sigma$

Normální rozložení

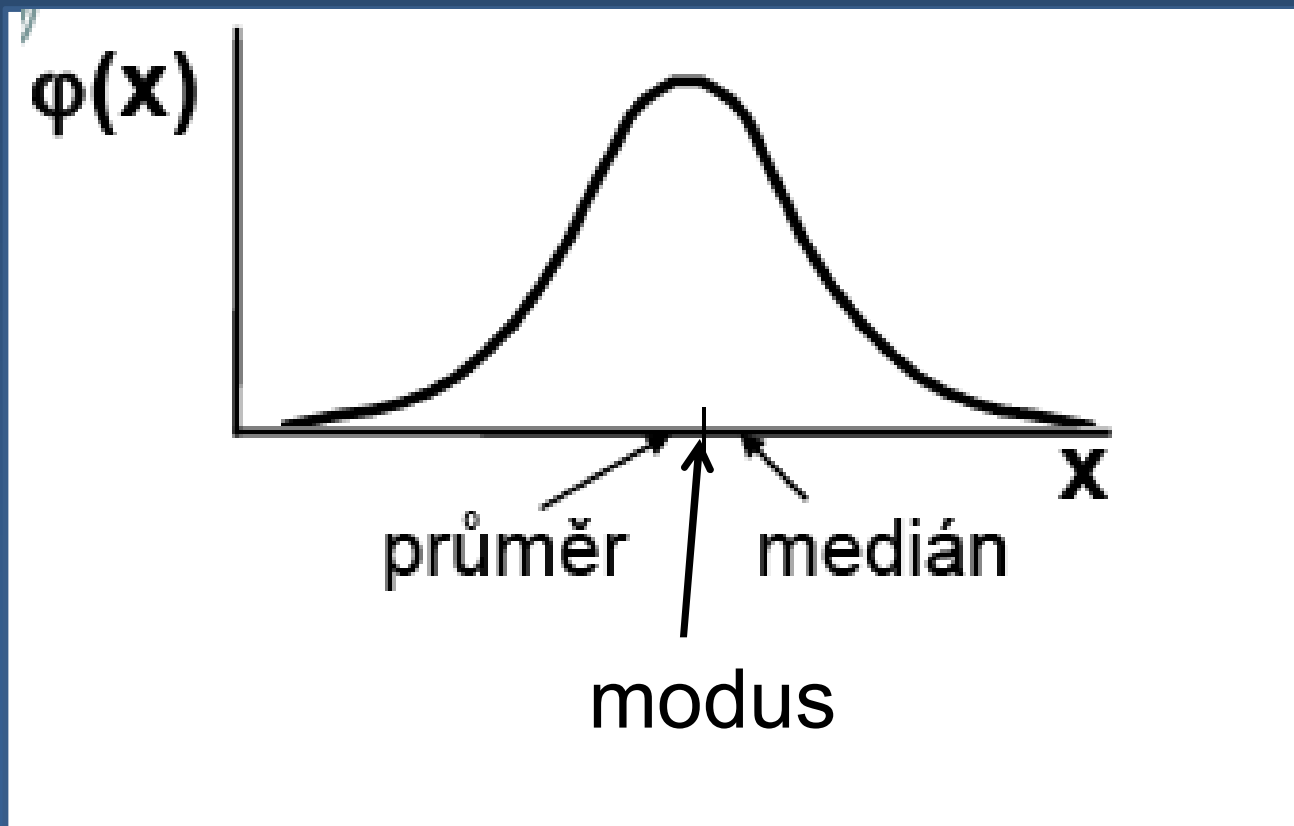


Různá μ , stejné σ



Různá σ , stejné μ

Normální rozložení

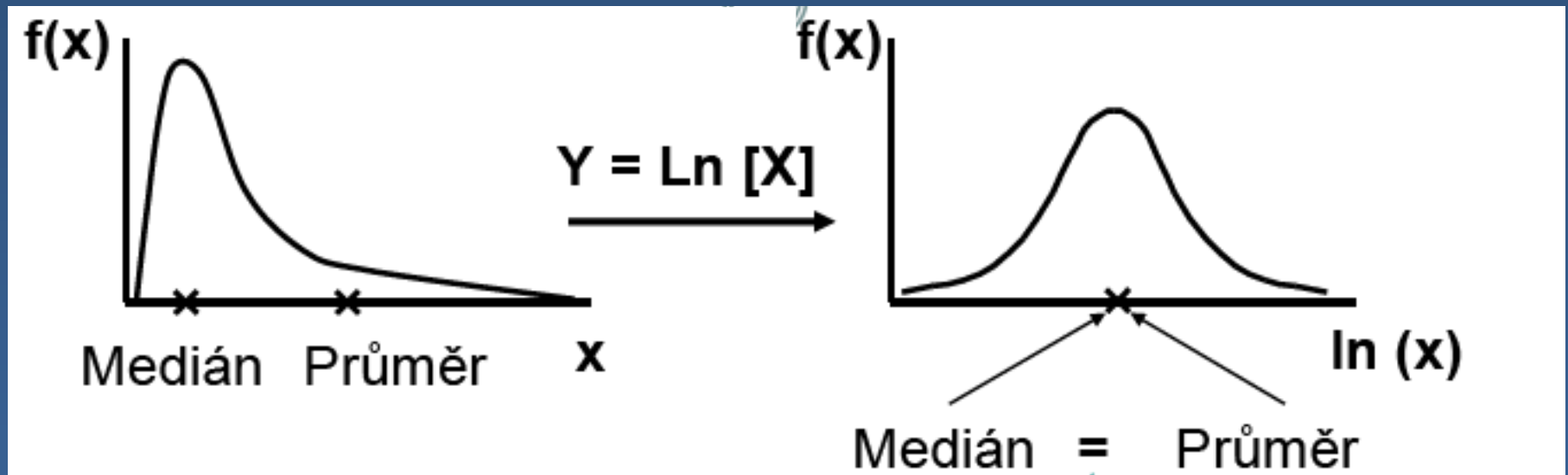


Standardizované normální rozložení

- $Z \sim N(0,1)$
- Libovolné $X \sim N(\mu, \sigma^2)$ můžeme transformovat na veličinu $Z = (X - \mu)/\sigma$, která má standardizované normální rozdělení $Z \sim N(0,1)$

Log-normální rozložení

- $X \sim LN(\mu, \sigma^2) \Leftrightarrow Y = \ln X \sim N(\mu, \sigma^2)$
- Tělesná hmotnost, doba přežití po jedné dávce ozáření,...



Zešikmená data

- Transformujeme původní veličinu na novou, pro kterou je model normálního rozložení přijatelný
- Analýzu provedeme na transformované veličině
- Výsledky analýzy (průměr, intervaly spolehlivosti) lze zpětně transformovat
- Pokud vhodná transformace neexistuje
=> neparametrické metody

