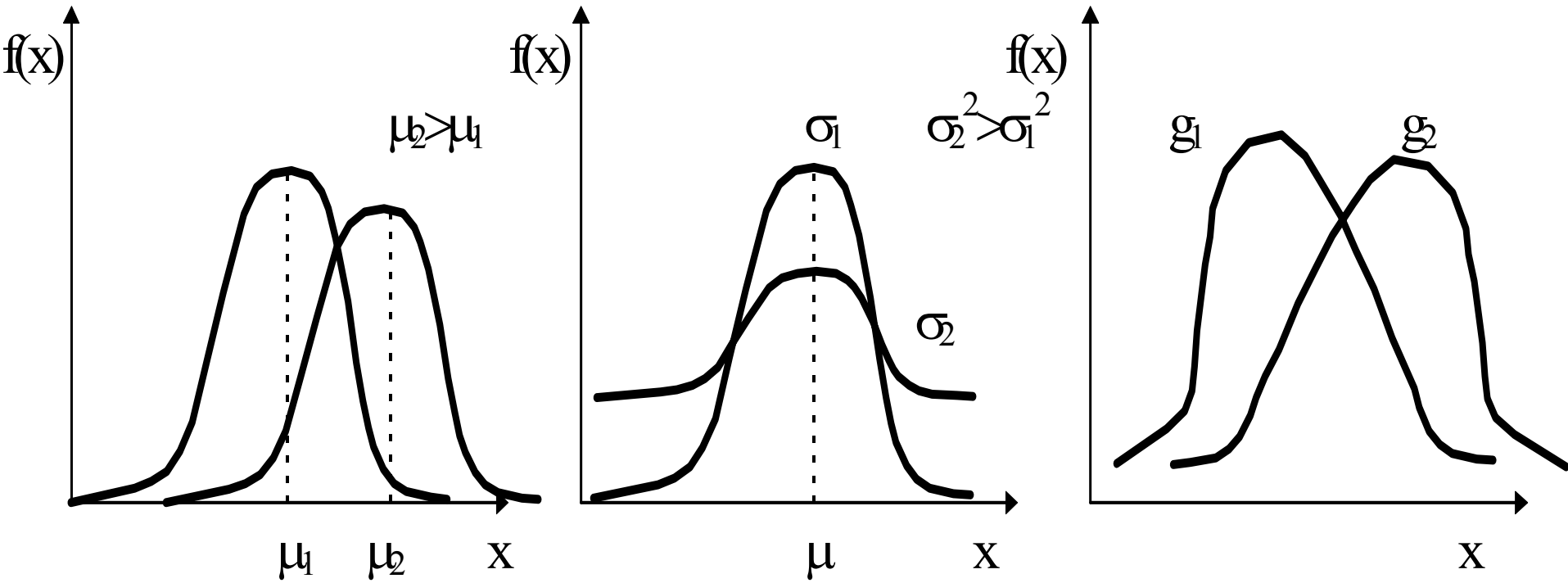
Charakterizace rozdělení

- **Momenty**



$$C_K = \int_{-\infty}^{\infty} (x - M_1)^K f(x)dx \qquad \hat{C}_K = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{M}_1)^K$$

$$M_K = \int_{-\infty}^{\infty} x^K f(x)dx \qquad \hat{M}_K = \frac{1}{N}\sum_{i=1}^{N} x_i^K$$

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx \qquad \sigma^2 = D(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

Speciální momenty

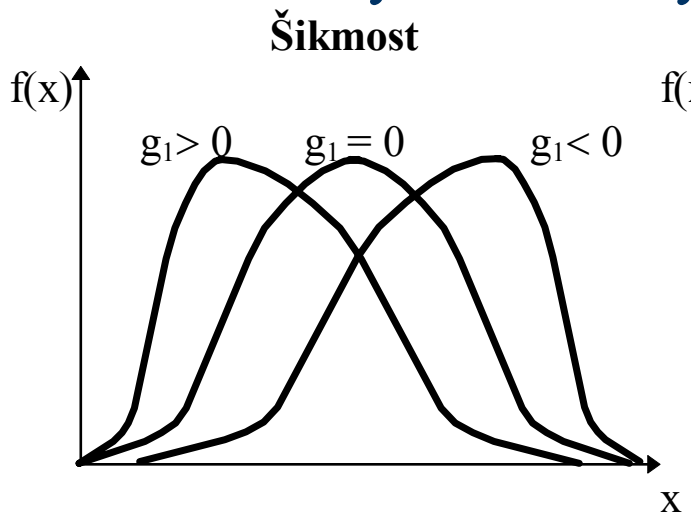
Střední hodnota: **(matematické očekávání) $M_1 = E(X) = \mu$**

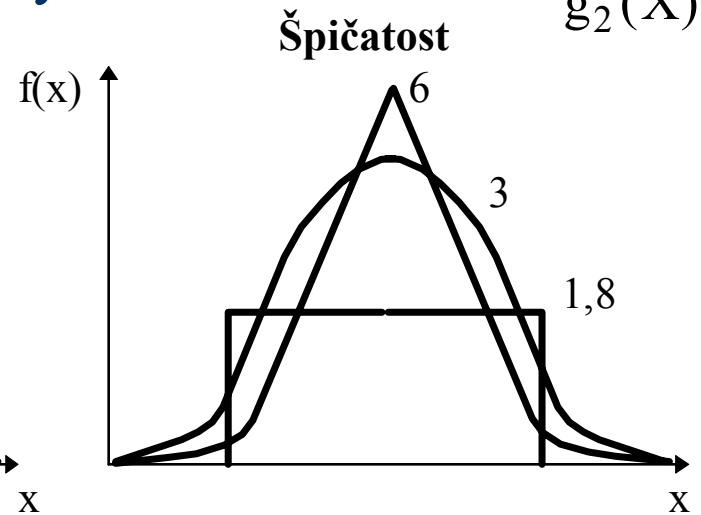
První centrální moment $C_1 = 0$

Rozptyl: **$C_2 = D(X) = \sigma^2$**

Šikmost: Třetí normovaný centrální moment $g_1(X) = \frac{C_3(X)}{\sigma^2 * \sigma}$

Špičatost: Čtvrtý normovaný centrální moment $g_2(X) = \frac{C_4(X)}{\sigma^2 * \sigma^2}$





Střední hodnota

Vlastnosti střední hodnoty: $E(X) \equiv M_1(\mu)$

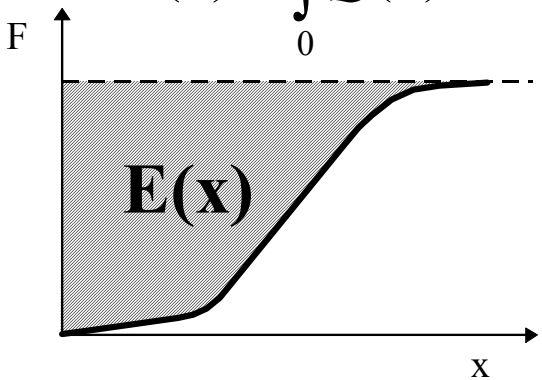
Konstanty A, B $\quad E(A*X \pm B) = A*.E(X) \pm B$

Nezávislé (X, Y) $\quad E(X \pm Y) = E(X) \pm E(Y)$

$E(X*Y) = E(X)*E(Y)$

Korelované (X, Y),

$$E(x) = \int_0^1 Q(u)du$$



Míra závislosti vyjádřená kovariancí

$cov(X, Y) = E(X*Y) - E(X)*E(Y)$

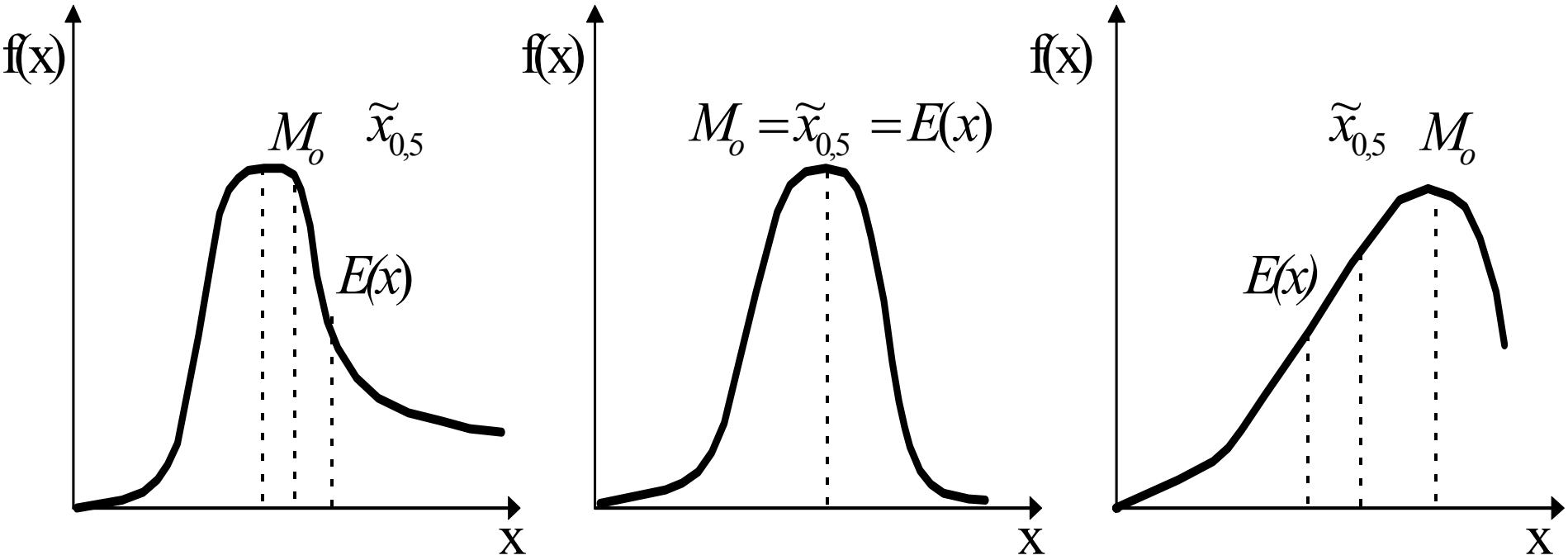
$E(X*Y) = E(X)*E(Y) + cov(X, Y)$

Centrování: $\quad X_0 = X - E(X) \Rightarrow E(X_0) = 0$

Pro positivní $X \geq 0$ platí $E(X) = \int Q(u)\, du$

Parametry polohy

- Módus: M_0 **... relativní maximum na hustotě pravděpodobnosti**
- Medián: $\tilde{x}_{0.5}$ **... 50%ní kvantil**



Pro kladná x platí: $\tilde{x}_\alpha = E(X) / (1-\alpha)$, pak maximálně $\tilde{x}_{0.5} = 2*E(X)$

Kvantilová metoda pro lognormální data

- **Tří parametrové lognormální rozdělení**

$$f(x) = \frac{1}{\sqrt{2} * \pi * (1 + \sigma * Y) * \tau} \exp\{-\frac{1}{2\sigma^2}[\ln(1 + \sigma * Y)]^2\}$$

$$P_i = \frac{i + 0.3133\sigma - 0.3927}{N + 0.2146}$$

Parametr $Y = x - \mu / \tau$ $\quad x \geq (\mu - \tau) / \sigma$

Transformace $z = \ln[(1 + \sigma * Y)^{1/\sigma}]$ **vede na N(0,1)**

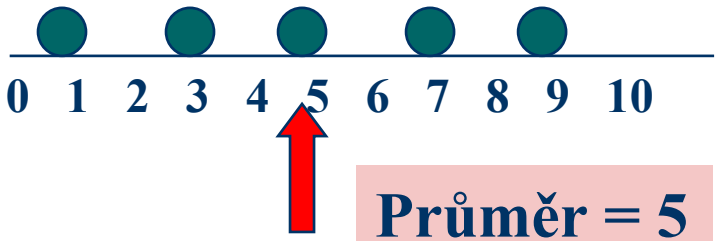
Pořádková statistika $x_{(i)}$ souvisí se standardizovanou normální pořádkovou statistikou $z_{(i)}$

$$x_{(i)} = \mu + \tau * \left[\frac{\exp(\sigma * z_{(i)}) - 1}{\sigma}\right] \approx \mu + \tau * g_i(\sigma)$$

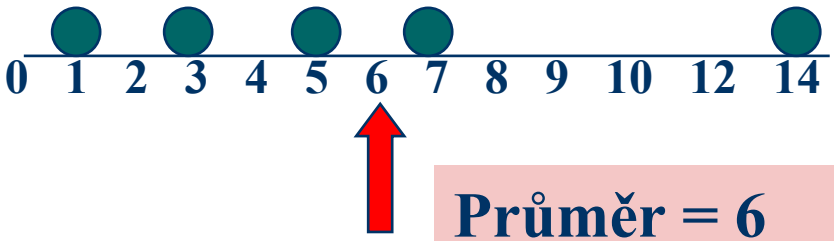
$$g_i(\sigma) = \left[\frac{\exp(\sigma * \Phi^{-1}(P_i) - 1}{\sigma}\right]$$

Pro známé σ je možné odhadnout parametry μ, τ jako směrnici a úsek regresní přímky $x_{(i)} \approx \mu + \tau * g_i(\sigma)$

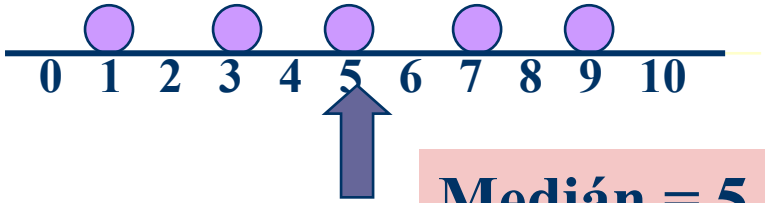
Centrální tendence



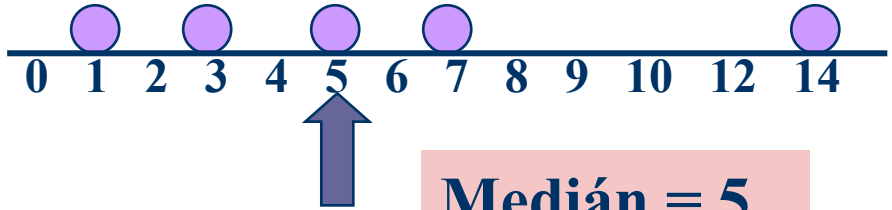
Průměr = 5



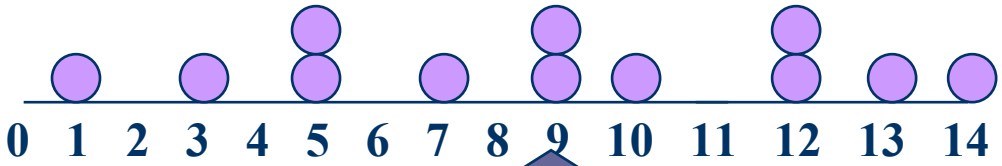
Průměr = 6



Medián = 5



Medián = 5



Módus = 9

Není módus

Rozptyl

$$\left|\tilde{x}_{0.5} - \mu \leq \sigma\right|$$

Vlastnosti rozptylu: $D(X) = C_2 = \sigma^2$

$\mathbf{D(X) = E((X - E(X))^2 = E(X^2 - 2 * X * E(X) + E(X)^2) = E(X^2) - E(X)^2}$

Konstanty A, B $\mathbf{D(A*X \pm B) = A^2 * D(X)}$

Nezávislé (X, Y) $\mathbf{D(X \pm Y) = D(X) + D(Y)}$

Pro centrované X a Y $\mathbf{D(X*Y) = D(X) * D(Y) + E(X)^2 * D(Y) + E(Y)^2 * D(X)}$

Lineárně korelované (X, Y) $\mathbf{D(X \pm Y) = D(X)+D(Y) + 2*cov\ (X, Y)}$

$\mathbf{D(X * y) = D(X)*D(Y) + E(X)^2 * D(Y) + E(Y)^2 * D(X) + 2 * E(X)*E(Y)*cov(X, Y)}$

Pro libovolné A platí $\mathbf{E[(X- A)^2] \geq D(X)}$

Rozptyl je nezáporný $\mathbf{\sigma^2 \geq 0}$

$$D(x) = \int_0^1 [Q(u) - \mu]^2 \, du$$

Standardizace

$U = X_0 / \sqrt{[D(X)]} = [X - E(X)] / \sqrt{[D(X)]}$

Platí, že:

1. $E(U) = 0$ a $D(U) = 1$

2. $f(x) = (1/\sigma) * f(u)$

3. $F(x) = F(u)$

4. $Q(Z) = \mu + \sigma * Q_0(Z)$, pro $0 \leq Z < 1$

Parametry variability

- Směrodatná odchylka $\Rightarrow \sigma = \sqrt{[D(X)]}$

- Průměrná absolutní odchylka $\Rightarrow d = E[abs(X - E(X))]$

- *Obecně platí* $d \leq \sqrt{D(X)}$.

Pro normální rozdělení: $d^2 / \sigma^2 = 2 / \pi$

- Interkvartilová odchylka

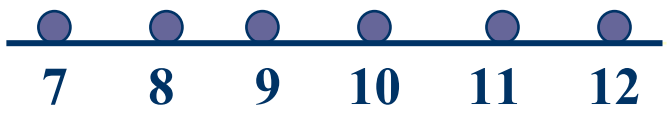
- *Obecně platí* $q \leq \sqrt{[2\sigma^2]}$.

Pro normální rozdělení $q \approx 0.675\ \sigma$

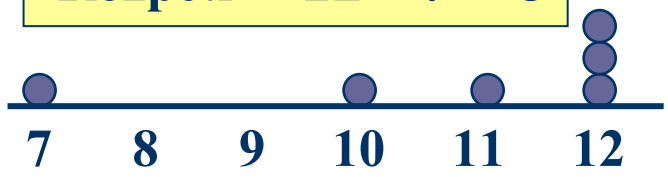
$$q = 0.5 * (\tilde{x}_{0.75} - \tilde{x}_{0.25})$$

Příklady variability

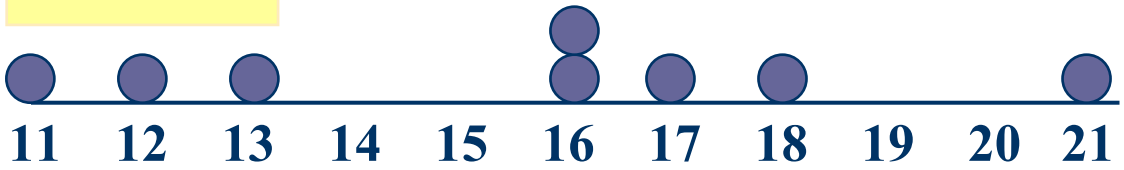
Rozpětí = 12 - 7 = 5



Rozpětí = 12 - 7 = 5



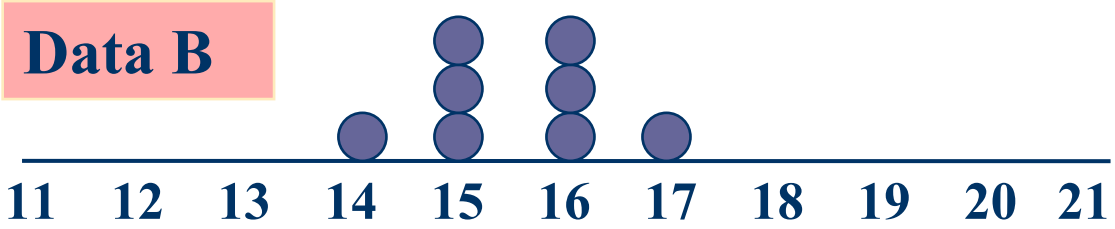
Data A



Průměr = 15.5

s = 3.338

Data B



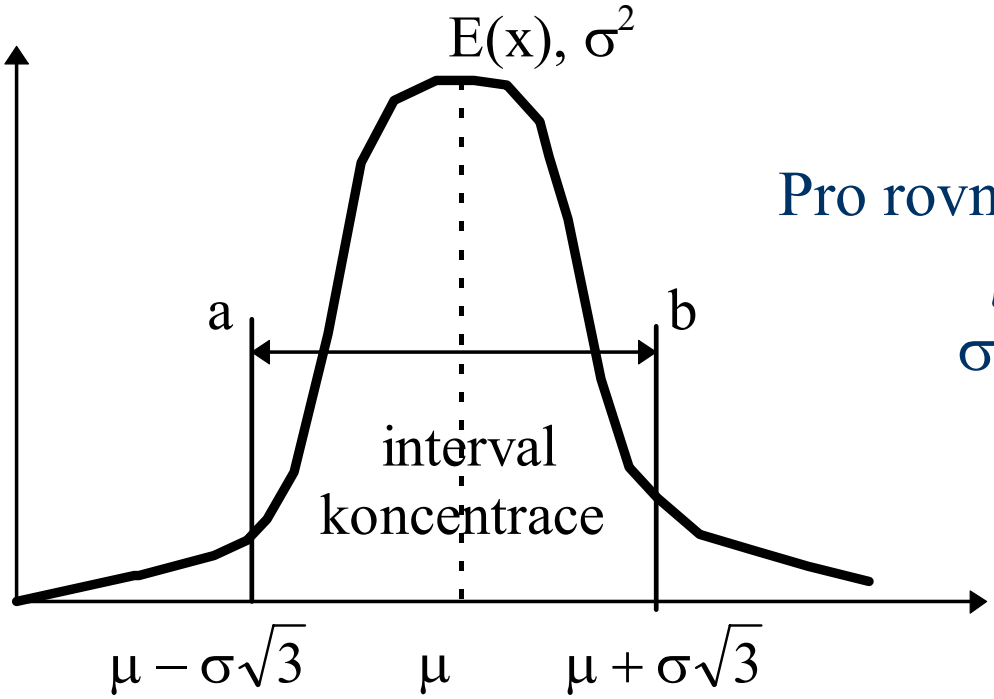
Průměr = 15.5

s = .9258

Aproximace rovnoměrným rozdělením

- Koncentrace:

$$K = \frac{1}{2 * \sigma * \sqrt{3}} \approx \frac{0.2887}{\sigma}$$



Pro rovnoměrné rozdělení je

$\mu = (a + b)/2$

$\sigma^2 = (b - a)^2/12$

Variační koeficient

Relativní mírou variability je **variační koeficient** $\mathbf{V = \sigma / E(X)}$

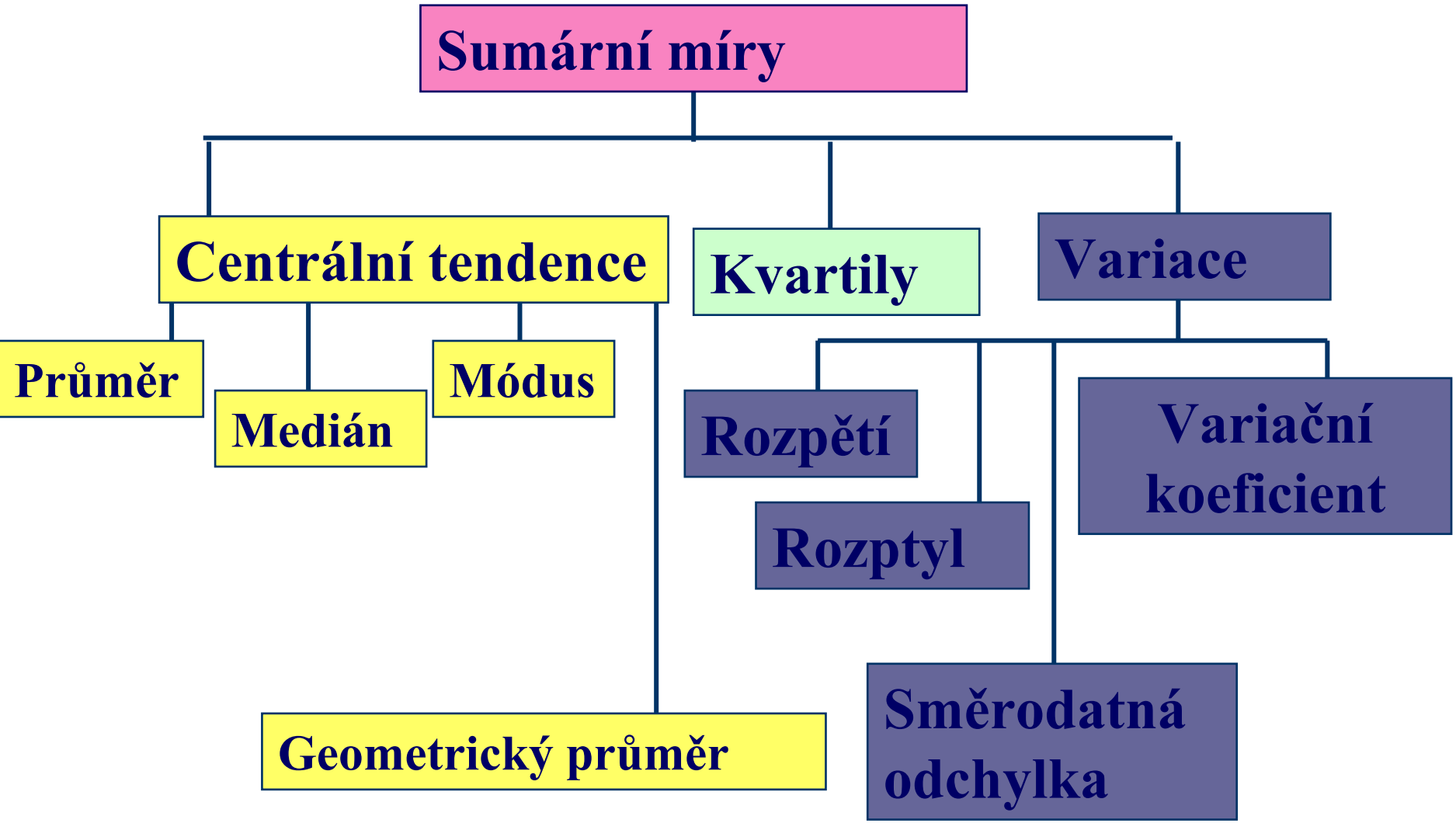
Vlastnosti variačního koeficientu:

1. $\mathbf{V(A) = 0,}$

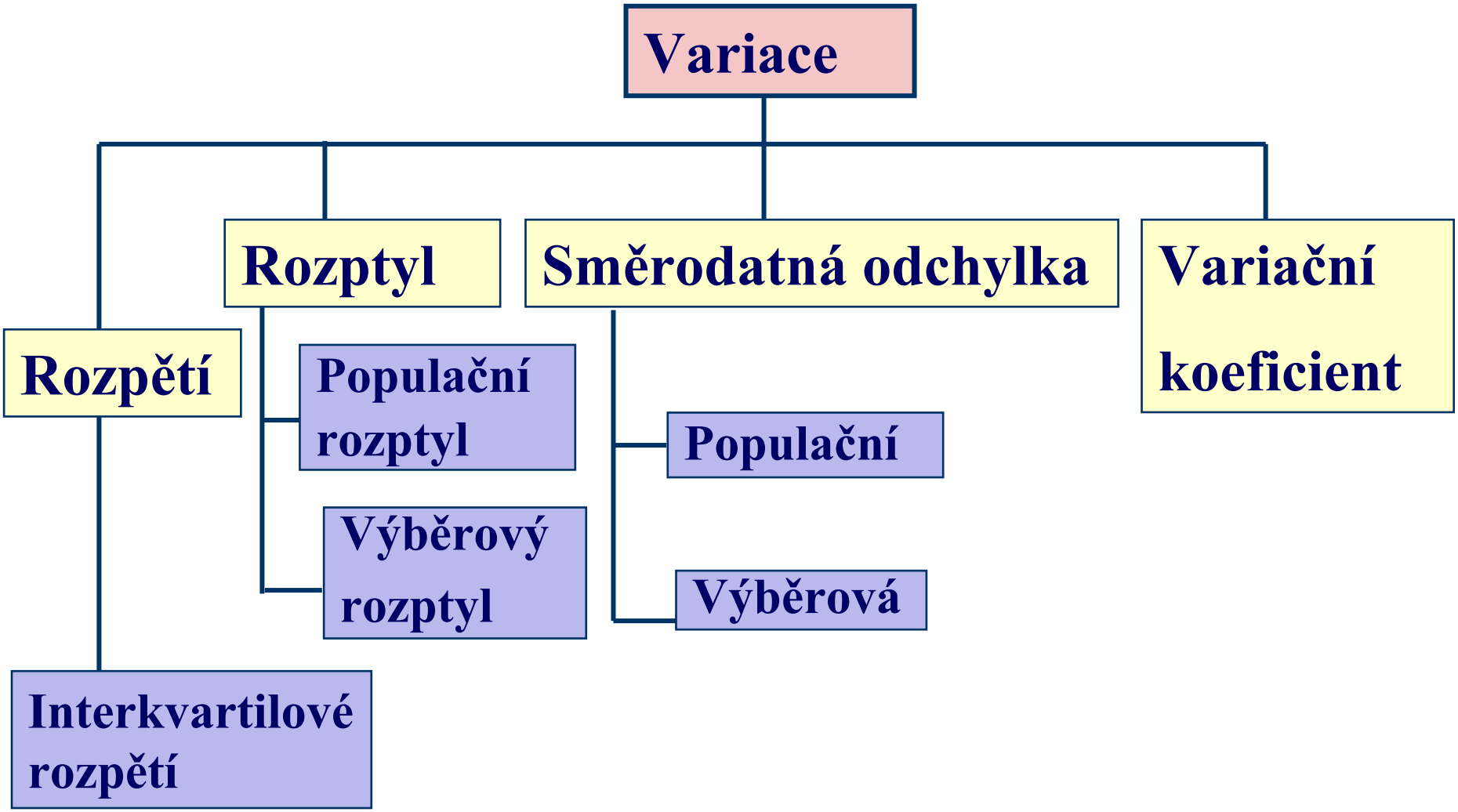
2. $\mathbf{V(A * X) = V(X)}$

3. $$V(X+Y) = \frac{\sqrt{V^2(X) * E^2(X) + V^2(Y) * E^2(Y)}}{E(X) + E(Y)}$$

Přehled sumarizace dat



Přehled měr variability



Popisné statistiky polohy

Centrální tendence

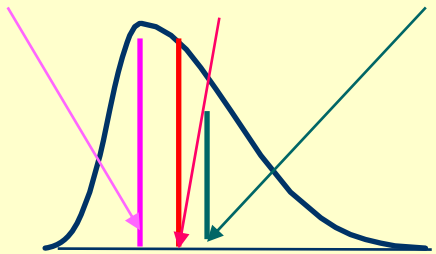
- Průměr aritmetický: ($\Sigma x/n$)
- Medián: Centrální bod (dělicí bod)

Variace

- směrodatná odchylka
- šikmost – náklon rozdělení
- špičatost – délka konců

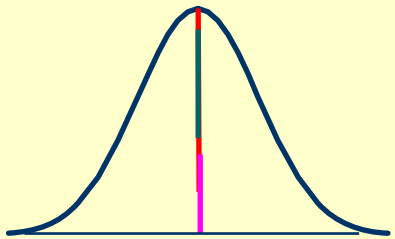
Zešikmené vpravo

Mode < Median < Mean



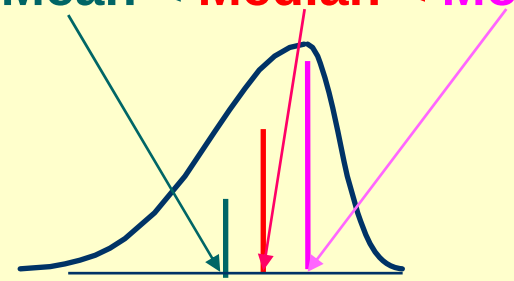
Symetrické

Mean = Median =Mode



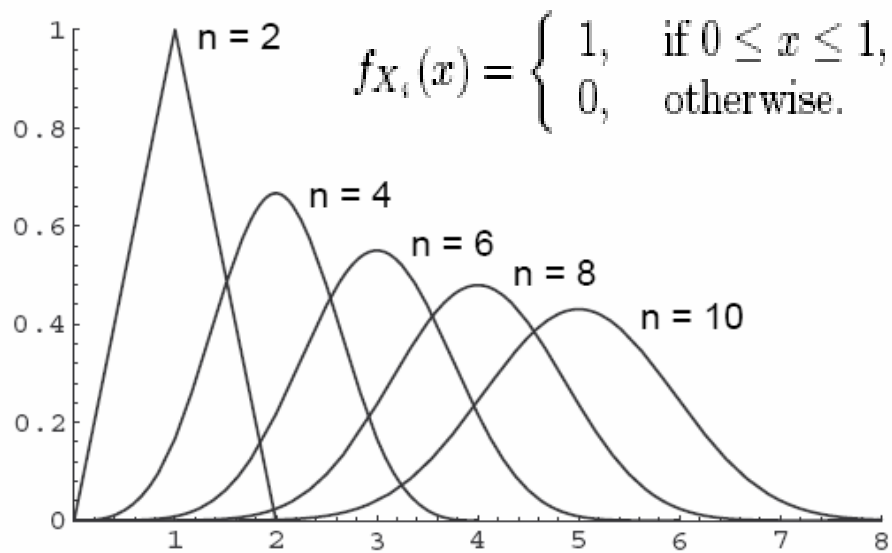
Zešikmené vlevo

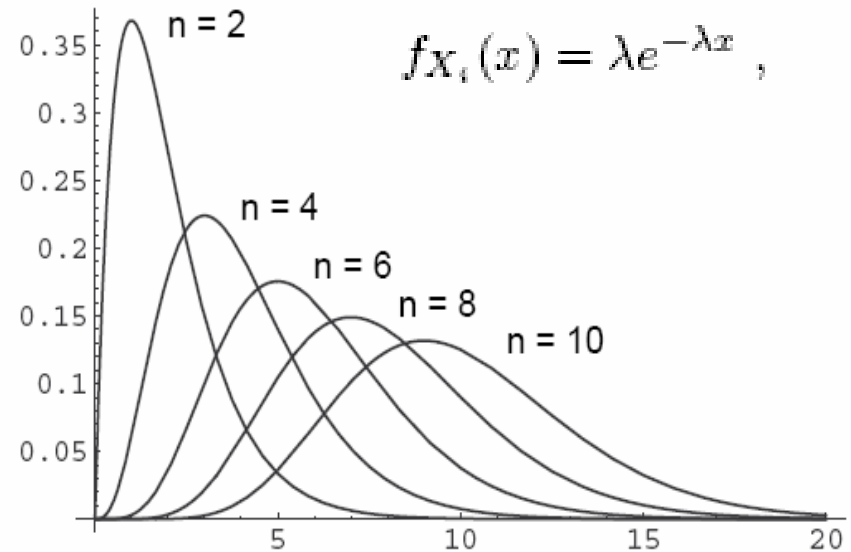
Mean < Median < Mode



Rozdělení součtu náhodných veličin

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y)\, dy$$

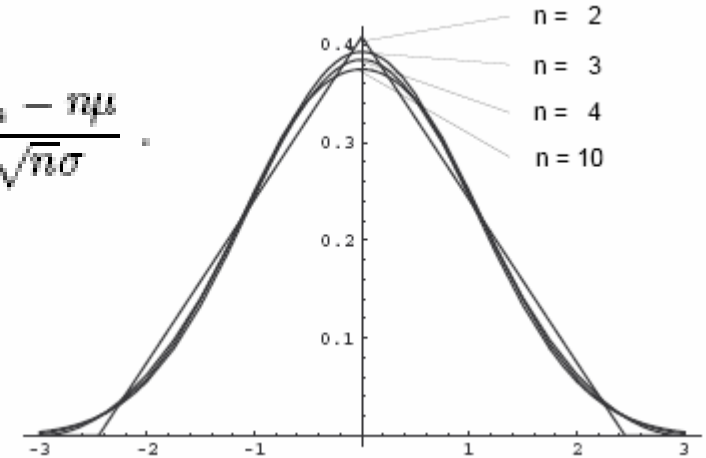




$$f_{X_i}(x) = \begin{cases} 1, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwis} \end{cases}$$

Centrální limitní věta

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma} .$$



- **Nechť** $X_1, X_2, \ldots X_n$ **jsou z nezávislých stejně rozdělených náhodných veličin se střední hodnotou** μ **a směrodatnou odchylkou** $\sigma > 0$

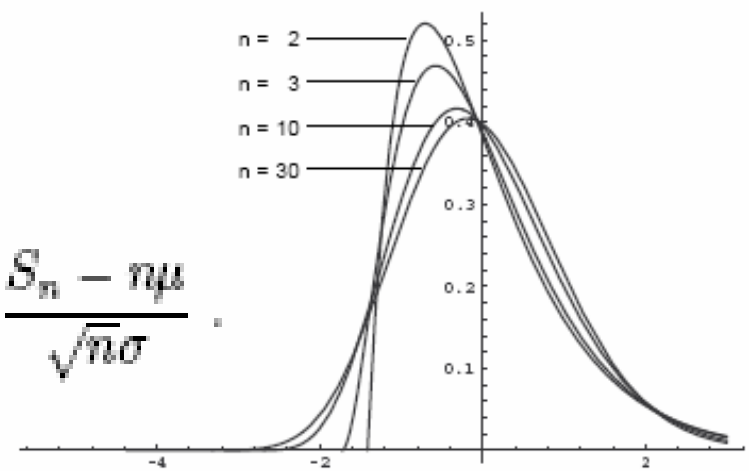
- **Nechť** $S_n = X_1 + X_2 + \ldots + X_n$
- **Pak pro každé x < y**

$$f_{X_i}(x) = \lambda e^{-\lambda x} ,$$

$$\lim_{n \to \infty} P[x \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le y] = \Phi(y) - \Phi(x)$$

kde Φ **je distribuční funkce normálního rozdělení**

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma} .$$



Transformace náhodné veličiny

Známe **X s rozdělením f(x) a transformaci** ("monotónní") **Y = g(X)**

Hledáme **rozdělení Y, tj. h(y)**

a) *diskrétní náhodné veličiny*

$$h(y) = f[\, g^{-1}(y)\,]$$

b) *spojité náhodné veličiny*

$$h(y) = f[\, g^{-1}(y)\,] * \left| d\, g^{-1}(y) / dy) \right|$$

kde $d\, g^{-1}(y) / dy)$ je (normalizační faktor) → Jakobián transformace

Příklady

a) $y = a*x + b$, $g^{-1}(y) = (y-b)/a$, $h(y) = (1/|a|)*.f((y-b)/a)$

b) $y = a*x^2$, $g^{-1}(y) = \sqrt{(y/a)}$, $h(y) = (1/(2\sqrt{|a*y|}))*f(\sqrt{(y/a)})$

c) $y = a * \exp(x)$,$g^{-1}(y) = \ln(y/a)$, $h(y) = f[\ln(y/a)*|1/y|$

- *Lineární transformace*: ***mění se pouze polohu a rozptýlení***
- *Nelineární transformace*: ***již dochází ke změně tvaru rozdělení***

Transformace distribuční funkcí vede na rovnoměrné rozdělení!!

$$y = F(x) \quad f(y) = f[F^{-1}(y)] * [d\,F^{-1}(y)] / dy = f[Q(y)].q(y) = 1$$

$$g_1 = \frac{1 - 2 * p}{\sqrt{N * p * (1 - p)}} \qquad g_2 = \frac{1 - 6 * p * (1 - p)}{N * p * (1 - p)} + 3$$

Binomické rozdělení

Binomické rozdělení B(N,p))

Binomické rozdělení má náhodná veličina X vyjadřující počet výskytu jevu A (příznivý výsledek) v N nezávislých pokusech. Pravděpodobnost výskytu jevu A (příznivý výsledek) v jednom pokusu je p a jevu $\overline{A}$ (nepříznivý výsledek) je q = 1 - p.

Pravděpodobnostní funkce $$P(x = i) = \binom{N}{x} * p^X * (1 - p)^{N - x}$$

Distribuční funkce $$F(x) = \sum_{i=0}^{x} \binom{N}{i} * p^i * (1 - p)^{N - i}$$

Střední hodnota: $E(X) = N * p$

x_p je počet příznivých jevů v *N* nezávislých pokusech

Rozptyl: $D(X) = N * p * (1 - p)$

Módus: $p*(N+1)-1 <= Mo < = p*(N+1)$

$$p = \frac{x_p}{N}$$

$$g_1 = \frac{1}{\sqrt{\lambda}} \qquad g_2 = 3 + \frac{1}{\lambda}$$

Poissonovo rozdělení

Poissonovo rozdělení Po(λ) **má náhodná veličina X, která je rovna počtu jevů v daném časovém nebo prostorovém intervalu,**

kde $\overline{x}$ **je aritmetický průměr počtu jevů v daném časovém nebo prostorovém intervalu**

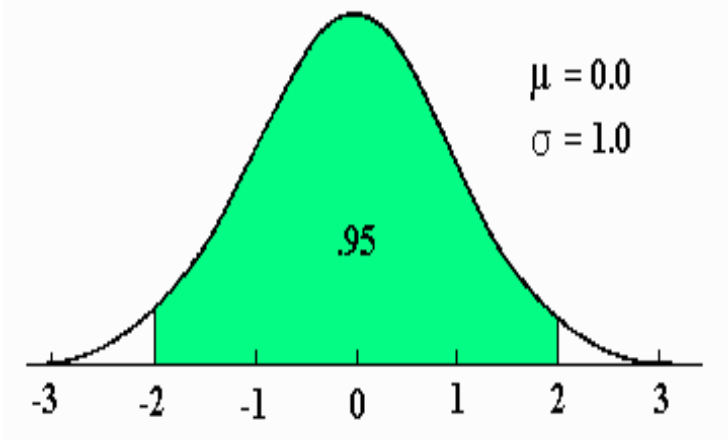
Pravděpodobnostní funkce $P(x = i) = \lambda^x * e^{-\lambda / x!}$

Distribuční funkce $F(x) = \sum_{i=0}^{x} \lambda^i * e^{-\lambda / i!}$

Střední hodnota **E(X) =λ** $\lambda = \overline{x}$

Rozptyl **D(X) =λ**

Normální rozdělení



Normované normální rozdělení U→N(0,1)

P(-1.65 ≤ U ≤ 1.65) = 0.90 **f(2) = 0.05**

P(-1.98 ≤ U ≤ 1.98) = 0.95 **f(3) = 0.04**

P(-3 ≤ U ≤ 3) = 0.9973 **f(4) = 0.00013**

- Normální rozdělení $N((\mu, \sigma^2)$:

Hustota pravděpodobnosti

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

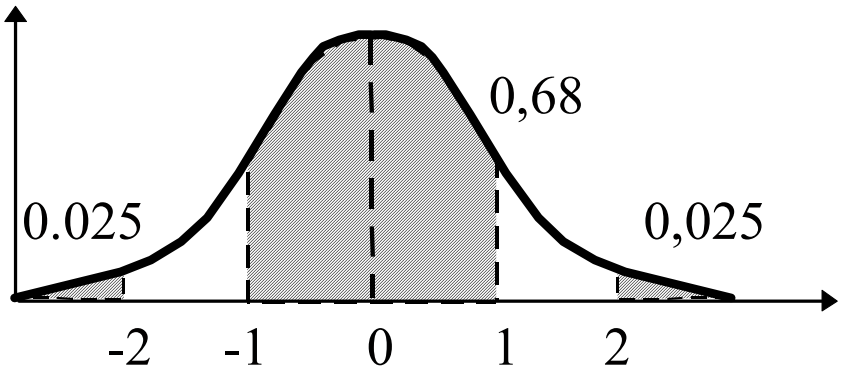
Distribuční funkce $F(x) = \Phi(\frac{x-\mu}{\sigma})$ $\Phi(x) = \frac{1}{\sqrt{2 * \pi}} \int_{-\infty}^{x} \exp(\frac{-y^2}{2}) dy$

Střední hodnota: $E(X) = \mu$

Rozptyl: $D(X) = \sigma^2$

Šikmost: $g_1 = 0$

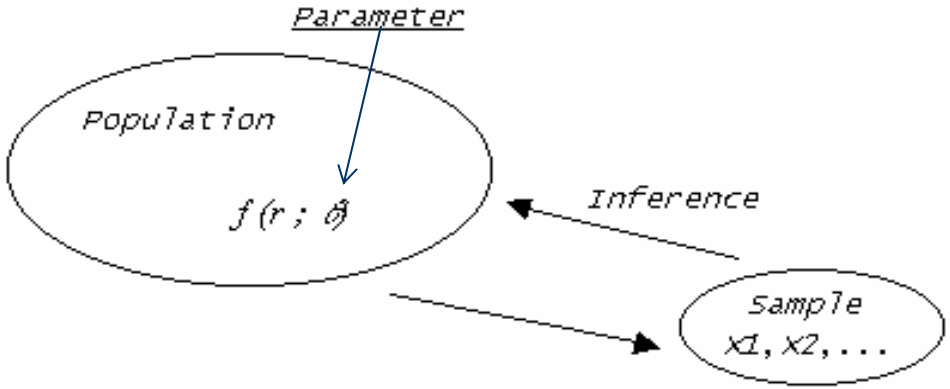
Špičatost: $g_2 = 3$



$$\hat{\mu} = \bar{x} \qquad \bar{x} = \Sigma x_i / N$$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

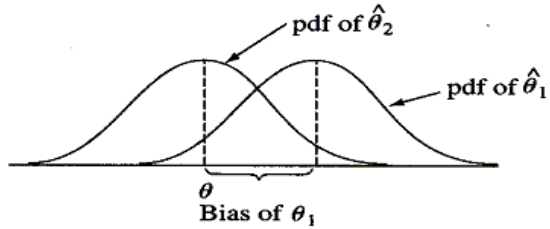
Matematická statistika



Populace X — *vzorkování* — *Výběr $\{x_i\}$ i=1,...N*

$f(x, \mu, \sigma^2, g_1, g_2)$ — $\hat{f}(x),\ \hat{\mu},\ \hat{\sigma}^2, \hat{g}_1,\ \hat{g}_2$

Symbol "∧" označuje odhady parametrů **nebo** hustoty pravděpodobnosti **z dat.**



Bodové odhady

Parametr a odhad $\hat{a}$ je náhodná proměnná.

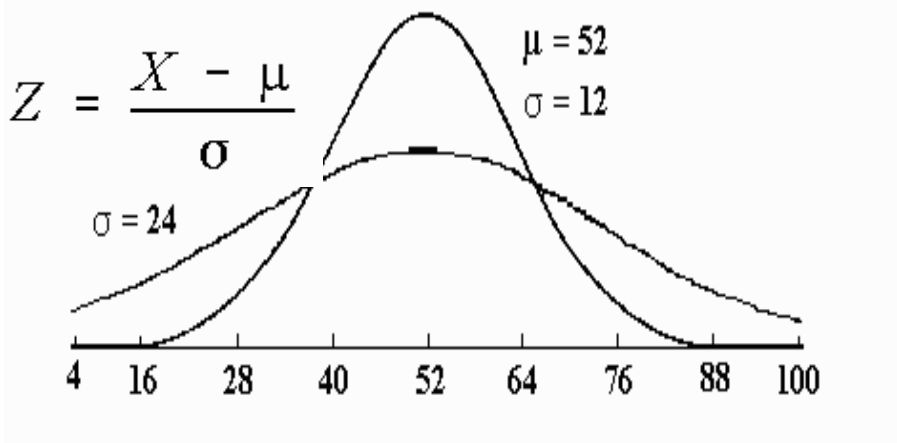
Vychýlení odhadu $b = a - E(\hat{a})$

Pokud je b = 0 jde o **nevychýlený odhad.**

Rozptyl odhadu $D(\hat{a})$ je charakterizací **"přesnosti odhadu"**

Normální rozdělení

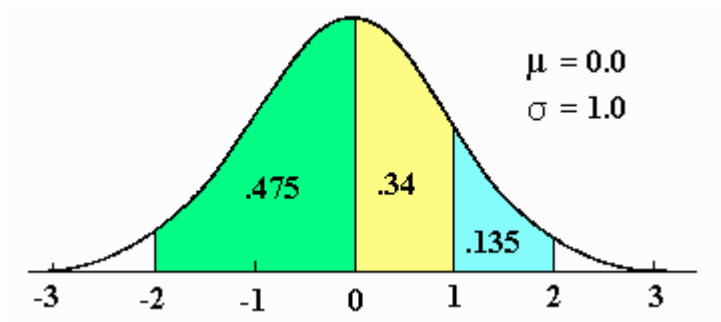
X_i mají rozdělení $N(\mu\ \sigma)^2$



Parametr μ **odhad** $\bar{x}$ **rozptyl** $D(\bar{x}) = \dfrac{\sigma^2}{N}$

$$D(\bar{x}) = \frac{1}{N^2} D(\Sigma x_i) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

Parametr σ^2 **odhad** s^2 **rozptyl** $D(s^2) = \dfrac{2 * \sigma^4}{N}$



Ne-normální rozdělení

Korelační koeficient

$$\rho(\bar{x}, s^2) = \frac{g_1(x)}{\sqrt{n}}$$

Pokud x_i nemají normální rozdělení

$$D(s) \approx \frac{\sigma^2 (g_2 - 1)}{4N}$$

Směrodatná odchylka

- **Směrodatná odchylka s je vychýleným odhadem veličiny σ.**
- **Platí, že E(s) < σ. Pro nevychýlený odhad lze odvodit**

$$\hat{\sigma} = K_u * s \approx \sqrt{\frac{s^2 (N-1)}{N - 1.45}}$$

kde $$K_u = \Gamma\left(\frac{N-1}{2}\right)\sqrt{\frac{N-1}{2}} \Big/ \Gamma\left(\frac{N}{2}\right) \approx \sqrt{\frac{N-1}{N}}\left(1 - \frac{3}{4N} - \frac{37}{32N^2} \cdots\right)$$

$$D(\hat{\sigma}) = \sigma^2 \big/ (2N)$$

$$D(s) \approx \frac{\sigma^2}{2(N-1)}$$

Variační koeficient

- **Výběrový odhad V a odpovídající rozptyl D(V) lze vyjádřit ve tvaru**

$$V = \frac{\bar{x}}{s}$$

$$\mathrm{D(V)} \approx \delta \left[\frac{N + (2N-1)\delta^2}{2N(n-1)} \right] \propto \frac{\delta^2}{2N}(1 + 2\delta^2)$$

Odhady jsou pouze asymptotické (platí pro velké N).

Šikmost a špičatost

- **Pro odhad šikmosti g_1 se používá výběrová šikmost $\hat{g}_1$ s rozptylem $D(\hat{g}_1)$, kde**

$$\hat{g}_1 = \frac{\sqrt{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{3/2}} \qquad D(\hat{g}_1) \approx \frac{6N(N-1)}{(N-2)(N+1)(N+3)}$$

- **Pro odhad špičatosti g_2 platí**

$$\hat{g}_2 = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{2}} \qquad D(\hat{g}_2) \approx \frac{24N(N-1)^2}{(N-3)(N-2)(N+5)(N+3)}$$

Omezení

x_i mají rozdělení N(μ , σ^2)

Maximální vybočující měření x_k je omezeno nerovností

$$(x_K - \bar{x})^2 \leq \frac{(n-1)^2}{n} s^2$$

Omezení na rozmezí dat

$$(x_{(n)} - x_{(1)})^2 \leq 2 * (n-1) * s^2$$

Kvantilová metoda pro lognormální data

- **Tří parametrové lognormální rozdělení** $P_i = \dfrac{i + 0.3133\sigma - 0.3927}{N + 0.2146}$

$$f(x) = \frac{1}{\sqrt{2} * \pi * (1 + \sigma * Y) * \tau} \exp\{-\frac{1}{2\sigma^2}[\ln(1 + \sigma * Y)]^2\}$$

Parametr $Y = x - \mu / \tau \qquad x \geq (\mu - \tau) / \sigma$

Transformace $z = \ln[(1 + \sigma * Y)^{1/\sigma}]$ **vede na N(0,1)**

Pořádková statistika $x_{(i)}$ souvisí se standardizovanou normální pořádkovou statistikou $z_{(i)}$

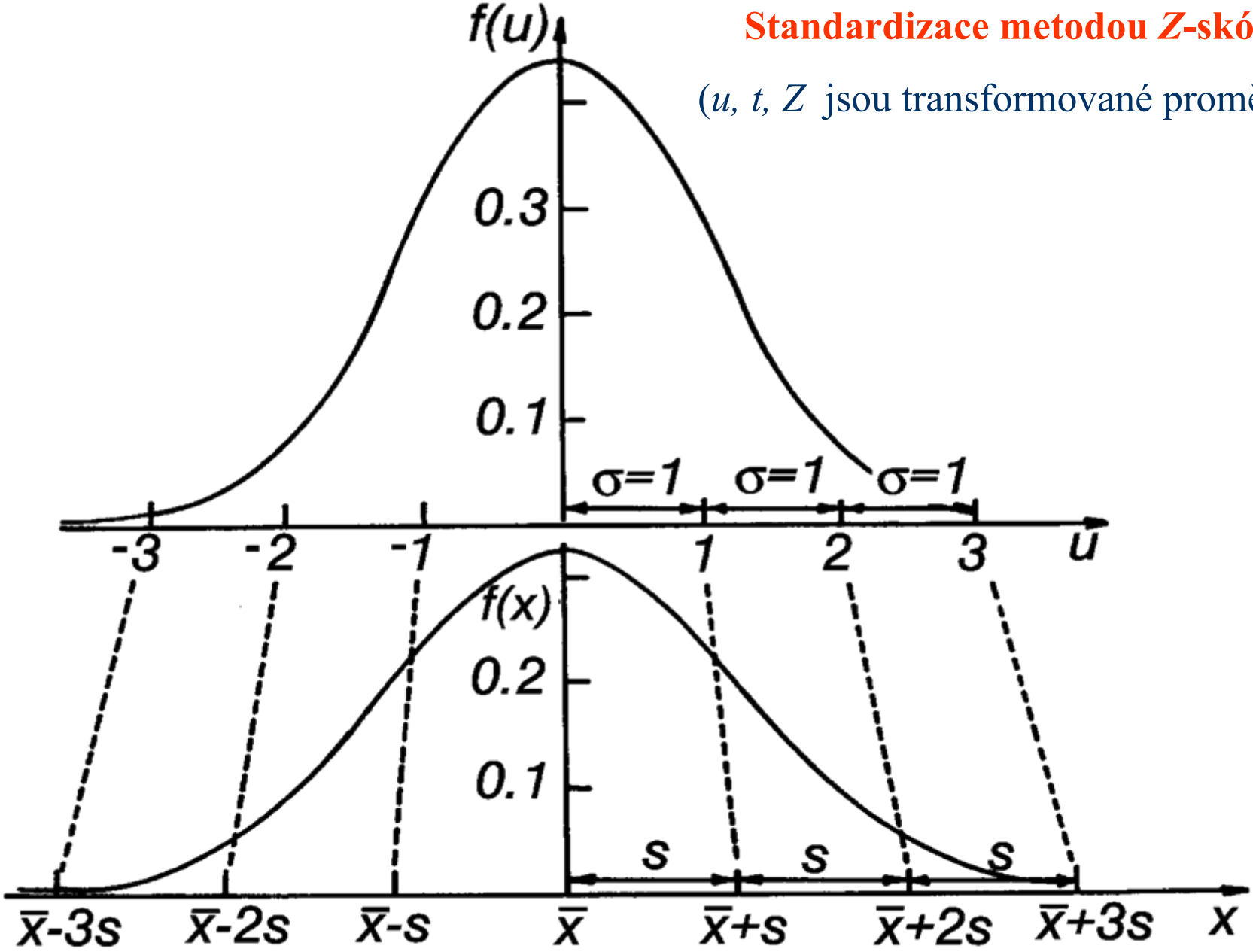
$$x_{(i)} = \mu + \tau * \left[\frac{\exp(\sigma * z_{(i)}) - 1}{\sigma}\right] \approx \mu + \tau * g_i(\sigma)$$

Pro známé σ je možné odhadnout parametry μ, τ jako směrnici a úsek regresní přímky $x_{(i)} \approx \mu + \tau * g_i(\sigma)$

$$g_i(\sigma) = \left[\frac{\exp(\sigma * \Phi^{-1}(P_i) - 1}{\sigma}\right]$$

Standardizace metodou *Z*-skóre

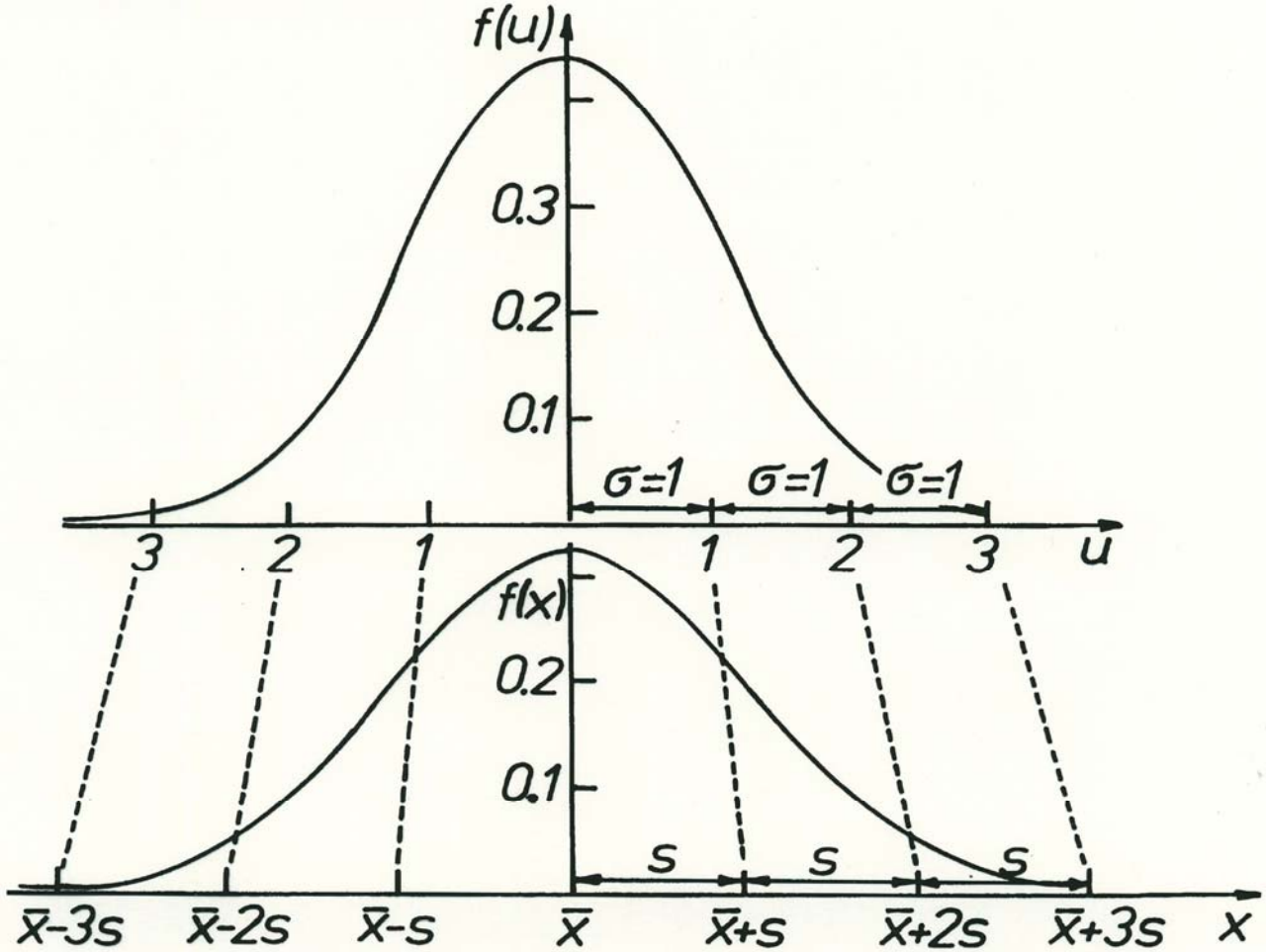
(*u, t, Z* jsou transformované proměnné)



Lineární transformace $u = (x - \mu)/\sigma$ se nazývá ***normování***,

Veličina u má střední hodnotu $E(u) = 0$ a rozptyl $D(u) = 1$.

Normované normální rozdělení N(0, 1) je součástí řady statistických tabulek.

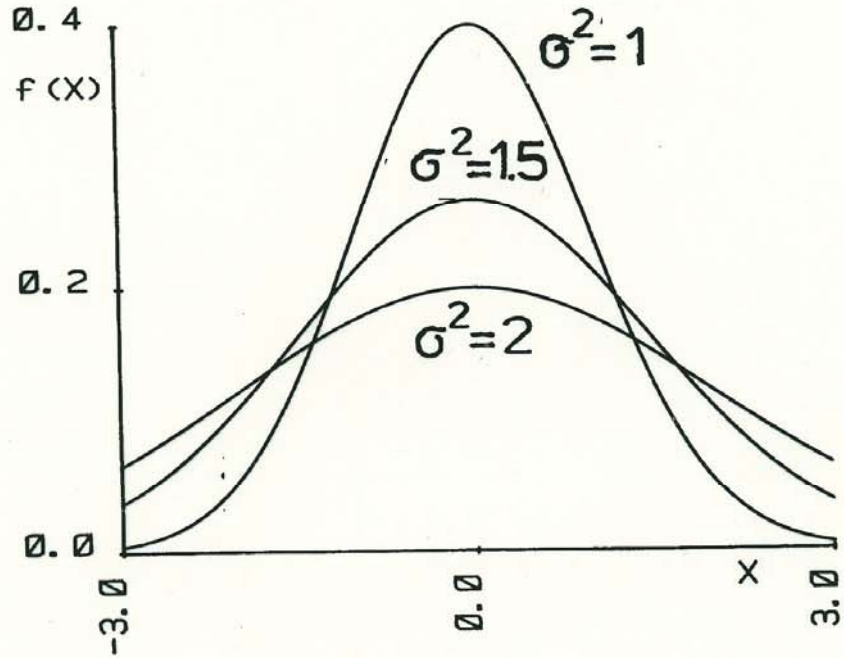


Normování normálního rozdělení

Normální rozdělení (Gaussovo)

Normálně rozdělená náhodná veličina: vzniká složením (součtem) více různých náhodných složek, vlivů a veličin, které jsou navzájem nezávislé a každá z nich ovlivňuje výslednou veličinu jen malým příspěvkem.

Nejčastěji v přírodních vědách (chemii, biologii, fyzice, medicině, atd.)



Hustota pravděpodobnosti normálního rozdělení pro $\sigma^2 = 1, 1.5, 2$

Výběr $\{x_i\}$, i = 1, ..., n, obsahuje prvky x_i, které jsou vzájemně nezávislé a pocházejí z normálního rozdělení.

Hustota pravděpodobnosti normálně rozdělené náhodné veličiny nabývající hodnot x v intervalu $(-\infty, \infty)$ má tvar

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

kde $\mu = E(x)$ představuje **střední hodnotu,**

$\sigma^2 = D(x)$ **rozptyl,**

Šikmost $g_1 = 0$,

Špičatost $g_2 = 3$.

Příklad 3.20 ***Pravděpodobnostní výpočty pro normálně rozdělená data***

Určete pro normální rozdělení naměřených hodnot souboru se střední hodnotou μ a rozptylem σ^2:

a) relativní podíl hodnot, které leží v intervalu od μ do $\mu + \sigma$,

b) relativní podíl hodnot, které leží v intervalu $\langle \mu - K\,\sigma, \mu + K\,\sigma \rangle$, K = 1 a 2,

c) relativní podíl hodnot, které leží nad mezí $\mu + 2\sigma$,

d) mezní hodnotu x_m, kterou přesáhne právě 1 % (5%) všech možných měření.

Řešení:

Pravděpodobnost P výskytu hodnot souboru v intervalu $\langle x_1, x_2 \rangle$ je rovna ***rozdílu hodnot distribuční funkce pro horní mez x_1 a dolní mez x_2,***

tj. $P = F(x_2) - F(x_1)$ ale vyčíslené pomocí normované veličiny $P = F(u_2) - F(u_1)$:

1) 100-násobek P [%] představuje relativní, procentuelní zastoupení hodnot souboru, vyskytujících se v daném intervalu.

2) Pro výpočet se užívá vždy **normovaná veličina** $\mathbf{u = (x - \mu)/\sigma}$, takže

horní mez intervalu bude $\mathbf{u_2 = (x_2 - \mu)/\sigma}$,

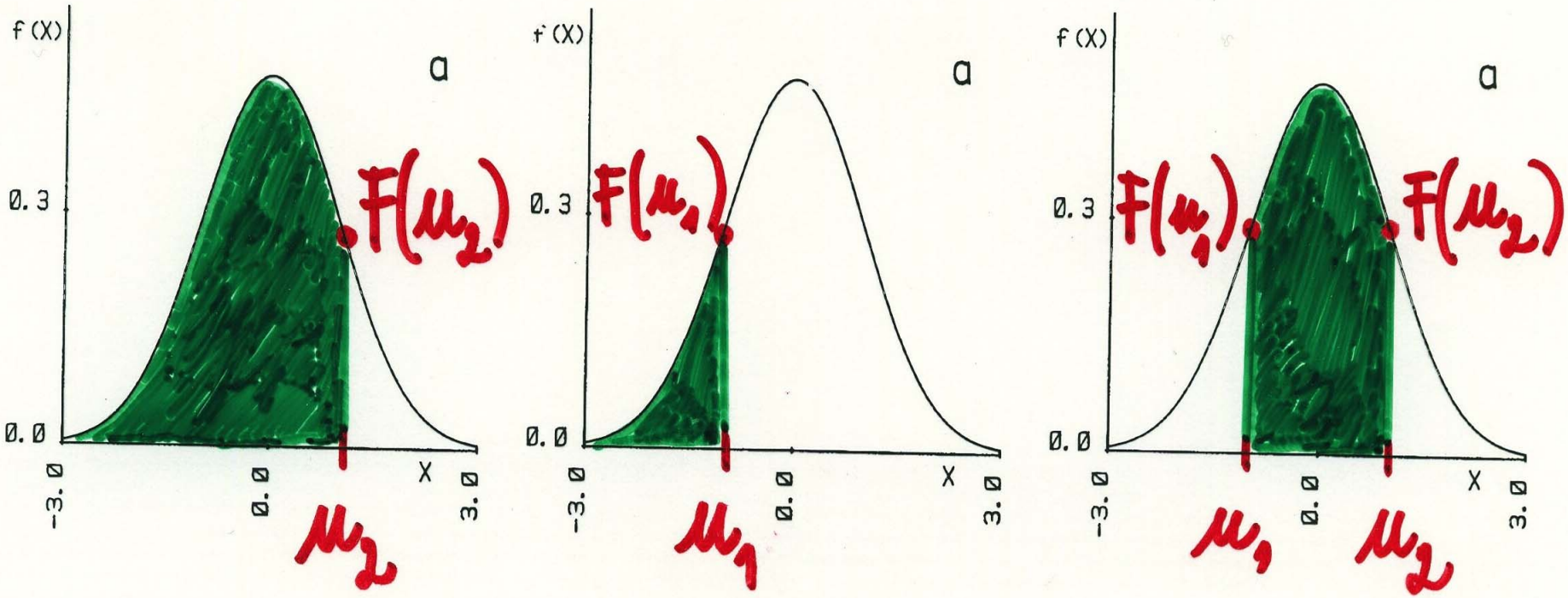
a dolní mez intervalu bude $\mathbf{u_1 = (x_1 - \mu)/\sigma}$.

3) **Hodnoty distribuční funkce F(u)** pro nezáporné normované u ≥ 0 jsou tabelovány. Pro záporné hodnoty $u < 0$ je hodnota rovna dle vztahu

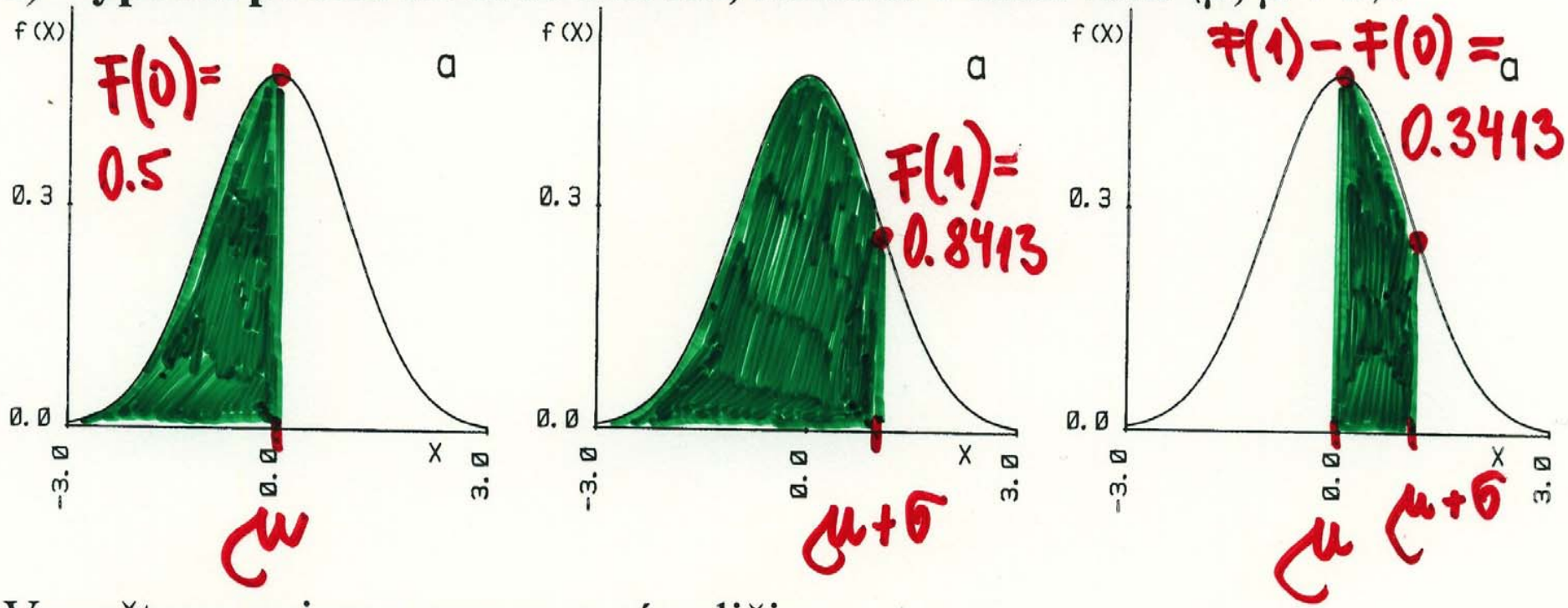
$$F(-u) = 1 - F(u)$$

a hodnota distribuční funkce pro interval od u_1 do u_2 (když $u_2 > u_1$) je

$$F(u_1, u_2) = F(u_2) - F(u_1)$$



a) Výpočet podílu hodnot měření, ležících v intervalu ⟨μ, μ + σ⟩:



Vypočteme nejprve normované veličiny, a to

dolní mez $u_1 = (\mu - \mu)/\sigma = 0$;

a horní mez $u_2 = (\mu + \sigma - \mu) / \sigma = 1$

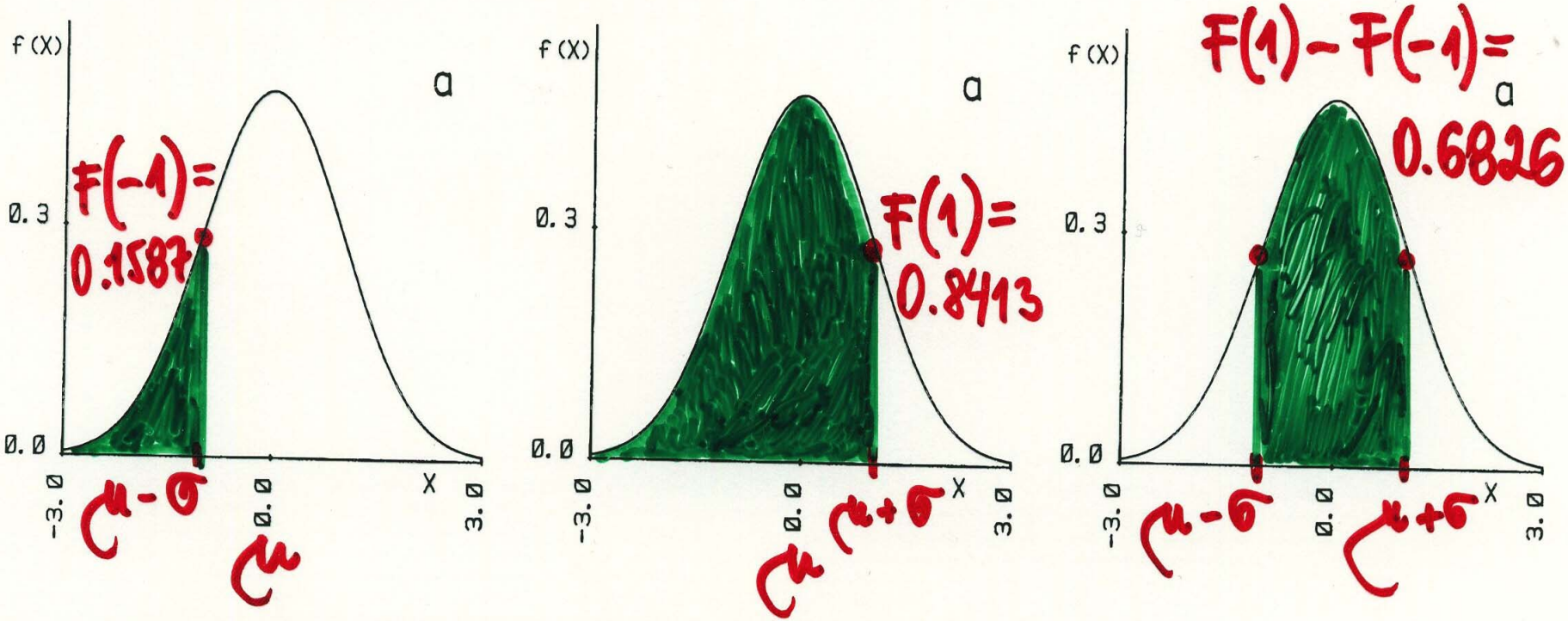
a pak vyčíslíme pravděpodobost P

$$\mathbf{P = F(u_1, u_2) = F(u_2) - F(u_1) = F(1) - F(0) = 0.8413 - 0.5 = 0.3413.}$$

$$\mathbf{P \times 100\ \% = 34.13\ \%}$$

Závěr a): V daném intervalu se bude vyskytovat 34.13 % všech měření.

b) **Výpočet podílu měření, ležících v intervalu ⟨μ - K σ, μ + K σ⟩**: podíl je symetrický kolem střední hodnoty



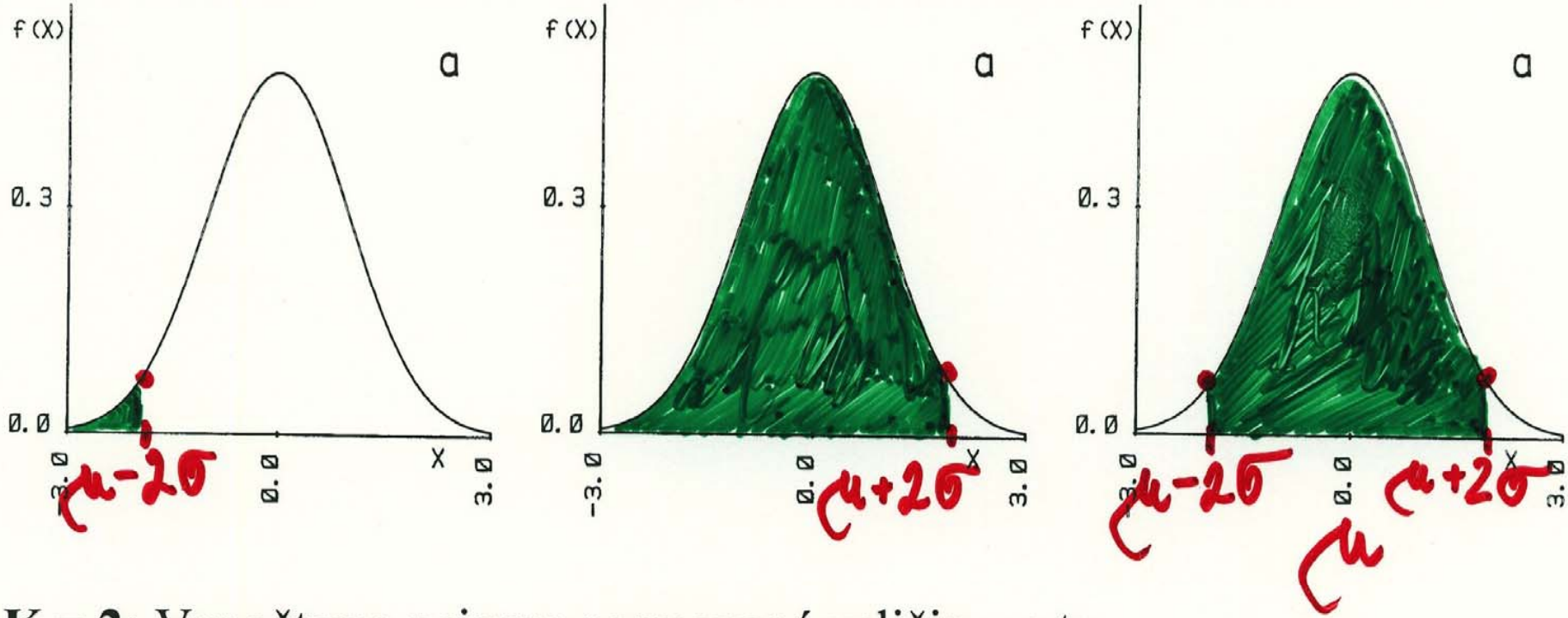
K = 1: Vypočteme nejprve normované veličiny, a to

dolní mez $u_1 = (\mu - \sigma - \mu)/\sigma = -1$;

a horní mez $u_2 = (\mu + \sigma - \mu)/\sigma = 1$

a pak vyčíslíme pravděpodobost P

$$\mathbf{P = F(u_1, u_2) = F(u_2) - F(u_1) = F(1) - F(-1) = 0.8413 - (1 - 0.8413) = 0.6826, \text{ tj. } 68.26\%.}$$



K = 2: Vypočteme nejprve normované veličiny, a to

dolní mez $u_1 = (\mu - 2\sigma - \mu)/\sigma = -2$;

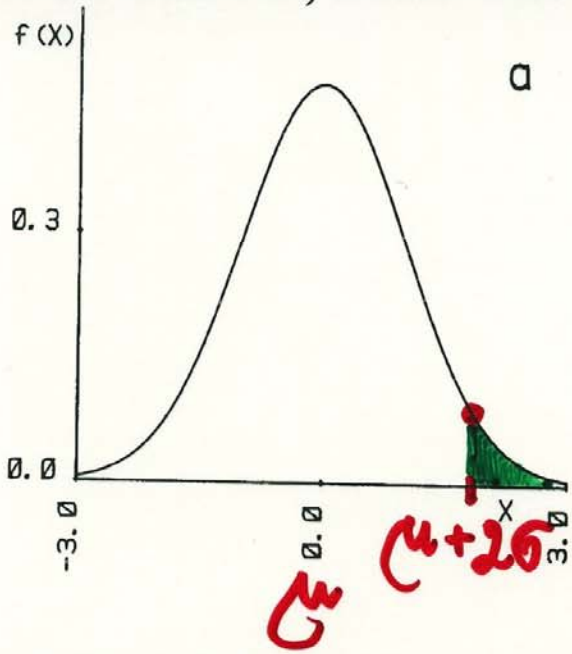
a horní mez $u_2 = (\mu + 2\sigma - \mu)/\sigma = 2$

a pak vyčíslíme pravděpodobost P

$$\mathbf{P = F(u_1, u_2) = F(u_2) - F(u_1) = F(2) - F(-2) = 0.97725 - (1 - 0.97725) = 0.9545, \text{ tj. } 95.45\%.}$$

Závěr b): V daném intervalu se pro K = 1 bude vyskytovat pro 68.26% a pro K = 2 pak 95.45 % všech měření.

c) **Výpočet relativního podílu hodnot, které leží nad mezí $\mu + 2\sigma$:**



Pro určení pravděpodobnosti P_v, s jakou převyšují naměřené hodnoty hranici $x_2 = \mu + 2\sigma$ platí $P_v = 1 - F(x_2)$ a je třeba určit odpovídající u_2:

$$\text{horní mez } u_2 = (\mu + 2\sigma - \mu)/\sigma = 2$$

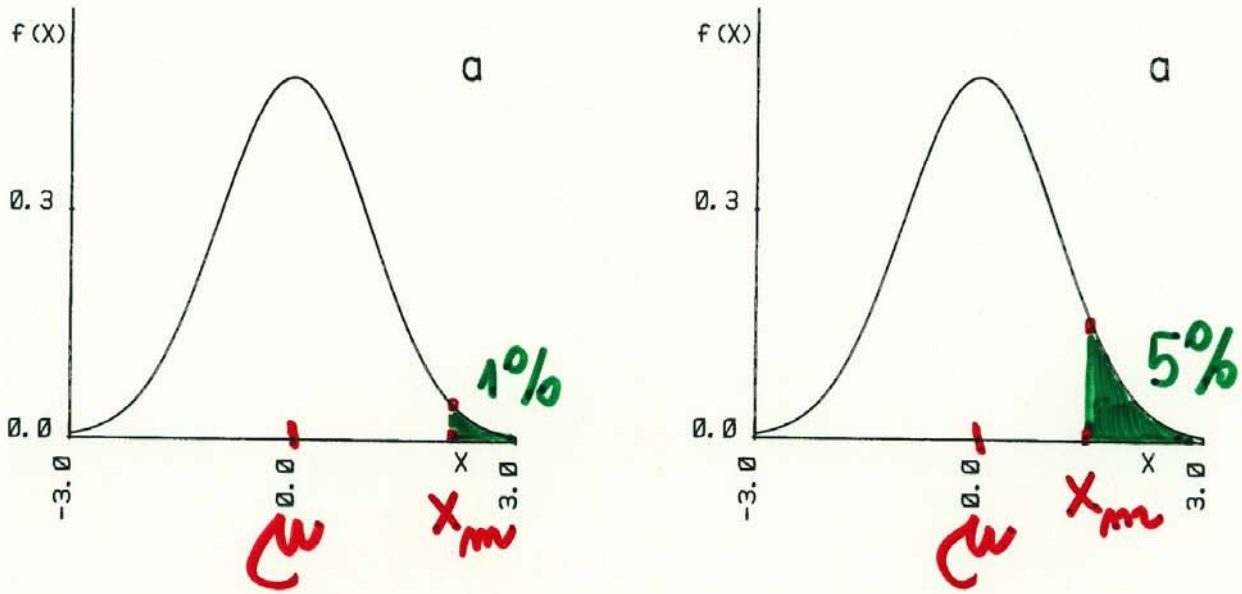
a pak vyčíslíme pravděpodobost P_v

$$\mathbf{P_v = F(u > u_2) = 1 - F(u_2) = 1 - F(2) = 1 - 0.97725 = 0.02275.}$$

$$\mathbf{P \times 100\ \% = 2.275\ \%}$$

Závěr c): Nad hodnotou x_2 bude ležet celkem 2.275 % všech měření.

d) **Výpočet mezní hodnoty x_m, kterou přesáhne pouze 100P % měření:**



V řadě měření se nachází **mezní hodnota** x_m, která se významně odlišuje od střední hodnoty.

Mezní hodnotu x_m přesáhne pouze 100 × P % všech měření.

Mezní hodnota x_m se vyjádří z normované veličiny $u_m = (x_m - \mu)/\sigma$ dle

$$\mathbf{x_m = \mu + \sigma \times F^{-1}(1 - P),}$$

kde symbol $F^{-1}(.)$ označuje kvantilovou funkci, která je inverzní k funkci distribuční a je tabelována.

1. Vypočteme o jakou hodnotu se bude lišit **1 měření (nejmenší nebo největší prvek výběru) od střední hodnoty**, když bylo změřeno právě 100 hodnot:

   1 měření ze 100 značí 1% čili $P = 0.01$ a $F(u) = 1 - P = 0.99$.

   Z tabulek nalezneme hodnotu u pro 0.99, tj. $F^{-1}(0.99) = 2.33$, a tím bude

   $$u_m = (x_m - \mu)/\sigma = 2.33 \quad \text{a úpravou} \quad \mathbf{x_m = \mu + 2.33\sigma}\ .$$

2. Analogicky najdeme i o jakou hodnotu se bude lišit **5 nejodlehlejších měření od střední hodnoty**, když bylo změřeno právě 100 hodnot:

   5 měření ze 100 značí 5% čili $P = 0.05$ a $F(u) = 1 - P = 0.95$.

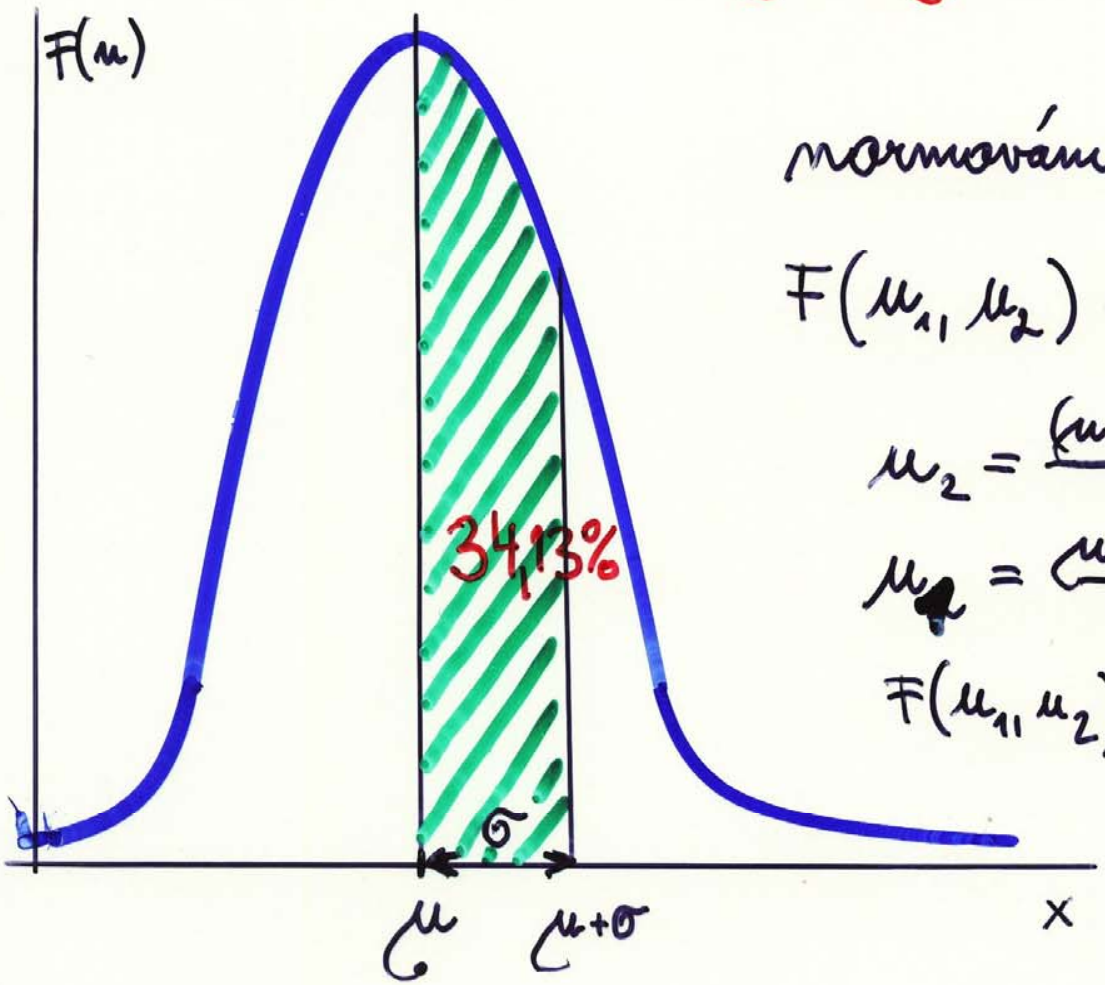
   Z tabulek nalezneme hodnotu u pro 0.95, tj. $F^{-1}(0.95) = 1.645$, a tím bude

   $$u_m = (x_m - \mu)/\sigma = 1.645 \quad \text{a úpravou} \quad \mathbf{x_m = \mu + 1.645\sigma}\ .$$

Závěr d): Mezní hodnotu $x_m = \mu + 2.33\sigma$ přesáhne pouze 1 % ze všech měření zatímco $x_m = \mu + 1.645\sigma$ pak 5% všech měření.

RELATIVNÍ PODÍL NAMĚŘENÝCH HODNOT VÝBĚRU, KTERÉ LEŽÍ

(1) V INTERVALU $\langle \mu ; \mu + \sigma \rangle$:



normování $u = \dfrac{x - \mu}{\sigma}$

$F(u_1, u_2) = F(u_2) - F(u_1)$ když $u_2 > u_1$

$u_2 = \dfrac{(\mu + \sigma) - \mu}{\sigma} = 1$; $F(1) = 0.8413$

$u_1 = \dfrac{\mu - \mu}{\sigma} = 0$; $F(0) = 0.5000$

$F(u_1, u_2) = 0.8413 - 0.5000 = 0.3413$

34,13% měření leží v intervalu $\langle \mu ; \mu + \sigma \rangle$

(2) V INTERVALU $\langle \mu - \sigma ; \mu \rangle$: také 34,13%

(3) V INTERVALU $\langle \mu - \sigma ; \mu + \sigma \rangle$:

$$u_2 = \frac{\mu + \sigma - \mu}{\sigma} = 1 \; ; \; F(1) = 0.8413$$

$$u_1 = \frac{\mu - \sigma - \mu}{\sigma} = -1 \; ; \; F(-1) = 1 - F(1) = 1 - 0.8413$$

$$F(-1; +1) = 0.8413 - 1 + 0.8413 = 0.6826 \text{ , t.j. } 68{,}26\%$$

(4) V INTERVALU $\langle \mu + 2\sigma \rangle$ nebo $\langle \mu - 2\sigma ; \mu \rangle$:

$$\left.\begin{array}{l} u_2 = (\mu + 2\sigma - \mu)/\sigma = 2 \\ u_1 = (\mu - \mu)/\sigma = 0 \end{array}\right\} F(0; 2) = F(2) - F(0) = 0.9773 - 0.5000 = 0.4773$$

t.j. 47,73%

$\langle \mu - 2\sigma ; \mu + 2\sigma \rangle$ t.j. $2 \times 47{,}73 = 95{,}46\%$

Příklad 3.4 *Vyšetření hmotnosti tablet aspirinu*

Byla vyšetřována hmotnost tablety u výběru n = 156 tablet aspirinu, o předepsané hmotnosti tablety 330 mg.

(a) Zvážením všech tablet vyšlo $\bar{x}$ = 330.43 mg a s^2 = 2.32.

(b) Z tohoto výběru bylo vybráno n_A = 32 tablet, pro které vyšlo $\bar{x}_A$ = 330.6 mg a s_A^2 = 2.135,

(c) Při výběru pouze 11 tablet ze 156 vyšlo $\bar{x}_B$ = 330.7 mg a s_B^2 = 2.05.

Stanovte 95%ní interval spolehlivosti střední hodnoty a rozptylu za předpokladu, že rozdělení hmotností tablet je normální.

Data: 1. Základní výběr o rozsahu n = 156.

328.99 329.75 331.62 333.08 330.61 331.35 328.42 330.63 332.17 330.15 331.28 330.92 329.36
329.62 329.61 329.17 330.39 333.47 330.59 330.52 329.49 329.01 331.63 330.64 330.85 326.06
329.92 330.66 328.57 331.45 331.54 332.20 329.43 327.76 334.06 331.25 328.43 330.57 329.68
330.27 328.81 332.26 332.60 327.32 331.28 330.92 332.66 329.88 329.84 329.92 329.32 333.37
330.28 330.78 333.19 330.84 330.70 329.73 328.87 331.71 329.76 329.82 330.59 328.57 332.20
328.03 330.28 331.02 330.58 333.35 329.86 331.22 329.99 330.34 331.85 332.88 331.99 330.02
328.14 330.03 330.10 330.03 330.47 330.62 331.78 329.33 330.16 329.46 331.89 330.65 329.35
331.84 330.31 331.31 328.06 332.59 327.57 329.10 331.61 331.69 329.47 332.09 330.45 329.41
331.78 330.50 330.23 329.89 331.53 331.49 330.52 329.59 334.53 329.04 330.88 330.08 330.11
331.38 331.85 328.51 328.56 332.26 330.98 330.91 330.18 325.47 330.99 330.54 329.74 332.55
329.70 328.99 330.63 330.69 331.00 329.29 328.02 330.16 333.56 331.72 325.47 330.72 331.93
329.23 327.87 331.83 330.58 330.94 331.51 330.00 331.21 331.23 330.57 329.59 327.88 328.86

2. Výběr ze základního výběru, rozsah výběru n_A = 32.

328.99 329.75 331.62 333.08 333.61 331.35 328.42 330.63 332.17 330.15 331.28 330.92 329.36
329.62 329.61 329.17 330.39 333.47 330.59 330.52 329.49 329.01 331.63 330.64 330.85 326.06
329.92 330.66 328.57 331.45 331.54 332.20

3. Výběr ze základního výběru, rozsah výběru n_B = 10

328.99 329.75 331.62 333.08 330.61 331.35 328.42 330.63 332.17 330.15

Řešení: Pro 95%ní interval spolehlivosti střední hodnoty se určí:

pro celý základní výběr interval $330.19 \leq \mu \leq 330.67$,

pro výběr o rozsahu $n_A = 32$ interval $329.96 \leq \mu \leq 331.09$,

pro výběr o rozsahu $n_B = 10$ interval $329.65 \leq \mu \leq 331.70$.

Pro 95%ní interval spolehlivosti rozptylu vyjde

pro celý základní výběr interval $1.879 \leq \sigma^2 \leq 2.937$,

pro výběr o rozsahu $n_A = 32$ interval $1.372 \leq \sigma^2 \leq 3.43$,

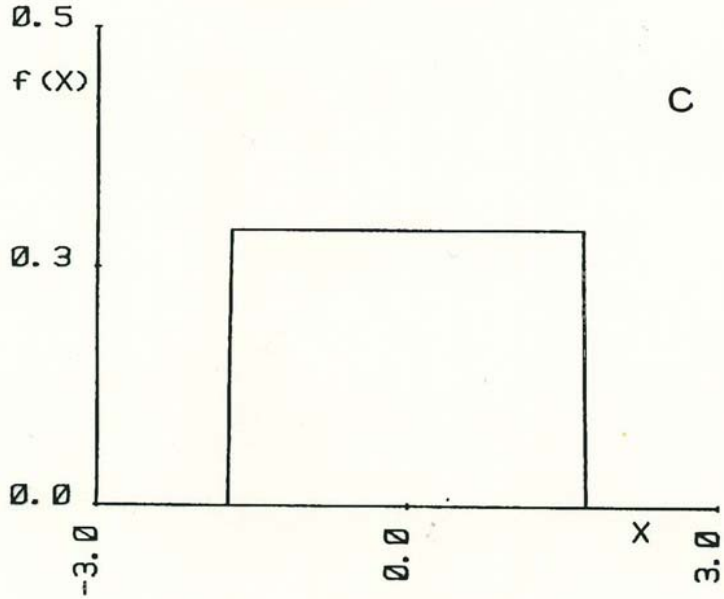
pro výběr o rozsahu $n_B = 10$ interval $0.970 \leq \sigma^2 \leq 6.83$.

Závěr: Pro malé rozsahy výběru z normálního rozdělení mohou vyjít zavádějící výsledky. Pro výběr o velikosti $n = 10$ však již 95 %ní interval spolehlivosti obsahuje předpokládané hodnoty $\mu = 331$ a $\sigma^2 = 1$.

Rovnoměrné rozdělení

Nejjednodušší typ rozdělení pro oboustranně omezenou náhodnou veličinu, která musí ležet v zadaném intervalu

$$a - h \leq x \leq a + h.$$



Hustota pravděpodobnosti
rovnoměrného rozdělení R(0, 1)

Týká se: náhodných veličin, které se v daném intervalu vyskytují se stejnou pravděpodobností.

Pokud je $a = 0$ a $h = 0.5 \cdot 10^{-K}$, popisuje rovnoměrné rozdělení chybu, vzniklou zaokrouhlením na K desetinných míst.

Hustota pravděpodobnosti rovnoměrného rozdělení má tvar

$$f(x) = \frac{1}{2h} \qquad a - h \leq x \leq a + h$$

Střední hodnota je $E(x) = a$,

Rozptyl $D(x) = h^2/3$,

Šikmost $g_1 = 0$,

Špičatost $g_2 = 1.8$.

Logaritmus věrohodnostní funkce má tvar

$$\ln L = -n \ln(2h)$$

pro $a - h \leq \min(x_1, \ldots, x_n) \leq \max(x_1, \ldots, x_n) \leq a + h$.

Vztah nabývá maxima při minimální velikosti h.

Je zřejmé, že $\min(x_1, \ldots, x_n) = x_{(1)}$ a $\max(x_1, \ldots, x_n) = x_{(n)}$.

Maximálně věrohodný **odhad parametru rozptýlení** $\hat{h}$ je roven

$$\hat{h} = 0.5\,(x_{(n)} - x_{(1)})$$

a

maximálně věrohodný **odhad parametru polohy a** je roven

$$\hat{a} = 0.5\,(x_{(n)} + x_{(1)})$$

Odhad $\hat{h}$ je vychýlený. **Nevychýlený odhad** $\hat{h}_0$ se získá násobením odhadu $\hat{h}$ faktorem $(n + 1)/(n - 1)$.

Pro **rozptyly těchto odhadů** platí vztahy

$$D(\hat{h}) = \frac{2h^2}{(n-1)(n+2)}$$

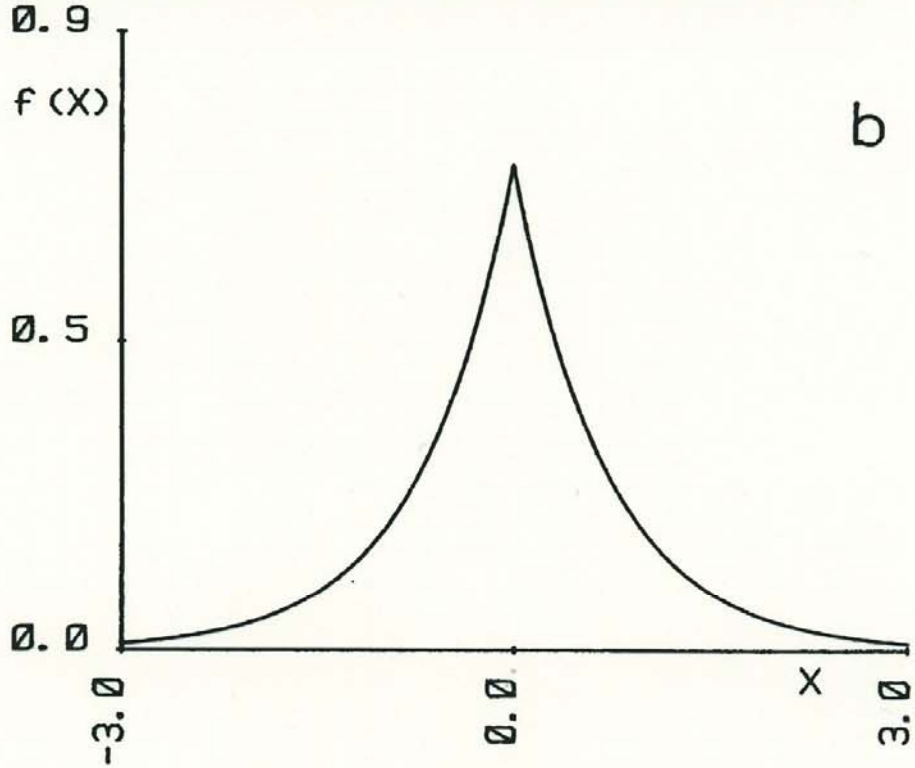
a

$$D(\hat{a}) = \frac{2h^2}{(n+1)(n+2)}$$

Odhady $\hat{a}$, $\hat{h}$ jsou nekorelované, ale nikoliv nezávislé.

Laplaceovo rozdělení
(oboustranné exponenciální)

Vyskytuje se: náhodné veličiny jsou měřeny za podmínek kolísání rozptylu kolem určité střední hodnoty.



Hustota pravděpodobnosti
Laplaceova rozdělení

Hustota pravděpodobnosti spojité náhodné veličiny x ležící v intervalu (- ∞, ∞) s Laplaceovým rozdělením má tvar

$$f(x) = 0.5\,\Phi^{-1} \exp\left(- \frac{|x - \Theta|}{\Phi} \right)$$

Střední hodnota je $E(x) = \Theta$,
Rozptyl $D(x) = 2\,\Phi^2$,
Šikmost $g_1 = 0$,
Špičatost $g_2 = 6$.

Ve srovnání s normálním rozdělením:

- je špičatější a má delší konce.
- 1%ní kvantil: $E(x) - 2.72\sqrt{D(x)}$ (Laplaceovo rozdělení),
  $E(x) - 2.33\sqrt{D(x)}$ (Gaussovo rozdělení).

Po zlogaritmování vyjde logaritmus věrohodnostní funkce

$$\ln L = -n \ln(2\Phi) - \Phi^{-1} \sum_{i=1}^{n} |x_i - \Theta|$$

Výběrový medián $\hat{\Theta} = \tilde{x}_{0.5}$

$$\hat{\Phi} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \Theta|$$

Přestože je medián $\tilde{x}_{0.5}$ nevychýlený odhad, nemá pro malé výběry minimální rozptyl

$$D(\hat{\Phi}) = \frac{\Phi^2}{n}$$

100(1 - α)%ní interval spolehlivosti pro Φ se vypočte podle

$$\frac{2\,n\,\hat{\Phi}}{\chi^2_{1-\alpha/2}(2\,n)} \le \Phi \le \frac{2\,n\,\hat{\Phi}}{\chi^2_{\alpha/2}(2\,n)}$$

Příklad 3.5 *Intervaly spolehlivosti Laplaceova rozdělení*

Z náhodného výběru velikosti n = 50 z Laplaceova rozdělení L(0, 2) byl určen odhad $\Theta = \tilde{x}_{0.5} = 0.0119$ a $\Phi = 1.0596$. Stanovte 95%ní interval spolehlivosti obou parametrů.

Řešení: Pro odhad rozptylu platí, že $s^2 = 2\,\Phi^2 = 2.246$ a po dosazení do prvního vztahu vyjde 95%ní interval spolehlivosti mediánu $-0.282 \le \hat{\Theta} \le 0.306$.

$$\hat{\Phi} - u_{1-\alpha/2}\frac{\hat{\Phi}}{\sqrt{n}} \le \Phi \le \hat{\Phi} + u_{1-\alpha/2}\frac{\hat{\Phi}}{\sqrt{n}}$$

a po dosazení vyjde 95%ní interval spolehlivosti $0.765 \le \hat{\Theta} \le 1.353$.

Je-li $\Theta = \tilde{x}_{0.5} = 0$, lze určit odhad Φ ze vztahu $\Phi = \frac{1}{n}\sum_{i=1}^{n}|x_i| = 1.0596$.

Dosazením vyjde 95%ní interval spolehlivosti $0.818 \le \hat{\Theta} \le 1.428$.

Závěr: Intervaly spolehlivosti parametru Φ vyčíslené dle různých rovnic se významně neliší, protože je střední hodnota $E(x) = \Theta$ rovna nule.

Exponenciální rozdělení

- Je jednostranně ohraničené zdola, má uplynulý čas, resp. obsazený prosto před tím, než nastal náhodný jev.

- Je typické pro životnost součástí strojů, vzdálenost, kterou urazí molekuly plynu při nízkém tlaku až do vzájemné srážky, doby mezi dopadem částic do čítače a doby bezporuchové činnosti.

Jednoparametrové exponenciální rozdělení

Popisuje statistické chování kladné náhodné veličiny pro $x \geq 0$.

Hustota pravděpodobnosti je definována vztahem

$$f(x) = \Theta^{-1} \exp\left(- \frac{x}{\Theta} \right)$$

Střední hodnota je $E(x) = \Theta$,

Rozptyl $D(x) = \Theta^2$,

Šikmost $g_1 = 2$,

Špičatost $g_2 = 9$.

Medián je roven $\tilde{x}_{0.5} = \Theta \ln 2$.

Logaritmus věrohodnostní funkce má tvar

$$\ln L = -n \ln \Theta - \sum_{i=1}^{n} \frac{x_i}{\Theta}$$

Maximálně věrohodný **odhad polohy** $\Theta = \dfrac{\sum_{i=1}^{n} x_i}{n}$

a odpovídající rozptyl $D(\hat{\Theta}) = \dfrac{\Theta^2}{n}$.

100(1-α)%ní interval spolehlivosti se vypočte

$$\frac{2\,n\,\hat{\Theta}}{\chi^2_{1-\alpha/2}(2\,n)} \leq \Theta \leq \frac{2\,n\,\hat{\Theta}}{\chi^2_{\alpha/2}(2\,n)}$$

Příklad 3.7 *Kinetika rozkladu terciárního dibutylperoxidu (DTBP)*

Při 154.6 °C dochází v plynné fázi k rozkladu DTBP reakcí prvního řádu s experimentálně určenou rychlostní konstantou $k = 3.46 \cdot 10^{-4}\ s^{-1}$. Stanovte dobu, za kterou dojde k rozkladu poloviny množství DTBP.

Řešení:

Počet nerozložených molekul N_t v čase t je roven

$$N_t = N_0 \exp(-k\,t),$$

kde N_0 je původní počet molekul DTBP.

Pravděpodobnost "přežití", tj. nerozložení, jedné molekuly DTBP v časovém intervalu $t \le T \le t + dt$ je rovna

$$P(t \le T \le t + dt) = -\frac{dN_t}{N_0} = k \exp(-k\,t)\,dt$$

Hustota pravděpodobnosti časů přežití molekul je dána rovnicí, kde $\Theta = 1/k$. Střední hodnota času přežití je $E(x) = 2.89 \cdot 10^3$ s.

Závěr: Doba, za kterou dojde k rozkladu právě poloviny množství DTBP, pak odpovídá mediánu, $\tilde{x}_{0.5} = \hat{\Theta} \ln 2 = 2 \cdot 10^3$ s.

Dvouparametrové exponenciální rozdělení

Popisuje: chování náhodné veličiny, která může nabývat hodnot $x \geq \mu$, tj. je zdola ohraničená.

Hustota pravděpodobnosti má tvar

$$f(x) = \Theta^{-1} \exp\left[-\frac{x - \mu}{\Theta}\right]$$

Střední hodnota je $E(x) = \mu + \Theta$,
Rozptyl $D(x) = \Theta^2$,
Šikmost $g_1 = 2$,
Špičatost $g_2 = 9$.
Logaritmus věrohodnostní funkce má tvar

$$\ln L = n \ln \Theta - \sum_{i=1}^{n} \frac{x_i - \mu}{\Theta}$$

Odhad $\hat{\mu}$

$$\hat{\mu} = x_{(1)} = \min(x_1, \ldots, x_n)$$

Maximálně věrohodný **odhad** $\hat{\Theta}$ parametru Θ se vypočte

$$\hat{\Theta} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}) \approx \bar{x} - x_{(1)}$$

Odhad $\hat{\mu}$ ***má střední hodnotu*** $E(\hat{\mu}) = \mu + \frac{\Theta}{n}$

a ***rozptyl*** $D(\hat{\mu}) = \frac{\Theta^2}{n^2}$

Odhad $\hat{\Theta}$ ***má střední hodnotu*** $E(\hat{\Theta}) = \Theta\left(1 - \frac{1}{n}\right)$

a ***rozptyl*** $D(\hat{\Theta}) = \Theta^2\left(\frac{1}{n} + \frac{1}{n^2} - \frac{2}{n^3}\right)$

Maximálně věrohodné odhady $\hat{\Theta}$ a $\hat{\mu}$ jsou vychýlené.

Nevychýlené odhady $\hat{\Theta}_0$ a $\hat{\mu}_0$ se vypočtou ze vztahů

$$\hat{\mu}_0 = \frac{n\,x_{(1)} - \bar{x}}{n - 1}$$

$$D(\mu_0) = \frac{\Theta^2}{n\,(n - 1)}$$

$$\hat{\Theta}_0 = \frac{n\,(\bar{x} - x_{(1)})}{n - 1}$$

$$D(\hat{\Theta}_0) = \frac{\Theta^2}{n - 1}$$

Odhady $\hat{\mu}_0$, Θ_0 jsou však korelované s korelačním koeficientem rovným $(-1/\sqrt{n})$.

100(1 - α)%ní oboustranný int. spolehlivosti parametru Θ se vypočte

$$\frac{2\ (n-1)\ \hat{\Theta}_0}{\chi^2_{1-\alpha/2}(2n-2)} \leq \Theta \leq \frac{2\ (n-1)\ \hat{\Theta}_0}{\chi^2_{\alpha/2}(2n-2)}$$

Spodní mez μ_1 je pro 100(1 - α)%ní interval spolehlivosti parametru μ vyjádřena vztahem

$$\mu_1 = x_{(1)} - \frac{\hat{\Theta}_0\ F_{1-\alpha}(2,\ 2n - 2)}{n}$$

Horní mez je s pravděpodobností blízkou jedné nejmenší prvek výběru $x_{(1)}$.

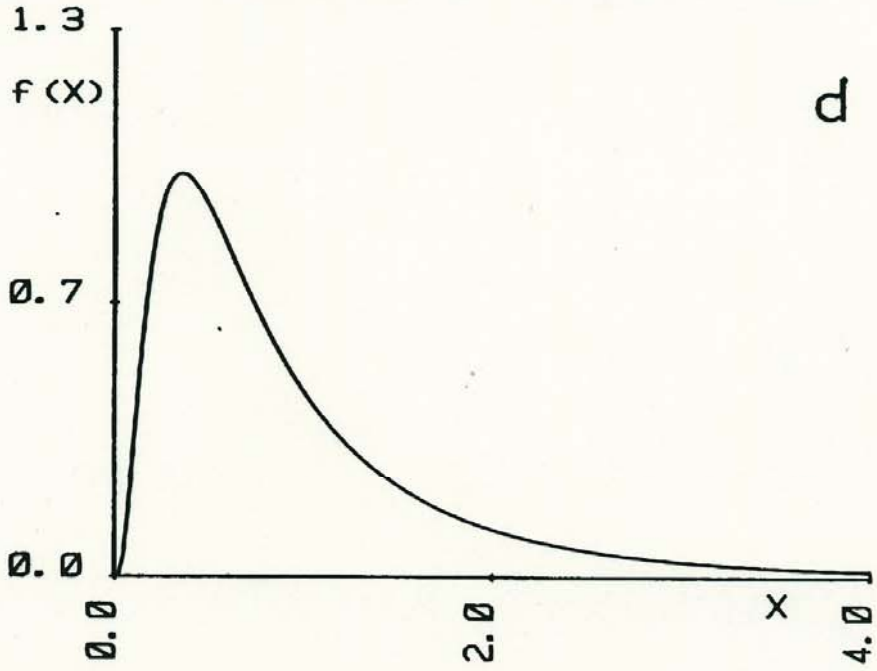
Pro určení kvantilů rozdělení F(2, 2n - 2) stačí dosadit do vztahu

$$F_P(2,\ 2n - 2) = (n - 1)\ [(1 - P)^{-\frac{1}{n-1}} - 1]$$

Logaritmicko-normální rozdělení

Nejrozšířenější alternativou Gauss. rozdělení pro jednostranně ohraničená data: ***fyzikální veličiny*** (teplota, tlak, objem, hmotnost, koncentrace) jsou buď kladné, nebo mají definovaný počátek (např. absolutní nula u teploty).

Používá se: kde se měří nízké koncentrace, malé hmotnosti, malé délky, v analytické chemii stopová analýza, velikosti prachových částic v atmosféře, distribuce velikosti práškových pigmentů.



Hustota pravděpodobnosti log.-normálního rozdělení

Transformace: náhodná veličina x souvisí s lognormálním rozdělením s náhodnou veličinou u s normovaným normálním rozdělením vztahem

$$u = \frac{\ln(x - \Theta) - \mu}{\sigma}$$, kde μ, σ, Θ jsou parametry.

Hustota pravděpodobnosti lognormálního rozdělení je

$$f(x) = \frac{1}{(x - \Theta)\,\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x - \Theta) - \mu)^2}{2\sigma^2}\right]$$

kde μ, σ, Θ jsou parametry polohy a rozptýlení.

Dvouparametrové lognormální rozdělení

Náhodná veličina, která může nabývat pouze kladných hodnot, tj. leží v intervalu $0 \leq x < \infty$.

Střední hodnota je $E(x) = \exp(\mu + 0.5\,\sigma^2)$

a **rozptyl náhodné veličiny x** se vyčíslí

$$D(x) = \exp(2\,\mu)\,\omega\,(\omega - 1)\ , \text{ kde } \omega = \exp(\sigma^2),$$

Šikmost g_1 je $g_1 = \sqrt{\omega - 1}\,(\omega + 2)$,

Špičatost g_2 je $g_2 = \omega^4 + 2\,\omega^3 + 3\,\omega^2 - 3$,

Variační koeficient δ je $\delta = \sqrt{\omega - 1}$,

Módus x_M se vyjádří $x_M = \exp(\mu - \sigma^2)$,

Medián med je $\text{med} = \exp(\mu)$

Logaritmus věrohodnostní funkce dvouparametrového lognormálního rozdělení má tvar

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\ln x_i -$$

$$- \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2$$

Maximálně věrohodný **odhad parametru polohy μ** je

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

Maximálně věrohodný **odhad parametru rozptýlení** σ^2 je

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2$$

který je však vychýlený.

Nevychýlený odhad $\hat{\sigma}_0^2$ se vyčíslí analogicky jako u normál-ního rozdělení $\hat{\sigma}_0^2 = \frac{n}{n-1} \hat{\sigma}^2$

100(1 - α)%ní oboustranný interval spolehlivosti mediánu se vypočte

$$\exp\left[\hat{\mu} - t_{1-\alpha/2}(n-1)\,\frac{\hat{\sigma}}{\sqrt{n}}\right] \le \text{med} \le \exp\left[\hat{\mu} + t_{1-\alpha/2}(n-1)\,\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

100(1 - α)%ní oboustranný interval spolehlivosti variačního koeficientu δ se vypočte

$$\sqrt{\exp\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(n-1)} - 1} \le \delta \le \sqrt{\exp\frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2}(n-1)} - 1}$$

Odhad **střední hodnoty původních dat M = E(x)** se vypočte

$$\hat{M} = \exp(\hat{\mu})\; g(0.5\,\sigma^2)$$

a odpovídající **rozptyl původních dat V** = D(x)

$$\hat{V} = \exp(2\,\hat{\mu}) \left[g\,(2\,\sigma^2) - g\left(\frac{(n-2)\,\hat{\sigma}^2}{n-1} \right) \right]$$

V obou vztazích je funkce g(t) vyjádřena nekonečnou řadou

$$g(t) = 1 + \frac{n-1}{n}\,t + \sum_{j=2}^{\infty} \frac{(n-1)^{2j-1}}{n^j\,(n+1)\,(n+3)\,\ldots\,(n+2j-1)}\,\frac{t^j}{j!}$$

Pro větší rozsahy výběru, n > 50, nebo pro dostatečně malé rozptyly σ^2 je možné použít aproximaci

$$g(t) \approx \exp(t) \left[1 - \frac{t\,(t+1)}{n} + \frac{t^2\,(3\,t^2 + 22\,t + 21)}{6\,n^2} \right]$$

Rozptyly odhadů $D(\hat{M})$ a $D(\hat{V})$ se vypočtou

$$D(\hat{M}) \approx \sigma^2 \exp(2\mu)\, \omega \frac{1 + 0.5\,\sigma^2}{n}$$

$$D(\hat{V}) \approx 2\,\sigma^2 \exp(\mu)\, \omega^2 \left[2\,(\omega - 1)^2 + \sigma^2 (2\,\omega - 1)^2\right]$$

Na základě odhadu $\hat{\tau}$ a jeho rozptylu $D(\hat{\tau})$, pro které platí

$$\hat{\tau} = \hat{\mu} + \frac{n\,\hat{\sigma}^2}{2\,(n - 1)}, \qquad D(\hat{\tau}) = \frac{\sigma^2}{n - 1} + \frac{n^2\,\sigma^4}{4\,(n - 1)^3}$$

se konstruuje pro výběry větších rozsahů **100(1 - α)%ní interval spolehlivosti střední hodnoty M** dle vztahu

$$\exp(\hat{\tau} - u_{1-\alpha/2} \sqrt{D(\hat{\tau})}) \le M \le \exp(\hat{\tau} + u_{1-\alpha/2} \sqrt{D(\hat{\tau})})$$

Tříparametrové lognormální rozdělení

Náhodná veličina může nabývat hodnot vyšších než spodní mez Θ, tj. leží v intervalu $\Theta \leq x < \infty$.

Náhodná veličina x má tříparametrové lognormální rozdělení, pokud má náhodná veličina $\ln(x - \Theta)$ normální rozdělení $N(\mu, \sigma^2)$.

Pro parametry polohy tříparametrového lognormálního rozdělení platí, že jsou o Θ vyšší než odpovídající parametry dvouparametrového lognormálního rozdělení.

Parametry rozptýlení, šikmost a špičatost jsou shodné.

Logaritmus věrohodnostní funkce má tvar

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(x_i - \Theta) -$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[\ln(x_i - \Theta) - \mu]^2$$

Pro dané Θ jsou maximálně věrohodné odhady $\hat{\mu}(\Theta)$ a $\hat{\sigma}^2(\Theta)$ určeny přímo při dosazení $(x_i - \Theta)$ místo hodnot x_i.

Pro $\Theta \to x_{(1)}$ roste maximum věrohodnostní funkce nade všechny meze.

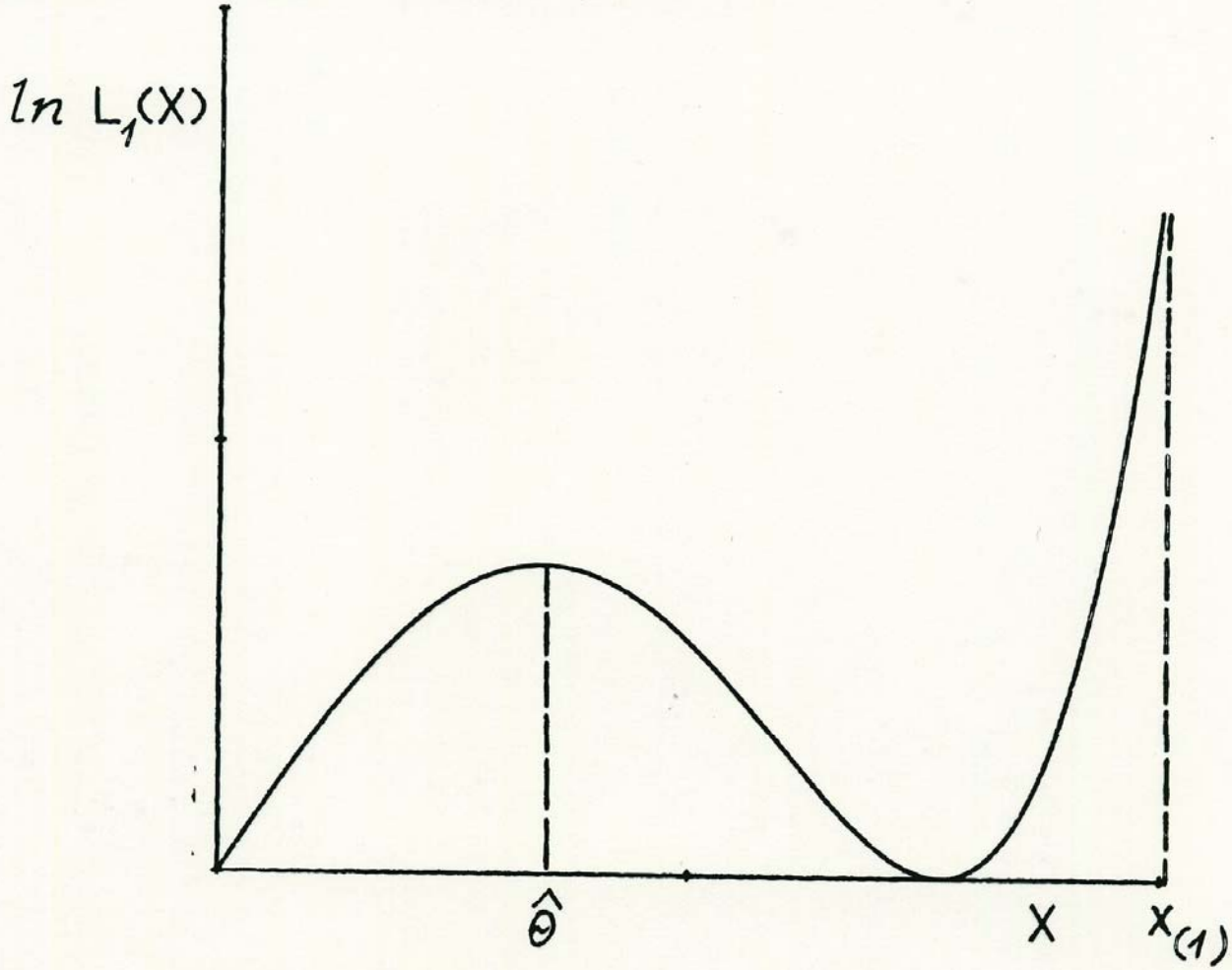
Navíc je nutné, aby $\Theta \leq x_{(1)}$.

Při numerické maximalizaci ln L je třeba zajistit, aby nebyl vyhledán tento globální extrém.

Výsledná modifikovaná věrohodnostní funkce má tvar

$$\ln L_1 = -n\,[\hat{\mu}(\Theta) + 0.5 \ln \hat{\sigma}(\Theta)]$$

Závisí pouze na jednom parametru Θ.



Průběh modifikované věrohodnostní funkce

Poissonovo rozdělení

Náhodná veličina vyjadřuje ***počet výskytů řídkých, málo pravděpodobných jevů*** v určitém časovém nebo objemovém intervalu,

např. počet výskytů náhodného jevu A v čas. intervalu Δt závisí na

a) pravděpodobnost výskytu A je úměrná délce Δt, přičemž pravděpodobnost výskytu více než jednoho jevu v tomto úseku je zanedbatelná;

b) pravděpodobnost výskytu A v daném časovém intervalu Δt je nezávislá na výskytu jevu A v předcházejícím intervalu Δt;

c) pravděpodobnost výskytu A v časovém intervalu Δt nezávisí na poloze počátku časového intervalu.

V praxi: počet α částic, které dopadnou za dobu Δt na plochu; výskyt poruch v plošných nebo objemových elementech při strukturních analýzách, počet kolonií bakterií na Petriho misce.

Pravděpodobnostní funkce je definována

$$p(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad \lambda > 0$$

kde x je diskrétní náhodná veličina, nabývající pouze celočíselných hodnot 0, 1, 2, ..., k, a λ je parametr.

Střední hodnota $E(x) = \lambda$,

Rozptyl $D(x) = \lambda$..

Data: pro různé hodnoty x, (x = 0, 1, ..., k) se určuje vždy počet prvků výběru, n_x, nabývající velikosti x.

Věrohodnostní funkce bude ve tvaru

$$L(\lambda) = \prod_{x=0}^{k} \lambda^{n_x x} \frac{\exp(-n_x \lambda)}{n_x x!}$$

Po zlogaritmování a derivování podle λ resultuje rovnice

$$\frac{d \ln L(\lambda)}{d \lambda} = \sum_{x=0}^{k} \left(\frac{n_x x}{\lambda} - n_x \right) = 0$$

a z ní bude max. věrohodný **odhad parametru polohy** λ

$$\hat{\lambda} = \frac{\sum_{x=0}^{k} n_x x}{\sum_{x=0}^{k} n_x} = \frac{\sum_{x=0}^{k} n_x x}{n}$$

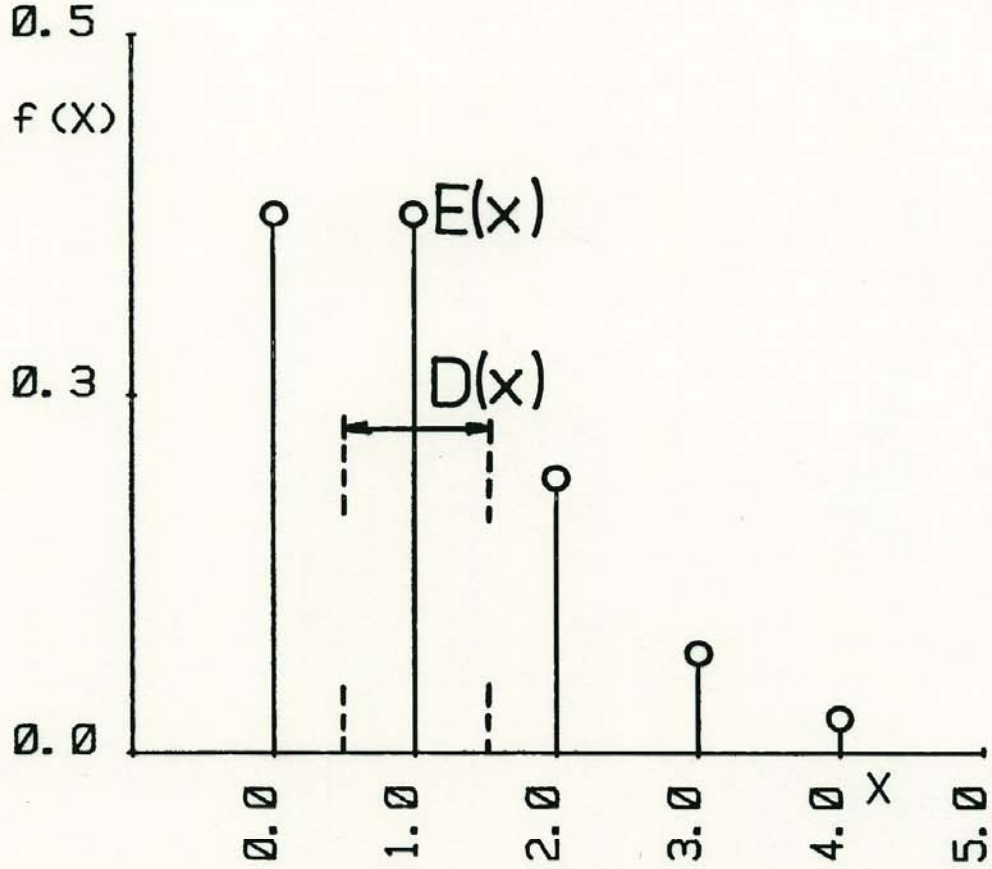
kde n je celkový rozsah výběru. Odhad parametru polohy $\hat{\lambda}$ zde odpovídá váženému aritmetickému průměru.

K určení rozptylu odhadu $\hat{\lambda}$ je třeba derivovat rovnici ještě jednou

$$\frac{d^2 \ln L(\lambda)}{d\lambda^2} = -\frac{1}{\lambda^2} \sum_{x=0}^{k} x\, n_x ,$$

a protože $E(\sum_{x=0}^{k} x\, n_x) \approx n\,\lambda$, vyjde **odhad rozptylu** λ

$$D(\hat{\lambda}) = \frac{\lambda}{n} .$$



Pravděpodobnostní funkce Poissonova rozdělení pro $\lambda = 1$

Meze λ_1 a λ_2 přibližného **100(1 - α)%ního int. spolehlivosti parametru** λ určit jako kořeny kvadratické rovnice

$$\lambda^2 - \left[2\,\hat{\lambda} + \frac{\chi^2_{1-\alpha}(1)}{n}\right]\lambda + \hat{\lambda}^2 = 0 \quad \text{vzhledem k } \lambda..$$

Asymptotický **interval spolehlivosti parametru** λ je

$$\hat{\lambda} - u_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + u_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}$$

Přibližný **100(1 - α)%ní interval spolehlivosti parametru** λ je

$$\frac{\chi^2_{\alpha/2}(2\,\hat{\lambda}\,n)}{2\,n} \leq \lambda \leq \frac{\chi^2_{1-\alpha/2}(2\,\hat{\lambda}\,n + 2)}{2\,n}$$

Příklad 3.3 *Int. spolehlivosti průměrného počtu částic kosmického záření*
Byly zaznamenány počty částic, emitovaných kosmickým zářením, v intervalech 0.1 s. Z 201 intervalů byl určen průměrný počet částic $\hat{\lambda} = 10.5$. Určete 95%ní interval spolehlivosti, ve kterém se nalézá skutečný počet částic v časových intervalech 0.1 s.

Data: počet částic k a odpovídající četnost výskytu počtu částic n_k ve výběru o rozsahu n = 200.

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_k	0	0	0	2	1	11	12	12	20	22	17	30	20	20	10

k	15	16	17	18	19	20	21	22	23
n_k	18	6	4	3	1	0	0	1	0

Řešení: K vyčíslení intervalu spolehlivosti parametru λ užijeme různých vztahů:

1. Dosazením za n = 200, $\chi^2_{0.95}(1) = 3.842$ a $\hat{\lambda} = 10.5$ dostaneme $\lambda^2 - 21.0192\lambda + 110.25 = 0$. Řešením jsou dva kořeny $\lambda_1 = 10.06$ a $\lambda_2 = 10.96$, takže 95%ní interval spolehlivosti bude $10.06 \leq \lambda \leq 10.96$.

2. Jiným dosazením bude 95%ní interval spolehlivosti $10.05 \leq \lambda \leq 10.95$.

3. Jelikož z tabulek χ^2 rozdělení nelze nalézt hodnotu $\chi^2_{0.025}(4200)$ a $\chi^2_{0.975}(4600)$, užijeme aproximativní formuli Wilson-Wilfertyové

$$\chi^2_P(v) \approx v\left(1 - \left(\frac{2}{9\,v} + u_P\sqrt{\frac{2}{9\,v}}\right)^3\right)$$

kde u_P je 100P%ní kvantil normovaného normálního rozdělení.

Pro v = 4200, $u_{0.025} = -1.96$ bude $\chi^2_{0.025}(4200) = 4022.26$ a při dosazení v = 4202, $u_{0.975} = 1.96$ bude $\chi^2_{0.975}(4202) = 4383.57$ a dosazením do bude 95%ní interval spolehlivosti $10.06 \le \lambda \le 10.96$.

Závěr: Interval spolehlivosti parametru λ vyčíslený podle různých vztahů vyšel prakticky stejný, $10.06 \le \lambda \le 10.96$.