

## 8. Normální rozdělení

**8.1. Definice: Normální (Gaussovo) rozdělení**  $N(\mu, \sigma^2)$  s parametry  $\mu$  a  $\sigma > 0$  je rozdělení určené hustotou

$$(\spadesuit) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in (-\infty, \infty).$$

Rozdělení  $N(0; 1)$  s parametry  $\mu = 0$  a  $\sigma = 1$  se nazývá *normované normální rozdělení*. V dalším textu budeme náhodnou veličinu, která má rozdělení  $N(0; 1)$  obvykle označovat písmenem  $U$ . Její hustota je pak

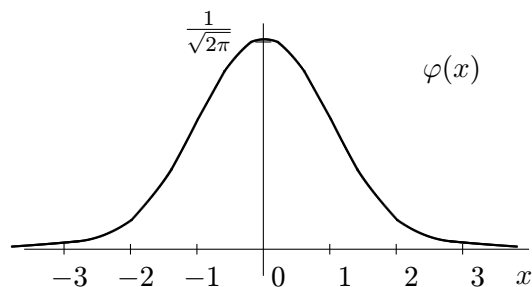
$$(\spadesuit\spadesuit) \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in (-\infty, \infty).$$

Distribuční funkci rozdělení  $N(0; 1)$ , která je definovaná vztahem

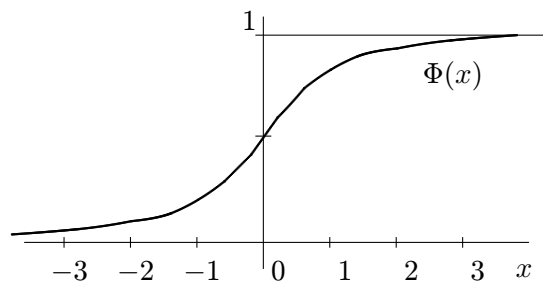
$$(\clubsuit) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in (-\infty, \infty),$$

budeme vždy označovat symbolem  $\Phi$ .

Graf hustoty  $\varphi$  normovaného normálního rozdělení  $N(0; 1)$  znázorníme na obrázku Obr. 8.1 a graf distribuční funkce  $\Phi$  je znázorněn na obrázku Obr. 8.2.



Obr. 8.1



Obr. 8.2

**Poznámka:** Normované normální rozdělení  $N(0; 1)$  je symetrické kolem nuly a tedy  $\varphi(x) = \varphi(-x)$ . Je tudíž  $E(X) = 0$  a dále je  $D(X) = 1$ . Odtud plyne, že pro distribuční funkci  $\Phi$  platí:

$$\Phi(x) + \Phi(-x) = 1 \Rightarrow \Phi(-x) = 1 - \Phi(x) \quad \text{a} \quad \Phi(0) = 0,5.$$

Obecné normální rozdělení  $N(\mu; \sigma^2)$  je posunuté o hodnotu  $\mu$ , je tedy symetrické vzhledem k této hodnotě. Je  $E(X) = \mu$  a  $D(X) = \sigma^2$ . Směrodatná odchylka  $\sigma(X) = \sigma$ . Rozdělení je koncentrováno ke střední hodnotě. I když nabývá náhodná veličina s tímto rozdělením teoreticky všech reálných hodnot je

$$P(|X - \mu| < 3\sigma) = 0,999 \quad \text{a} \quad P(|X - \mu| < 3,5\sigma) = 0,9999.$$

**Poznámka:** Normální rozdělení si zachovává svůj charakter při lineární transformaci. Platí totiž následující tvrzení.

**8.2. Věta:** Jestliže má náhodná veličina  $X$  rozdělení  $N(\mu, \sigma^2)$ , má pak náhodná veličina  $Y = \alpha X + \beta$  rozdělení  $N(\alpha\mu + \beta, \alpha^2\sigma^2)$ . **Speciálně platí**, že náhodná veličina

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

má normované normální rozdělení  $N(0, 1)$ .

*Důkaz:* Je-li  $X$  náhodná veličina, která má rozdělení  $N(\mu, \sigma^2)$  a je-li  $f$  její hustota a  $F$  je její distribuční funkce, pak pro hustotu  $g$  a distribuční funkci  $G$  náhodné veličiny  $Y$  platí:

$$G(y) = P(Y \leq y) = P(\alpha X + \beta \leq y) = P(\alpha X \leq y - \beta) =$$

$$= \begin{cases} P(X \leq \frac{y-\beta}{\alpha}) = F(\frac{y-\beta}{\alpha}), & \alpha > 0 \\ P(X \geq \frac{y-\beta}{\alpha}) = 1 - F(\frac{y-\beta}{\alpha}), & \alpha < 0. \end{cases}$$

Potom pro hustotu  $g$  rozdělení náhodné veličiny  $Y$  dostaneme:

$$\begin{aligned} g(y) = G'(y) &= \left\langle \frac{\frac{d}{dy} \left( F\left(\frac{y-\beta}{\alpha}\right) \right)}{\frac{d}{dy} \left( 1 - F\left(\frac{y-\beta}{\alpha}\right) \right)} = \frac{\frac{1}{\alpha} f\left(\frac{y-\beta}{\alpha}\right)}{\frac{-1}{\alpha} f\left(\frac{y-\beta}{\alpha}\right)} \right\rangle = \frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right) = \\ &= \frac{1}{|\alpha|\sigma\sqrt{2\pi}} e^{-\frac{(\frac{y-\beta}{\alpha}-\mu)^2}{2\sigma^2}} = \frac{1}{|\alpha|\sigma\sqrt{2\pi}} e^{-\frac{(y-(\alpha\mu+\beta))^2}{2(\alpha\sigma)^2}}. \end{aligned}$$

To je ovšem hustota normálního rozdělení  $N(\alpha\mu + \beta; (\alpha\sigma)^2)$ .

Jestliže zvolíme  $\alpha = \frac{1}{\sigma}$  a  $\beta = -\frac{\mu}{\sigma}$  dostaneme  $\alpha\mu + \beta = 0$  a  $\alpha\sigma = 1$ . Náhodná veličina  $U$  má tudíž normované normální rozdělení  $N(0; 1)$ .

**Poznámka: Transformace na normované rozdělení.** Jestliže má náhodná veličina  $X$  normální rozdělení  $N(\mu; \sigma^2)$ , pak má náhodná veličina

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

normované normální rozdělení  $N(0; 1)$ . Pro její distribuční funkci  $F$  a hustotu  $f$  platí:

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

a

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) \Leftrightarrow \sigma f(\sigma u + \mu) = \varphi(u).$$

Je tedy

$$P(a < X < b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right),$$

kde hodnoty funkce  $\Phi$  odečteme z tabulek hodnot distribuční funkce  $\Phi$ .

**8.3. Kvantily normálního rozdělení.** Pro kvantily  $u_p$  normovaného normálního rozdělení  $N(0; 1)$  platí, že:

$$\Phi(u_p) = p, \quad 0 < p < 1 \Leftrightarrow u_p = \Phi^{-1}(p)$$

a jejich hodnoty nalezneme v tabulkách kvantilů. Všimneme si, že platí:

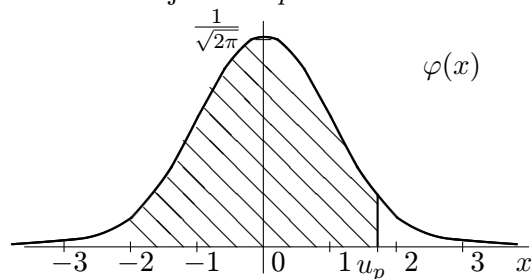
$$u_{0,5} = 0 \quad \text{a} \quad u_{1-p} = -u_p.$$

Pro kvantily  $x_p$  obecného normálního rozdělení  $N(\mu; \sigma^2)$  platí:

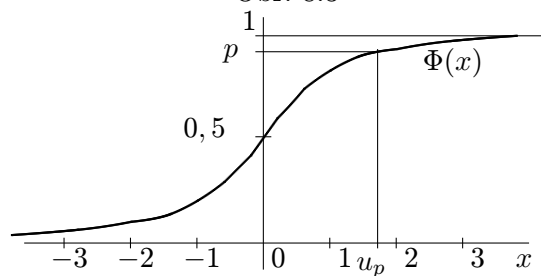
$$F(x_p) = p \Leftrightarrow \Phi\left(\frac{x_p - \mu}{\sigma}\right) = p \Leftrightarrow \frac{x_p - \mu}{\sigma} = u_p \Leftrightarrow x_p = u_p\sigma + \mu.$$

Odtud plyne, že medián  $\tilde{x} = x_{0,5} = \hat{x} = E(X) = \mu$  je zároveň modus.

Význam kvantilů si znázorníme na obrázku hustoty  $\varphi$  a distribuční funkce  $\Phi$  normovaného normálního rozdělení. Obsah obrazce vyznačeného šrafováním je roven  $p$ .



Obr. 8.3



Obr. 8.4

## 8.4. Výpočet distribuční a kvantilové funkce normálního rozdělení.

### I. Přímý výpočet

hustota:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbf{R};$$

pro rozdělení s parametry  $\mu$ ,  $\sigma$  je :

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbf{R};$$

distribuční funkce:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbf{R};$$

pro rozdělení s parametry  $\mu$ ,  $\sigma$  je :

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbf{R};$$

kvantilovou funkci dostaneme jako řešení rovnice:

$$\Phi(u_p) = p, \quad x_p = \mu + u_p \sigma, \quad 0 < p < 1.$$

### II. Pomocí aproximace

distribuční funkce:

$$\text{a) } \Phi(x) = \frac{1}{2} \left( 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right), \quad x \in \mathbf{R},$$

$$\text{kde } \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad x \in \mathbf{R}.$$

b) pro  $x \geq 0$  je:

$$\Phi(x) \doteq 1 - \varphi(x) \sum_{k=1}^5 a_k w^k,$$

kde

$$w = \frac{1}{1 + a_0 x},$$

$$a_0 = 0,231\,641\,9, \quad a_1 = 0,319\,381\,5, \quad a_2 = -0,356\,563\,8,$$

$$a_3 = 1,781\,478, \quad a_4 = -1,821\,256, \quad a_5 = 1,330\,274$$

Pro  $x \leq 0$  je  $\Phi(x) = 1 - \Phi(-x)$ . Pro  $x \geq 6$  je  $\Phi(x) = 1$ . Maximální chyba aproximace je menší než  $10^{-6}$ .

kvantilová funkce:

pro  $0 < p \leq \frac{1}{2}$  je

$$\Phi^{-1}(p) = u_p \doteq -w + \frac{\sum_{k=0}^2 a_k w^k}{\sum_{k=0}^3 b_k w^k},$$

kde

$$w = \sqrt{-2 \ln p}, \quad a_0 = 2,515\,517, \quad a_1 = 0,802\,853, \quad a_2 = 0,010\,328$$

$$b_0 = 1, \quad b_1 = 1,432\,788, \quad b_2 = 0,189\,269, \quad b_3 = 0,001\,308$$

Pro  $\frac{1}{2} < p < 1$  je  $u_p = -u_{1-p}$ . Přesnost aproximace je 0,000 45 pro  $-0,999 < P < 0,999$ .

### III. V programu MAPLE

*with(stats) :*

$\varphi(x) = \text{stateval}f[\text{pdf}, \text{normald}](x);$

$f(x) = \text{stateval}f[\text{pdf}, \text{normald}[\mu, \sigma]](x);$

$\Phi(x) = \text{stateval}f[\text{cdf}, \text{normald}](x);$

$F(x) = \text{stateval}f[\text{cdf}, \text{normald}[\mu, \sigma]](x);$

$u_p = \Phi^{-1}(p) = \text{stateval}f[\text{icdf}, \text{normald}](p);$

$x_p = F^{-1}(p) = \text{stateval}f[\text{icdf}, \text{normald}[\mu, \sigma]](p).$

**8.5. Intervaly spolehlivosti.** Ve statistice se setkáváme s úlohou, kdy potřebujeme k dané pravděpodobnosti určit interval, ve kterém se hodnota náhodné veličiny vyskytuje. Vyřešíme tuto úlohu pro normované normální rozdělení. Pro jiná rozdělení se princip řešení zachová,

jenom hraniční hodnoty hledaného intervalu se určí z kvantilů odpovídajícího rozdělení.

**8.6. Příklad:** K danému číslu  $\alpha$ ,  $0 < \alpha < 1$ , určete interval tak, aby pro náhodnou veličinu  $U$ , která má normované normální rozdělení  $N(0; 1)$  platilo:

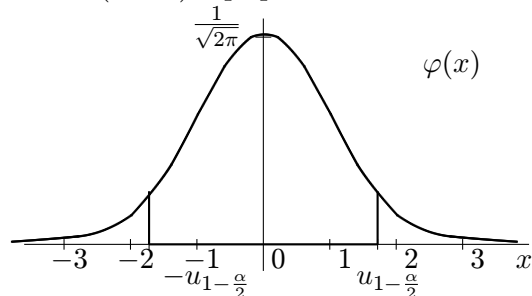
- a) ( $\clubsuit$ )  $P(|U| < a) = 1 - \alpha;$   
 b) ( $\clubsuit\clubsuit$ )  $P(U < a) = 1 - \alpha;$   
 c) ( $\clubsuit\clubsuit\clubsuit$ )  $P(U > a) = 1 - \alpha.$

*Řešení:* a) Z podmínky vyplývá  $1 - \alpha = P(-a < U < a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1 \Rightarrow \Phi(a) = 1 - \frac{\alpha}{2}.$  Odtud plyne, že  $a = u_{1-\frac{\alpha}{2}}$  kvantil. Je tedy  $-a < U < a \Leftrightarrow -u_{1-\frac{\alpha}{2}} < U < u_{1-\frac{\alpha}{2}}.$

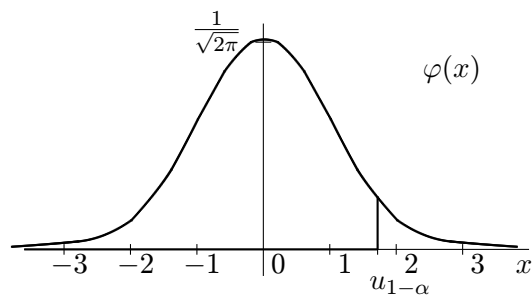
b) Obdobně jako v a) dostaneme  $1 - \alpha = P(U < a) = \Phi(a) \Rightarrow a = u_{1-\alpha}.$  Je tedy  $U < a \Leftrightarrow U < u_{1-\alpha}.$

c) Z podmínky pro interval plyne  $1 - \alpha = P(U > a) = 1 - \Phi(a) \Rightarrow \Phi(a) = \alpha \Rightarrow a = u_{\alpha}.$  Je tedy  $U > a \Leftrightarrow U > u_{\alpha}.$

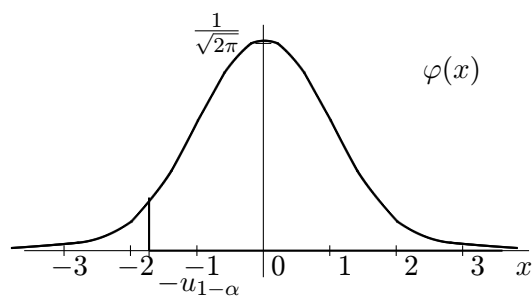
**Poznámka:** Číslo  $\alpha$  volíme malé, obvykle  $0 < \alpha \leq 0,1$  a číslo  $1 - \alpha$  se nazývá *koefficient spolehlivosti (konfidenční koeficient)*. Získaný interval nazýváme  $100(1 - \alpha)$  *procentním intervalem spolehlivosti*. Interval ( $\clubsuit$ ) je oboustranný interval, intervaly ( $\clubsuit\clubsuit$ ) a ( $\clubsuit\clubsuit\clubsuit$ ) jsou jednostranné. První je *pravostranný* a druhý *levostranný*. Jsou to intervaly, ve kterých se hodnota náhodné veličiny bude vyskytovat s pravděpodobností  $(1 - \alpha)$ , tedy ve  $100(1 - \alpha)\%$  případech.



Obr. 8.5



Obr. 8.6



Obr. 8.7.

**8.7. Sčítání náhodných veličin.** Náhodnou veličinu  $Z$ , která je definována pomocí náhodných veličin  $X$  a  $Y$  tak, že

$$X = x, \quad Y = y \Rightarrow Z = x + y$$

nazýváme součtem náhodných veličin  $X$  a  $Y$  a značíme ji  $Z = X + Y$ . Obdobně definujeme i jiné funkce náhodných veličin. Později ukážeme, že platí:

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

a pro nezávislé náhodné veličiny i

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

V případě nezávislých náhodných veličin, které mají normální rozdělení ještě platí:



**8.8. Věta:** Jsou-li  $X$ , resp.  $Y$  nezávislé náhodné veličiny s normálními rozděleními  $N(\mu_1, \sigma_1^2)$ , resp.  $N(\mu_2, \sigma_2^2)$ , pak má náhodná veličina  $X + Y$  normální rozdělení s parametry  $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ .

**8.9. Náhodný výběr.** Ve statistice zpracováváme data, která jsou souborem výsledků náhodného pokusu. Soubor dat je tedy realizací uspořádané  $n$ -tice náhodných veličin  $\{X_1, X_2, \dots, X_n\}$ . Všechny náhodné veličiny mají shodné rozdělení a jsou na sobě nezávislé. Takovou uspořádanou  $n$ -tici náhodných veličin nazýváme *prostým náhodným výběrem* z daného rozdělení. Z vět 8.8 a 8.2 vyplývá toto tvrzení.

**8.10. Věta:** Jestliže mají nezávislé náhodné veličiny  $X_i$ ,  $1 \leq i \leq n$  normální rozdělení  $N(\mu, \sigma^2)$  (náhodný výběr z normálního rozdělení), má pak

$$\text{výběrový úhrn } \tilde{X} = \sum_{i=1}^n X_i \text{ normální rozdělení } N(n\mu, n\sigma^2)$$

$$\text{a výběrový průměr } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ normální rozdělení } N(\mu, \frac{\sigma^2}{n}).$$

**Poznámka: K terminologii:** O náhodné veličině, která je funkcí náhodného výběru mluvíme jako o *statistice*. Častou úlohou je nalezení vhodné statistiky, z jejíchž hodnot můžeme odvodit vlastnosti sledovaného rozdělení. Z vlastností normálního rozdělení vidíme, že statistika  $\bar{X}$ , výběrový průměr je dobrým odhadem střední hodnoty  $\mu$ , neboť při dostatečně rozsáhlém výběru, velké hodnotě  $n$ , se bude hodnota  $\bar{X}$  jen velmi málo lišit od střední hodnoty  $\mu$ .

### Centrální limitní věty

Uvedeme dvě formulace tzv. centrálních limitních vět, které mají obdobný charakter jako věta 8.8.

**8.11. Věta: Čebyševův zákon velkých čísel.** Je-li  $\{X_1, X_2, \dots\}$  posloupnost nezávislých náhodných veličin, pro které je  $E(X_i) = \mu$  a  $D(X_i) \leq C$ , pak pro  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$  a každé číslo  $\varepsilon > 0$  je

$$\lim_{n \rightarrow \infty} P(|Y_n - \mu| > \varepsilon) = 0, \quad \lim_{n \rightarrow \infty} P(|Y_n - \mu| < \varepsilon) = 1.$$

**8.12. Věta: Lindeberg, Lévy** Je-li  $\{X_1, X_2, \dots\}$  posloupnost nezávislých náhodných veličin, pro které je

$E(X_i) = \mu$  a  $D(X_i) = \sigma^2$ , pak pro náhodnou veličinu

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)$$

je

$$\lim_{n \rightarrow \infty} P(\xi_n \leq x) = \Phi(x),$$

kde  $\Phi$  je distribuční funkce normovaného normálního rozdělení  $N(0, 1)$ .

**Důsledek:** Je-li  $\{X_1, X_2, \dots, X_n\}$  prostý náhodný výběr z rozdělení, pro které je  $E(X_i) = \mu$  a  $D(X_i) = \sigma^2$ , pak má v limitě výběrový úhrn  $\tilde{X}$  normální rozdělení  $N(n\mu; n\sigma)$  a výběrový průměr  $\bar{X}$  normální rozdělení  $N(\mu, \frac{\sigma^2}{n})$ . Je tedy

$$\lim_{n \rightarrow \infty} P\left(\frac{\tilde{X} - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad x \in \mathbf{R},$$

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq x\right) = \Phi(x), \quad x \in \mathbf{R}.$$

**Binomické rozdělení**  $Bi(n; p)$  je úhrnem alternativního rozdělení, pro které je  $E(X_i) = p$  a  $D(X_i) = p(1 - p)$ . Centrální limitní věta je Moivre-Laplaceova věta, která byla první z vět tohoto charakteru.

**8.13. Věta: Moivre - Laplace** Jestliže má náhodná veličina  $X$  binomické rozdělení  $Bi(n; p)$ , pak

$$\lim_{n \rightarrow \infty} P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x), \quad x \in \mathbf{R},$$

tedy

$$\begin{aligned} P(a < X < b) &= \\ &= \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

**Poznámka:** Aproximace je jako dobré náhrady možné použít, jsou-li splněny podmínky  $np \geq 5$  a  $n(1 - p) \geq 5$ .

**Poissonovo rozdělení** Jestliže má náhodná veličina  $X$  Poissonovo rozdělení  $Po(\lambda)$ , pak

$$\lim_{\lambda \rightarrow \infty} P\left(\frac{X - \lambda}{\sqrt{\lambda}} \leq x\right) = \Phi(x), \quad x \in \mathbf{R}.$$

Je tedy

$$P(a < X < b) = \Phi\left(\frac{b - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \lambda}{\sqrt{\lambda}}\right).$$