# Některá speciální rozdělení používaná při řešení matematicko-statistických úloh

## Gama rozdělení

**Definice 1 (gama funkce).** Pro všechna reálná čísla $s > 0$ definujme funkci $\Gamma(s)$ předpisem

$$(*) \qquad \Gamma(s) = \int_0^\infty x^{s-1} e^{-x} \mathrm{d}x \,.$$

Lze ukázat, že pro $s > 0$ je integrál $(*)$ konvergentní (tj. existuje a je konečný). Pro $s \le 0$ má tento integrál hodnotu $\infty$.

**Tvrzení 1 (vlastnosti gama funkce).**

($a$) $\Gamma(1) = 1$,

($b$) $\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$,

($c$) $\Gamma(s+1) = s\Gamma(s)$.

**Důkaz.** (a) Je

$$\Gamma(1) = \int_0^\infty e^{-x} \mathrm{d}x = -\left[e^{-x}\right]_0^\infty = -(0-1) = 1 \,.$$

(b) Dle $(*)$ je

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty x^{-1/2} e^{-x} \mathrm{d}x \,.$$

Použijeme-li substituci $x = y^2/2$, dostaneme

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty \sqrt{2} e^{-y^2/2} \mathrm{d}y = \sqrt{2} \int_0^\infty e^{-y^2/2} \mathrm{d}y = \frac{\sqrt{2}}{2} \int_{-\infty}^\infty e^{-y^2/2} \mathrm{d}y = \sqrt{\pi} \,.$$

(c) Integrace per partes ($u = x^s$, $v' = e^{-x}$) dává

$$\Gamma(s+1) = \int_0^\infty x^s e^{-x} \mathrm{d}x = -\left[x^s e^{-x}\right]_0^\infty + s \int_0^\infty x^{s-1} e^{-x} \mathrm{d}x = -(0-0) + s\Gamma(s) = s\Gamma(s) \,. \; \square$$

Dle vlastnosti (c) z předchozího tvrzení k vyčíslení hodnot gama funkce stačí znát hodnoty $\Gamma(s)$ pro $0 < s \le 1$. Například

$$\Gamma(5) = 4 \cdot \Gamma(4) = 4 \cdot 3 \cdot \Gamma(3) = 4 \cdot 3 \cdot 2 \cdot \Gamma(2) = 4 \cdot 3 \cdot 2 \cdot 1 \cdot \Gamma(1) = 4! \,,$$

$$\Gamma\left(\frac{7}{2}\right) = \frac{5}{2} \cdot \Gamma\left(\frac{5}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \Gamma\left(\frac{3}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} \,.$$

Proto bývají hodnoty $\Gamma(s)$ pro $0 < s \le 1$ tabelizovány. (Lze se přitom omezit na tabelování hodnot $\Gamma(s)$ pro $0 < s \le \tfrac{1}{2}$, neboť pro $0 < s < 1$ platí, že $\Gamma(s)\Gamma(1-s) = \pi / \sin \pi s$ a na základě tohoto vzorce

lze vyjádřit hodnoty funkce gama pro $s$ ležící mezi ½ a 1 pomocí jejích hodnot pro $s$ ležící mezi 0 a ½.) Obecněji pro libovolné přirozené číslo $n$ platí:

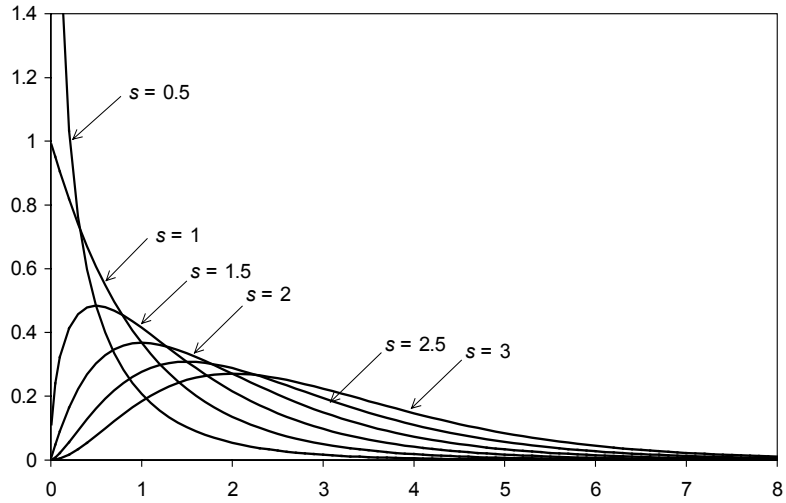
(1) $\Gamma(n) = (n-1)!$,

(2) $\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\cdot\left(n-\frac{3}{2}\right)\cdot\ldots\cdot\frac{1}{2}\sqrt{\pi}$.

**Definice 2 (gama rozdělení).** Z rovnosti $(*)$ vyplývá, že předpisem

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(s)} x^{s-1} e^{-x} & \text{pro } x > 0, \\ 0 & \text{pro } x \le 0, \end{cases}$$

kde $s > 0$ je daný parametr, je definována hustota rozdělení pravděpodobností nějaké náhodné veličiny. Toto rozdělení se nazývá *gama rozdělení s parametry* $s, 1$ a označuje se $\Gamma(s, 1)$. Grafy hustot rozdělení $\Gamma(s, 1)$ pro různé hodnoty parametru $s$ jsou na následujícím obrázku.




Nechť $X$ je náhodná veličina s rozdělením $\Gamma(s, 1)$. Položme $Y = X/\alpha$, kde $\alpha > 0$ je dané číslo. Veličina $Y$ má hustotu rozdělení

$$f(x) = \begin{cases} \dfrac{\alpha^s}{\Gamma(s)} x^{s-1} e^{-\alpha x} & \text{pro } x > 0, \\ 0 & \text{pro } x \le 0. \end{cases}$$

Toto rozdělení nazýváme *gama rozdělení s parametry* $s, \alpha$ a označujeme je $\Gamma(s, \alpha)$. Parametr $s$ určuje „*tvar*“ tohoto rozdělení, tj. tvar funkce $f(x)$, zatímco parametr $\alpha$ má význam „*měřítka*“. Skutečnost, že náhodná veličina $X$ má rozdělení $\Gamma(s, \alpha)$ vyjadřujeme zápisem $X \sim \Gamma(s, \alpha)$.

Rozdělení $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, kde $n$ je přirozené číslo, se nazývá $\chi^2$ *rozdělení* (čti chí-kvadrát rozdělení) s $n$ stupni volnosti a značí se $\chi^2_n$. Hustota rozdělení $\chi^2_n$ je tedy dána předpisem

$$f_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2} & \text{pro } x > 0, \\ 0 & \text{pro } x \le 0. \end{cases}$$

Rozdělení $\Gamma(n,\alpha)$, kde $n$ je přirozené číslo, se nazývá *Erlangovo*; jeho speciální případ $\Gamma(1,\alpha)$ je *exponenciální rozdělení* s parametrem $\alpha$, které se značí $Exp(\alpha)$ a má hustotu definovanou předpisem

$$f(x)=\begin{cases}\alpha e^{-\alpha x} & \text{pro } x>0\,,\\ 0 & \text{pro } x\le 0\,.\end{cases}$$

Zcela speciálně pak $\Gamma\left(1,\frac{1}{2}\right)\sim\chi^2_2\sim Exp\left(\frac{1}{2}\right)$.

**Tvrzení 2.** *Nechť $\beta$ je kladné reálné číslo.*

(*a*) *Jestliže $X\sim\Gamma(s,\alpha)$, pak $\beta X\sim\Gamma(s,\alpha/\beta)$.*

(*b*) *Jestliže $X\sim Exp(\alpha)$, pak $\beta X\sim Exp(\alpha/\beta)$.*

**Důkaz.** Nechť $X$ veličina s rozdělením $\Gamma(s,\alpha)$ a $f(x)$ je hustota tohoto rozdělení. Budiž $\beta$ kladné reálné číslo. Hustota $g(x)$ veličiny $\beta X$ je pak dána předpisem

$$g(x)=\tfrac{1}{\beta}\cdot f(x/\beta)=\begin{cases}\dfrac{(\alpha/\beta)^s}{\Gamma(s)}x^{s-1}e^{-(\alpha/\beta)\cdot x} & \text{pro } x>0\,,\\ 0 & \text{pro } x\le 0\,.\end{cases}$$

To ale znamená, že $\beta X\sim\Gamma(s,\alpha/\beta)$, čímž je dokázáno tvrzení (a). Tvrzení (b) je speciálním případem tvrzení (a). □

**Tvrzení 3.** *Nechť $X$ je náhodná veličina s rozdělením $\Gamma(s,\alpha)$. Pak*

$$\hat{x}=\frac{s-1}{\alpha},\quad E(X)=\frac{s}{\alpha}\quad\text{a}\quad D(X)=\frac{s}{\alpha^2}\,.$$

**Důkaz.** Je-li $f(x)$ hustota rozdělení $\Gamma(s,\alpha)$, pak pro libovolné $k=0,1,2,\ldots$ je

$$E(X^k)=\int_{-\infty}^{\infty}x^k f(x)\mathrm{d}x=\frac{\alpha^s}{\Gamma(s)}\int_0^{\infty}x^{k+s-1}e^{-\alpha x}\mathrm{d}x=\frac{\alpha^s}{\Gamma(s)}\cdot\frac{\Gamma(k+s)}{\alpha^{k+s}}=\frac{\Gamma(k+s)}{\alpha^k\Gamma(s)}\,.$$

Speciálně

$$E(X)=\frac{s}{\alpha},\quad E(X^2)=\frac{s(s+1)}{\alpha^2},$$

a tedy

$$D(X)=E(X^2)-E^2(X)=\frac{s}{\alpha^2}\,.$$

Dále pro $x>0$ je

$$f'(x)=\frac{\alpha^s}{\Gamma(s)}\left[(s-1)x^{s-2}e^{-\alpha x}-\alpha x^{s-1}e^{-\alpha x}\right]=\frac{\alpha^s}{\Gamma(s)}x^{s-2}e^{-\alpha x}(s-1-\alpha x)\,.$$

Funkce $f(x)$ tudíž nabývá svého maxima v bodě $x=(s-1)/\alpha$. □

**Důsledek.**

(*a*) *Nechť $X\sim\chi^2_n$. Pak $\hat{x}=n-2$, $E(X)=n$, $D(X)=2n$.*

(*b*) *Nechť $X\sim Exp(\alpha)$. Pak $\hat{x}=0$, $E(X)=1/\alpha$, $D(X)=1/\alpha^2$.*

Konvolucí dvou (či více) gama rozdělení s týmž měřítkem je opět gama rozdělení a konvolucí dvou (či více) $\chi^2$ rozdělení je opět $\chi^2$ rozdělení. Poněkud přesnější vyjádření této skutečnosti je obsahem následující věty:

**Věta 1.** *Nechť $X$ a $Y$ jsou nezávislé náhodné veličiny, přičemž $X \sim \Gamma(s_1, \alpha)$ a $Y \sim \Gamma(s_2, \alpha)$. Pak*

$$X + Y \sim \Gamma(s_1 + s_2, \alpha).$$

*Speciálně, jestliže $X_1, X_2, \ldots, X_n$ jsou vzájemně nezávislé náhodné veličiny s rozdělením $Exp(\alpha)$, pak veličina $X = X_1 + X_2 + \ldots + X_n$ má rozdělení $\Gamma(n, \alpha)$.*

Důsledkem věty 1 je následující tvrzení.

**Tvrzení 4.** *Nechť $X$ a $Y$ jsou nezávislé náhodné veličiny, přičemž $X \sim \chi^2_m$ a $Y \sim \chi^2_n$. Pak $X + Y \sim \chi^2_{m+n}$.*

Jádrem aplikovatelnosti rozdělení $\chi^2$ při řešení matematicko-statistických úloh je následující věta.

**Věta 2.** *Nechť $X_1, X_2, \ldots, X_n$ jsou vzájemně nezávislé náhodné veličiny s rozdělením $N(0,1)$. Pak*

$$X_1^2 + X_2^2 + \ldots + X_n^2 \sim \chi^2_n.$$

**Důkaz.** Stačí ukázat, že druhá mocnina veličiny s rozdělením $N(0,1)$ má rozdělení $\chi^2_1$ a poté použít tvrzení 4. Nuže nechť $Y = X^2$, kde $X \sim N(0,1)$. Označme po řadě $g(x)$ a $G(x)$ hustotu a distribuční funkci veličiny $Y$, pro hustotu a distribuční funkci rozdělení $N(0,1)$ zvolme standardní označení $\varphi(x)$ a $\Phi(x)$. Veličina $Y$ nabývá pouze nezáporných hodnot, přičemž pro $x > 0$ je

$$G(x) = P(Y < x) = P(X^2 < x) = P\left(-\sqrt{x} < X < \sqrt{x}\right) = 2\Phi\left(\sqrt{x}\right) - 1.$$

Tudíž

$$g(x) = G'(x) = 2\Phi'\left(\sqrt{x}\right) \cdot \left(\sqrt{x}\right)' = 2\varphi\left(\sqrt{x}\right) \cdot \frac{1}{2} x^{-1/2} = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}.$$

To ale znamená, že funkce $g(x)$ je totožná s hustotou rozdělení $\chi^2_1$, což bylo dokázat. □

Za použití centrální limitní věty lehce vyvodíme z věty 2 následující důsledek.

**Tvrzení 5.** *Pro $n \to \infty$ se rozdělení $\chi^2_n$ asymptoticky blíží k rozdělení normálnímu. To znamená, že pro velké $n$ lze rozdělení $\chi^2_n$ nahradit při výpočtech rozdělením normálním, konkrétně rozdělením $N(n, 2n)$.*

## Beta rozdělení

**Definice 3 (beta funkce).** Pro všechna reálná čísla $r, s > 0$ definujme funkci $\mathrm{B}(r, s)$ předpisem

$$(**) \qquad \mathrm{B}(r, s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x.$$

Lze ukázat, že integrál $(**)$ je konvergentní právě tehdy, když obě čísla $r, s$ jsou kladná.

**Tvrzení 6.** *Pro libovolná dvě kladná čísla* $r, s$ *platí*:

$$\mathrm{B}(r, s) = \frac{\Gamma(r) \cdot \Gamma(s)}{\Gamma(r+s)}.$$

**Důkaz.** Nejprve odvodíme jisté alternativní vyjádření pro gama a beta funkci. Substitucí $x = y^2$ ve formuli $(*)$, dostaneme, že

$$\Gamma(s) = 2 \int_0^\infty y^{2s-1} e^{-y^2} \mathrm{d}y .$$

Substitucí $x = \cos^2 \varphi$ ve formuli $(**)$ pak dospějeme k vyjádření

$$\mathrm{B}(r, s) = 2 \int_0^{\pi/2} \cos^{2r-1} \varphi \cdot \sin^{2s-1} \varphi \, \mathrm{d}\varphi .$$

Nyní již můžeme přistoupit k vlastnímu důkazu. Je

$$\Gamma(r) = 2 \int_0^\infty x^{2r-1} e^{-x^2} \mathrm{d}x \quad \text{a} \quad \Gamma(s) = 2 \int_0^\infty y^{2s-1} e^{-y^2} \mathrm{d}y .$$

Vynásobením těchto rovnic a použitím Fubiniovy věty o integraci funkce dvou proměnných dostaneme

$$\Gamma(r) \cdot \Gamma(s) = 4 \int_M x^{2r-1} y^{2s-1} e^{-(x^2+y^2)} \mathrm{d}x\mathrm{d}y ,$$

kde $M = \{(x, y);\, x > 0 \ \ \& \ \ y > 0\}$. Zavedením polárních souřadnic, tj. substitucí

$$x = \rho \cos\varphi , \ \ y = \rho \sin\varphi$$

dále dostaneme

$$\Gamma(r) \cdot \Gamma(s) = 4 \int_N \left(\rho^{2r-1} \cos^{2r-1} \varphi\right) \cdot \left(\rho^{2s-1} \sin^{2s-1} \varphi\right) \cdot e^{-\rho^2} \rho \, \mathrm{d}\rho\mathrm{d}\varphi ,$$

kde $N = \left\{(\rho, \varphi);\, \rho > 0 \ \ \& \ \ 0 < \varphi < \frac{\pi}{2}\right\}$. Opětovným použitím Fubiniovy věty pak konečně máme

$$\Gamma(r) \cdot \Gamma(s) = 2 \int_0^\infty \rho^{2(r+s)-1} e^{-\rho^2} \mathrm{d}\rho \cdot 2 \int_0^{\pi/2} \cos^{2r-1} \varphi \cdot \sin^{2s-1} \varphi \, \mathrm{d}\varphi = \Gamma(r+s) \cdot \mathrm{B}(r, s) ,$$

což bylo dokázat. □

**Poznámka.** Pro nezáporná celá čísla $r, s$ je

$$\mathrm{B}(r+1, s+1) = \frac{r! \cdot s!}{(r+s+1)!} .$$

**Definice 4 (beta rozdělení).** Z rovnosti $(**)$ vyplývá, že předpisem

$$f(x) = \begin{cases} \dfrac{1}{\mathrm{B}(r, s)} x^{r-1} (1-x)^{s-1} & \text{pro } \ 0 < x < 1 , \\ 0 & \text{pro } \ x \notin (0, 1) , \end{cases}$$

kde $r>0$, $s>0$ jsou dané parametry, je definována hustota rozdělení pravděpodobností nějaké náhodné veličiny. Toto rozdělení se nazývá *beta rozdělení* s parametry $r, s$ a značí se $\mathrm{B}_{r,s}$.

(Nakreslete si, jaký tvar má hustota beta rozdělení při různých hodnotách parametrů $r, s$. Speciálně se zaměřte na hodnoty $r, s \in \{½, 1, 2, 3\}$. Jde celkem o šestnáct obrázků.)

## $t$ rozdělení

**Definice 5.** Řekneme, že spojitá náhodná veličina $X$ má *$t$ rozdělení o $n$ stupních volnosti*, jestliže má hustotu

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}.$$

Toto rozdělení budeme označovat $t_n$ a skutečnost, že náhodná veličina $X$ má rozdělení $t_n$ budeme vyjadřovat zápisem $X \sim t_n$. Starší název pro $t$ rozdělení je Studentovo rozdělení. Speciálně je

$$f_1(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2},$$

což znamená, že rozdělení $t_1$ je totožné s Cauchyovým rozdělením (s parametry $a=0,\ b=1$).

**Tvrzení 7.** *Nechť $X \sim t_n$.*
(*a*) *Jestliže $n>1$, pak $E(X)=0$.*
(*b*) *Jestliže $n>2$, pak $D(X)=n/(n-2)$.*

Použití $t$ rozdělení v matematické statistice se zpravidla opírá o následující větu:

**Věta 3.** *Nechť $X$ je náhodná veličina s rozdělením $N(0,1)$ a $Y$ je náhodná veličina nezávislá na $X$ a mající rozdělení $\chi^2_n$. Pak*

$$\frac{X}{\sqrt{Y/n}} \sim t_n.$$

Spojení věty 3 a tvrzení 5 vede k následujícímu důsledku.

**Tvrzení 8.** *Pro $n\to\infty$ se rozdělení $t_n$ asymptoticky blíží k rozdělení $N(0,1)$. To znamená, že pro velké $n$ lze rozdělení $t_n$ nahradit při výpočtech rozdělením $N(0,1)$.*

## $F$ rozdělení

**Definice 6.** Řekneme, že spojitá náhodná veličina $X$ má *$F$ rozdělení o $m$ a $n$ stupních volnosti*, jestliže má hustotu

$$f_{m,n}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}\cdot x\right)^{-(m+n)/2}.$$

Toto rozdělení budeme označovat $F_{m,n}$ a skutečnost, že náhodná veličina $X$ má rozdělení $F_{m,n}$ budeme vyjadřovat zápisem $X \sim F_{m,n}$. Starší název pro $F$ rozdělení je *Fisherovo-Snedecorovo rozdělení*.

Aplikace $F$ rozdělení při řešení úloh z matematické statistiky jsou založeny na následující větě.

**Věta 4.** *Nechť $X$ a $Y$ jsou nezávislé náhodné veličiny, přičemž $X \sim \chi^2_m$ a $Y \sim \chi^2_n$. Pak*

$$\frac{X/m}{Y/n} \sim F_{m,n} \ .$$

## Předmět matematické statistiky

Metody teorie pravděpodobnosti slouží k výpočtu pravděpodobností náhodných jevů, speciálně pak k hledání rozdělení pravděpodobností náhodných veličin a určování charakteristik těchto rozdělení. Při řešení pravděpodobnostních úloh vycházíme vždy z určitých známých skutečností či apriorních předpokladů o okolnostech, za nichž nastávají zkoumané náhodné jevy a z těchto předpokladů pak vyvozujeme závěry o pravděpodobnostech jevů a o rozdělení veličin, jejichž hodnoty jsou při pozorování těchto jevů zaznamenávány. To znamená, že metody teorie pravděpodobnosti jsou *deduktivní* (umožňují totiž určit, s jakou pravděpodobností dojde v důsledku jistých příčin k těm či oněm následkům).

Na druhou stranu metody matematické statistiky jsou ve své podstatě *induktivní*. Jejich prostřednictvím totiž z výsledků pozorování náhodných jevů a z hodnot náhodných veličin, které byly při těchto pozorováních zaznamenány, usuzujeme na příčiny, v jejichž důsledku zkoumané jevy nastaly. Přestože však je způsob statistického usuzování povahy induktivní, lze matematickou statistiku považovat za exaktní vědeckou disciplínu. Je tomu tak proto, že závěry, k nimž při zpracování dat pomocí matematicko-statistických metod docházíme, jsou založeny právě na výsledcích exaktně vybudované teorie pravděpodobnosti. Tyto závěry však nejsou nikdy absolutní; vždy jsou poněkud nepřesné a jen do jisté míry spolehlivé, míra jejich nepřesnosti a nespolehlivosti bývá ovšem známa.

Popišme nyní formálně obecné schéma matematicko-statistické metody. Dejme tomu, že na jednotkách (resp. jedincích) vybraných z jisté populace měříme hodnoty nějaké náhodné veličiny $X$ podléhající určitému zákonu rozdělení $\mathcal{L}$ charakterizujícímu vyšetřovanou populaci. Tento zákon rozdělení přitom buď vůbec neznáme, nebo jej známe pouze částečně (známe například typ rozdělení, nikoliv však jeho parametry). Naším cílem je získat o rozdělení $\mathcal{L}$ prostřednictvím měření hodnot veličiny $X$ nějaké nové informace.

Statistické jednotky, na nichž budeme hodnoty náhodné veličiny $X$ měřit, musíme přitom vybrat z vyšetřované populace nezaujatě, tj. zcela náhodně, neboť jinak by získané informace nebyly objektivní, tj. nebyly by charakteristické pro celou populaci. Změříme-li přitom $n$ jednotek, obdržíme ve skutečnosti posloupnost hodnot ne jedné, nýbrž $n$ náhodných veličin $X_1, X_2, \ldots, X_n$. Všechny tyto veličiny mají však stejné rozdělení pravděpodobností $\mathcal{L}$, neboť předpokládáme, že měření jsou prováděna uvnitř jedné a té samé populace a za stále stejných podmínek. Skutečnost, že jednotky, na nichž jsou měření prováděna, jsou vybrány zcela náhodně, přitom znamená, že veličiny $X_1, X_2, \ldots, X_n$ jsou vzájemně nezávislé. V této souvislosti zavádíme následující definici.

**Definice 7.** Posloupnost $X_1, X_2, \ldots, X_n$ vzájemně nezávislých náhodných veličin s týmž rozdělením pravděpodobností $\mathcal{L}$ se nazývá (*náhodný*) *výběr z rozdělení* $\mathcal{L}$.

Zaměříme se dále na pouze na dva okruhy matematicko-statistických metod, a to na metody *odhadu neznámých parametrů* a *statistické testování hypotéz*. Speciální skupinou metod vyvinutých pro statistické testování hypotéz jsou *metody parametrické*, sloužící k testování hypotéz o neznámých

parametrech pravděpodobnostních rozdělení; ostatní metody testování hypotéz se nazývají *neparametrické*. Dříve než přistoupíme k vlastnímu studiu matematicko-statistických metod, uvedeme několik ilustračních příkladů.

**Příklad 1 (klíčivost semen).** Klíčivost daného druhu semen definujeme jako pravděpodobnost $p$, že semeno vyklíčí. Chceme-li tuto pravděpodobnost odhadnout, vybereme zcela náhodně $n$ semen a zaznamenáme, která semena vyklíčí a která nikoliv. Takový záznam lze pořídit tak, že vyšetříme postupně všechna vybraná semena a v případě, že semeno vyklíčilo, napíšeme na papír jedničku, v opačném případě pak nulu. Obdržíme posloupnost $X_1, X_2, \ldots, X_n$ složenou z jedniček a nul, přičemž libovolný prvek této posloupnosti je náhodnou veličinou s alternativním rozdělením s parametrem $p$. Veličiny $X_1, X_2, \ldots, X_n$ jsou přitom vzájemně nezávislé, neboť semena byla vybrána zcela náhodně. Posloupnost $X_1, X_2, \ldots, X_n$ je tedy náhodným výběrem z alternativního rozdělení s parametrem $p$. To znamená, že odhadujeme-li klíčivost semen, odhadujeme ve skutečnosti parametr $p$ alternativního rozdělení, a to na základě náhodného výběru z tohoto rozdělení. Přitom intuitivně cítíme, že veškerá informace o parametru $p$ je obsažena v údaji o počtu vyklíčených semen, tedy v hodnotě veličiny

$$X = X_1 + X_2 + \ldots + X_n .$$

Jelikož však veličina $X$ má binomické rozdělení s parametry $n, p$, lze na odhad klíčivosti pohlížet též jako na odhad parametru $p$ binomického rozdělení.

**Příklad 2 (hmotnost mincí).** Při výrobě mincí je stanovena hmotnost mince pět gramů. Je podezření, že na materiálu se systematicky šetří. Cílem je toto podezření prokázat či vyvrátit.

Je zřejmé, že není jiné možnosti, jak ohledně vzneseného podezření získat nějakou objektivní informaci, než vybrat náhodně určitý počet mincí a ty zvážit. Označme $n$ počet vybraných mincí a $X_1, X_2, \ldots, X_n$ jejich hmotnosti. Předpokládejme, že zařízení na výrobu mincí je seřízeno tak, aby střední hmotnost mince byla $\mu$ gramů. Při výrobě však dochází k náhodné chybě, a proto skutečná hmotnost mince se od střední hmotnosti zpravidla o trochu liší. Je přitom rozumné předpokládat, že odchylka $e_i$ hmotnosti $i$-té mince od střední hmotnosti $\mu$ je čistě náhodné povahy, což znamená, že

$$e_i \sim N(0, \sigma^2) ,$$

kde $\sigma > 0$ je parametr vyjadřující přesnost výrobní technologie. Jelikož však $X_i = \mu + e_i$, je

$$X_i \sim N(\mu, \sigma^2) .$$

Jinak řečeno, hmotnosti $X_1, X_2, \ldots, X_n$ vybraných mincí představují náhodný výběr z rozdělení $N(\mu, \sigma^2)$. Skutečnost, že se mince vyrábějí systematicky lehčí přitom znamená, že $\mu < 5$. Abychom mohli kvalifikovaně posoudit, zda tomu tak je či nikoliv, musíme ze zaznamenaných hodnot veličin $X_1, X_2, \ldots, X_n$ vytěžit pokud možno maximum informace o parametru $\mu$. Lze přitom ukázat, že v případě výběru z normálního rozdělení, je veškerá informace o parametru $\mu$ obsažena ve výběrovém průměru

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} .$$

Veličina $\overline{X}$ slouží jako odhad parametru $\mu$ rozdělení $N(\mu, \sigma^2)$. K posouzení otázky, zda se mince vyrábějí systematicky lehčí, je přitom nutné vědět, jak je tento odhad přesný.

Je však možno postupovat též tak, že budeme důsledně respektovat pravidlo presumpce neviny a předpokládat, že mince se nevyrábějí systematicky lehčí, tj. že $\mu = 5$. (Možnost, že se mince vyrábějí systematicky těžší a priori vylučujeme.) Takový předpoklad je naší pracovní hypotézou, kterou posléze konfrontujeme s experimentálními daty, tj. se zaznamenanými hmotnostmi $X_1, X_2, \ldots, X_n$ vy-

braných mincí. Je-li průměrná hmotnost vybraných mincí je „významně“ menší než pět gramů, pak naši hypotézu zamítneme. Kritérium, které stanoví, při jak velkém rozdílu mezi průměrnou hmotností vybraných mincí a stanovenou hmotností pěti gramů budeme hypotézu zamítat, je nutné objektivně stanovit ještě předtím než začneme realizovat vlastní experiment, a to tak, aby pravděpodobnost, že zamítneme správnou hypotézu, tj. že obviníme obsluhu výrobní linky ze systematického šetření materiálem, přestože na materiálu se ve skutečnosti systematicky nešetří, byla hodně malá. Na základě tohoto stanoveného kritéria hypotézu buď zamítneme nebo nezamítneme; takový postup se přitom nazývá testem hypotézy. V našem případě jde o test parametrický, konkrétně o test hypotézy o parametru $\mu$ (tj. o střední hodnotě) normálního rozdělení.

**Příklad 3 (výška hladiny piva v půllitru).** Existuje podezření, že v jisté restauraci se systematicky točí pivo pod míru. Cílem je toto podezření prokázat či vyvrátit.

Jde o analogický úkol jako v příkladu 2 s tím rozdílem, že nám není dovoleno přesně změřit vzdálenost hladiny piva od rysky. Můžeme pouze decentně zaznamenat, zda pivo bylo natočeno pod míru či nad míru. Nechť $p$ je pravděpodobnost, že pivo je natočeno pod míru a $n$ je počet piv, která byla během našeho šetření prozkoumána. Testujeme hypotézu, že $p = 1/2$. Veškerá informace o správnosti či nesprávnosti této hypotézy je nyní obsažena v údaji o počtu piv natočených pod míru; označme tento počet $X$. Veličina $X$ má rozdělení $Bi(n, p)$, testujeme tudíž hypotézu o parametru $p$ binomického rozdělení.

## Základní pojmy teorie odhadu

Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr z rozdělení $\mathcal{L}$ závislého na nějakém parametru $\theta$. Předpokládejme přitom, že hodnota parametru $\theta$ není známa. Dále nechť $g(x_1, x_2, \ldots, x_n)$ je funkce $n$ proměnných, jejíž hodnoty na parametru $\theta$ nezávisí. Veličinu $T = g(X_1, X_2, \ldots, X_n)$ nazýváme (*bodovým*) *odhadem* parametru $\theta$. Chceme-li obdržet dobrý odhad parametru $\theta$, musíme dobře zvolit funkci $g$.

Tak například, jestliže $X_1, X_2, \ldots, X_n$ je výběr z rozdělení $N(\mu, \sigma^2)$, pak veličina

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

se jeví být dobrým odhadem parametru $\mu$. Funkce $g$ je v tomto případě definována předpisem

$$g(x_1, x_2, \ldots, x_n) = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Podobně od veličiny

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n}$$

lze očekávat, že bude dobrým odhadem parametru $\sigma^2$.

Existují přitom různá kritéria jak posoudit, zda je nějaký (bodový) odhad neznámého parametru odhadem dobrým. Zformulujme tato kritéria.

**Definice 8.** Řekneme, že $T$ je *nestranný* (*nevychýlený*) odhad parametru $\theta$, jestliže $E(T) = \theta$.

Je-li tedy odhad $T$ nestranný, pak při odhadování parametru $\theta$ veličinou $T$ nedochází k systematickému podhodnocování ani nadhodnocování.

V případě, že jsme schopni pořídit si libovolně veliký náhodný výběr ze zkoumaného rozdělení $\mathcal{L}$, je důležité vědět, zda s rostoucí velikostí výběru se odhad $T$ v nějakém smyslu blíží k hodnotě odhadovaného parametru $\theta$. Takový odhad se nazývá konzistentní. Přesná definice je následující.

**Definice 9.** Nechť $X_1, X_2, \ldots$ je náhodný výběr z rozdělení $\mathcal{L}$ (závislého na parametru $\theta$). Pro každé $n$ nechť je definován odhad $T_n = g_n(X_1, X_2, \ldots, X_n)$ parametru $\theta$. Řekneme, že tento odhad je *konzistentní*, jestliže

$$T_n \xrightarrow{n\to\infty} \theta$$

podle pravděpodobnosti. To znamená, že pro libovolné kladné číslo $\varepsilon$

$$(*) \qquad P(|T_n - \theta| > \varepsilon) \xrightarrow{n\to\infty} 0 .$$

Vztah $(*)$ lze rozepsat takto: Pro libovolné kladné číslo $\varepsilon$ a pro libovolné číslo $\alpha$, $0 < \alpha < 1$, existuje přirozené číslo $n_0$ takové, že pro $n \geq n_0$ je

$$P(|T_n - \theta| > \varepsilon) < \alpha$$

čili

$$(**) \qquad P(|T_n - \theta| < \varepsilon) = P(\theta - \varepsilon < T_n < \theta + \varepsilon) \geq 1 - \alpha .$$

To znamená, že je-li číslo $n$ dostatečně veliké, pak chyba $|T_n - \theta|$ odhadu neznámého parametru $\theta$ veličinou $T_n = g_n(X_1, X_2, \ldots, X_n)$ je s pravděpodobností alespoň $1-\alpha$ menší než libovolná předepsaná mez $\varepsilon$.

Vztah $(**)$ lze zapsat též takto:

$$P(T_n - \varepsilon < \theta < T_n + \varepsilon) \geq 1 - \alpha .$$

Slovně vyjádřeno: Interval $(T_n - \varepsilon, T_n + \varepsilon)$ pokrývá hodnotu neznámého parametru $\theta$ s pravděpodobností alespoň $1-\alpha$. Takový interval nazýváme *intervalovým odhadem* parametru $\theta$. Číslo $\varepsilon$ vyjadřuje přesnost a číslo $\alpha$ spolehlivost odhadu. Hodně spolehlivý odhad přitom nemůže být příliš přesný a naopak hodně přesný odhad nemůže být příliš spolehlivý.

Každý interval $I = I(X_1, X_2, \ldots, X_n)$, který pokrývá hodnotu neznámého parametru $\theta$ s pravděpodobností $1-\alpha$ se nazývá *konfidenčním intervalem* (neboli *intervalem spolehlivosti*) *pro parametr $\theta$ s koeficientem spolehlivosti $\alpha$*. Používán je též název $(1-\alpha)\cdot 100\%$ *interval spolehlivosti*.

Konzistence odhadů se zpravidla ověřuje pomocí následující věty.

**Věta 5.** *Jestliže*

$$(1) \qquad E(T_n) \xrightarrow{n\to\infty} \theta$$

*a*

$$(2) \qquad D(T_n) \xrightarrow{n\to\infty} 0 ,$$

*pak $T_n$ je konzistentním odhadem parametru $\theta$.*

**Důkaz.** Předpokládejme nejprve, že $T_n$ jsou spojité náhodné veličiny s hustotou rozdělení $f_n(x)$. Pak pro libovolné kladné číslo $\varepsilon$ platí:

$$P(|T_n - \theta| > \varepsilon) = P(T_n < \theta - \varepsilon) + P(T_n > \theta + \varepsilon) = \int_{-\infty}^{\theta-\varepsilon} f_n(x)\,\mathrm{d}x + \int_{\theta+\varepsilon}^{\infty} f_n(x)\,\mathrm{d}x = \int_M f_n(x)\,\mathrm{d}x ,$$

kde $M = (-\infty, \theta - \varepsilon) \cup (\theta + \varepsilon, \infty)$.

Pro $x \in M$ je $|x-\theta| > \varepsilon$, tj. $(x-\theta)^2 > \varepsilon^2$, a tedy $\varepsilon^{-2}(x-\theta)^2 f_n(x) \geq f_n(x)$. Odtud plyne, že

$$P(|T_n - \theta| > \varepsilon) = \int_M f_n(x)\mathrm{d}x \leq \varepsilon^{-2} \int_{-\infty}^{\infty} (x-\theta)^2 f_n(x)\mathrm{d}x = \varepsilon^{-2} \int_{-\infty}^{\infty} [x - E(T_n) + E(T_n) - \theta]^2 f_n(x)\mathrm{d}x$$

$$= \varepsilon^{-2} \left\{ \int_{-\infty}^{\infty} [x - E(T_n)]^2 f_n(x)\mathrm{d}x + 2[E(T_n) - \theta] \int_{-\infty}^{\infty} [x - E(T_n)] f_n(x)\mathrm{d}x + [E(T_n) - \theta]^2 \int_{-\infty}^{\infty} f_n(x)\mathrm{d}x \right\}$$

$$= \varepsilon^{-2} \left\{ D(T_n) + [E(T_n) - \theta]^2 \right\} \xrightarrow{n \to \infty} 0 .$$

Jsou-li veličiny $T_n$ diskrétní, provede se důkaz obdobně. □

Nechť nyní $X_1, X_2, \ldots, X_n$ je náhodný výběr z rozdělení $\mathcal{L}$ se střední hodnotou $\mu$ a rozptylem $\sigma^2$. Náhodná veličina

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

se nazývá *výběrový průměr* a veličina

$$S^2 = S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

se nazývá *výběrový rozptyl*. Ukážeme, že za velmi obecných předpokladů je výběrový průměr nestranným a konzistentním odhadem parametru $\mu$ a výběrový rozptyl je nestranným odhadem parametru $\sigma^2$. K tomu, abychom to nahlédli stačí spojit předchozí větu s následujícím tvrzením.

**Tvrzení 8.** *Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr z rozdělení $\mathcal{L}$ majícího konečnou střední hodnotu $\mu$ a konečný rozptyl $\sigma^2$. Pak platí*:

(*a*) $E(\overline{X}) = \mu$,

(*b*) $D(\overline{X}) = \dfrac{\sigma^2}{n}$,

(*c*) $E(S^2) = \sigma^2$ *pro* $n \geq 2$.

**Důkaz.** (a) $E(\overline{X}) = E\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \cdot E\left( \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$.

(b) $D(\overline{X}) = D\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \cdot D\left( \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} D(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

(c) Je

$$E\left[ \sum_{i=1}^{n} (X_i - \overline{X})^2 \right] = E\left[ \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right] = \sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)$$

$$= \sum_{i=1}^{n} \left[ D(X_i) + E^2(X_i) \right] - n\left[ D(\overline{X}) + E^2(\overline{X}) \right] = n(\sigma^2 + \mu^2) - n\left( \frac{\sigma^2}{n} + \mu^2 \right) = (n-1)\sigma^2 ,$$

tudíž

$$E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \frac{1}{n-1}\cdot E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sigma^2,$$

což bylo dokázat. □

**Poznámka.** Označíme-li

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

pak

$$S_n^2 = \frac{n-1}{n}\cdot S_{n-1}^2.$$

To ale znamená, že $S_n^2$ není nestranným odhadem parametru $\sigma^2$, neboť

$$E(S_n^2) = E\left(\frac{n-1}{n}\cdot S_{n-1}^2\right) = \frac{n-1}{n}\cdot E(S_{n-1}^2) = \frac{n-1}{n}\cdot\sigma^2 < \sigma^2.$$

Veličina $S_n^2$ je tedy vždy podhodnoceným odhadem parametru $\sigma^2$, toto podhodnocení přitom ztrácí na významu pro $n \to \infty$. Jinak řečeno veličina $S_n^2$ je *asymptoticky nestranným odhadem* parametru $\sigma^2$. Poznamenejme ještě, že výběrový rozptyl není obecně konzistentním odhadem parametru $\sigma^2$.

Závěr oddílu o základních pojmech teorie odhadu je matematicky poněkud obtížnější a čtenář, který není příliš cvičený v abstraktním myšlení možná udělá lépe, když jej přeskočí a bude pokračovat ve čtení až závěrečným oddílem o principu statistického testování hypotéz.

Nechť $X_1, X_2, \ldots$ je náhodný výběr z rozdělení $\mathscr{L}$ (závislého na parametru $\theta$) a pro každé $n$ je definován odhad $T_n = g_n(X_1, X_2, \ldots, X_n)$ parametru $\theta$. Je-li tento odhad konzistentní, pak chyba, které se při odhadování parametru $\theta$ veličinou $T_n$ dopouštíme je pro $n \to \infty$ s velikou pravděpodobností hodně malá. Je přirozené se přitom ptát, jak je tato chyba veliká při daném počtu měření $n$ a jak rychlá je konvergence odhadu $T_n$ k parametru $\theta$. Ukazuje se, že za velmi obecných předpokladů o rozdělení $\mathscr{L}$ je tato konvergence v průměru nejrychlejší možná pro tzv. maximálně věrohodné odhady parametru $\theta$. Těmto odhadům se proto dává v moderní statistice zpravidla přednost před všemi ostatními odhady a je velmi důležité umět takové odhady zkonstruovat. Podáme nyní definici maximálně věrohodného odhadu a doplníme ji podrobným návodem k použití a několika příklady.

**Definice 10.** Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr z nějakého diskrétního rozdělení pravděpodobností závislého na parametru $\theta$ a $T = g(X_1, X_2, \ldots, X_n)$ je odhad tohoto parametru. Dále nechť $x_1, x_2, \ldots, x_n$ je soubor konkrétních čísel představující experimentálně zaznamenané hodnoty veličin $X_1, X_2, \ldots, X_n$. Pravděpodobnost, že výsledkem provedených pozorování bude soubor hodnot $x_1, x_2, \ldots, x_n$ veličin $X_1, X_2, \ldots, X_n$, závisí na hodnotě parametru $\theta$. Odhad $T$ nazýváme *maximálně věrohodným odhadem* parametru $\theta$, jestliže je tato pravděpodobnost maximální právě tehdy, když za $\theta$ dosadíme hodnotu $g(x_1, x_2, \ldots, x_n)$ tohoto odhadu; tuto hodnotu pak zpravidla značíme $\hat{\theta}$.

Je-li $X_1, X_2, \ldots, X_n$ náhodný výběr ze spojitého rozdělení pravděpodobností závislého na parametru $\theta$, definuje se maximálně věrohodný odhad $T = g(X_1, X_2, \ldots, X_n)$ parametru $\theta$ obdobně s tím rozdílem, že je požadováno, aby dosazení hodnoty $g(x_1, x_2, \ldots, x_n)$ za parametr $\theta$ maximalizovalo nikoliv pravděpodobnost, že výsledkem pozorování bude soubor hodnot $x_1, x_2, \ldots, x_n$ (ta je totiž v tomto případě nulová), nýbrž hustotu této pravděpodobnosti.

Dříve než vyslovíme větu vystihující skutečnost, že maximálně věrohodné odhady jsou v jistém smyslu nejlepší možné, zavedeme pojem *eficientního* neboli *vydatného odhadu.*

**Definice 11.** Řekneme, že nestranný odhad $T$ parametru $\theta$ je *eficientní* (*vydatný*), jestliže střední kvadratická chyba $E(T-\theta)^2$ tohoto odhadu je pro nestranné odhady parametru $\theta$ nejmenší možná. Je-li tedy odhad $T$ eficientní, dopouštíme se při jeho použití v průměru nejmenší možné chyby.

**Věta 6.** *Maximálně věrohodné odhady jsou za velmi obecných předpokladů eficientní* (*a navíc normálně rozdělené*), *pokud* $n \to \infty$. *Říkáme též, že tyto odhady jsou asymptoticky eficientní.*

Hledání maximálně věrohodných odhadů spočívá v hledání kořenů tzv. *věrohodnostních rovnic*, které dále popíšeme. Dejme tomu, že $X_1, X_2, \ldots, X_n$ je náhodný výběr z nějakého *diskrétního* rozdělení $\mathcal{L}$ závislého na parametru $\theta$. Nechť $x_1, x_2, \ldots, x_n$ je libovolná $n$-tice čísel. Ptáme se, pro jakou hodnotu parametru $\theta$ je pravděpodobnost, s níž veličiny $X_1, X_2, \ldots, X_n$ nabývají hodnot $x_1, x_2, \ldots, x_n$, maximální. Tato pravděpodobnost je funkcí $L(x_1, x_2, \ldots, x_n, \theta)$ proměnných $x_1, x_2, \ldots, x_n, \theta$; při známých zaznamenaných hodnotách $x_1, x_2, \ldots, x_n$ a neznámé hodnotě parametru $\theta$ jde pak o funkci jediné proměnné $\theta$. Taková funkce se nazývá *věrohodnostní funkce*. Formálně zapsáno

$$L(x_1, x_2, \ldots, x_n, \theta) = P[(X_1 = x_1) \,\&\, (X_2 = x_2) \,\&\, \ldots \,\&\, (X_n = x_n)]$$

$$= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n),$$

neboť veličiny $X_1, X_2, \ldots, X_n$ jsou vzájemně nezávislé. Hodnotou maximálně věrohodného odhadu parametru $\theta$ je takové číslo $\hat{\theta}$, pro které nabývá funkce $L(x_1, x_2, \ldots, x_n, \theta)$ v proměnné $\theta$ při pevně zadaných hodnotách proměnných $x_1, x_2, \ldots, x_n$ svého maxima. Hledáme-li tento odhad v nějakém otevřeném intervalu $J$ a předpokládáme-li, že funkce $L(x_1, x_2, \ldots, x_n, \theta)$ má v intervalu $J$ derivaci, je číslo $\hat{\theta}$ řešením rovnice

$$(*) \qquad \frac{\partial L(x_1, x_2, \ldots, x_n, \theta)}{\partial \theta} = 0.$$

V praxi se obvykle postupuje tak, že namísto hodnoty proměnné $\theta$, pro níž nabývá maxima funkce $L(x_1, x_2, \ldots, x_n, \theta)$ hledáme takovou hodnotu této proměnné, pro níž nabývá maxima funkce $\ln L(x_1, x_2, \ldots, x_n, \theta)$. (Vzhledem k tomu, že funkce logaritmus je rostoucí, jde zřejmě o úlohy ekvivalentní.) Hledáme tedy kořen rovnice

$$(**) \qquad \frac{\partial \ln L(x_1, x_2, \ldots, x_n, \theta)}{\partial \theta} = 0.$$

Rovnice $(**)$ je tzv. *věrohodnostní rovnice.* Rovnice $(*)$ se nahrazuje rovnicí $(**)$ proto, že funkce $\ln L(x_1, x_2, \ldots, x_n, \theta)$ má zpravidla výrazně jednodušší analytické vyjádření oproti funkci $L(x_1, x_2, \ldots, x_n, \theta)$. Konkrétně

$$L(x_1, x_2, \ldots, x_n, \theta) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n)$$

$$= p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta),$$

kde $p(x, \theta)$ pravděpodobnost, s níž náhodná veličina s rozdělením $\mathcal{L}$ nabývá hodnoty $x$. Tudíž

$$\ln L(x_1, x_2, \ldots, x_n, \theta) = \ln p(x_1, \theta) + \ln p(x_2, \theta) + \ldots + \ln p(x_n, \theta)$$

a věrohodnostní rovnice má tvar

$$\sum_{i=1}^{n} \frac{\partial \ln p(x_i, \theta)}{\partial \theta} = 0 .$$

Je-li $\mathscr{L}$ *spojité* rozdělení pravděpodobností s hustotou $f(x, \theta)$, lze psát

$$P[(X_1 = x_1) \,\&\, (X_2 = x_2) \,\&\, \ldots \,\&\, (X_n = x_n)]$$

$$= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n)$$

$$= f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n .$$

K určení maximálně věrohodného odhadu parametru $\theta$ je tedy třeba maximalizovat *věrohodnostní funkci*

$$L(x_1, x_2, \ldots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) .$$

Odpovídající věrohodnostní rovnice je

$$\sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta} = 0 .$$

Poznamenejme nakonec, že označení $L$ pro věrohodnostní funkci pochází z anglického termínu „*likelihood*“, což je v podstatě ekvivalent pojmu pravděpodobnost, v české matematicko-statistické literatuře se však překládá jako věrohodnost. Rozlišení termínů pravděpodobnost a věrohodnost je podstatné. O maximálně věrohodném odhadu neznámého parametru $\theta$ nelze mluvit jako o odhadu maximálně pravděpodobném, neboť tento parametr není náhodnou veličinou, nýbrž konstantou (byť neznámou.) Pravděpodobnost je v tomto kontextu vztažena k experimentálně zaznamenaným hodnotám $x_1, x_2, \ldots, x_n$ veličin $X_1, X_2, \ldots, X_n$, které hodnotu parametru $\theta$ reflektují.

**Příklad 1 (odhad parametru *p* alternativního a binomického rozdělení).** Připomeňme, že klíčivost semen je definována jako pravděpodobnost $p$, že náhodně vybrané semeno vyklíčí. Předpokládejme přitom, že $0 < p < 1$. Budeme odhadovat parametr $p$. Vybereme proto náhodně $n$ semen a pořídíme si záznam o tom, které semeno vyklíčí a které ne. Tímto záznamem je posloupnost $x_1, x_2, \ldots, x_n$ hodnot náhodných veličin $X_1, X_2, \ldots, X_n$ definovaných předpisem

$$X_i = \begin{cases} 1, & \text{jestliže } i\text{-té semeno vyklíčí,} \\ 0, & \text{jestliže } i\text{-té semeno nevyklíčí.} \end{cases}$$

Položme dále

$$X = X_1 + X_2 + \ldots + X_n .$$

Veličina $X$ vyjadřuje počet vyklíčených semen a má binomické rozdělení s parametry $n$, $p$, zatímco náhodný vektor $X_1, X_2, \ldots, X_n$ je výběrem z alternativního rozdělení s parametrem $p$. Máme tedy co do činění s odhadem parametru $p$ alternativního, resp. binomického rozdělení.

Výchozí datovou strukturou pro odhad neznámého parametru $p$ je záznam hodnot $x_1, x_2, \ldots, x_n$ náhodného vektoru $X_1, X_2, \ldots, X_n$, resp. hodnota $X = k$ počtu vyklíčených semen. Pohlížíme-li na číslo $p$ jako na parametr alternativního rozdělení, je odpovídající věrohodnostní funkce dána předpisem

$$L(x_1, x_2, \ldots, x_n, p) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n) = p^k (1 - p)^{n-k} ,$$

kde $k = x_1 + x_2 + \ldots + x_n$ je zaznamenaný počet vyklíčených semen. Pohlížíme-li na číslo $p$ jako na parametr rozdělení binomického, je odpovídající věrohodnostní funkce dána předpisem

$$L(k, p) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} .$$

Hledáme přitom takovou hodnotu parametru $p$, pro níž nabývá věrohodnostní funkce maximální hodnoty; přitom je zřejmé, že nezáleží na tom, kterou z uvedených věrohodnostních funkcí použijeme. (V obou případech hledáme takovou hodnotu parametru $p$, pro kterou je nejvíce pravděpodobná zaznamenaná hodnota $k$ náhodné veličiny $X$.) Použijeme-li první z uvedených věrohodnostních funkcí, dospějeme k věrohodnostní rovnici

$$\frac{\partial \ln\left[p^k (1-p)^{n-k}\right]}{\partial p} = 0$$

neboli

$$\frac{\partial [k \ln p + (n-k)\ln(1-p)]}{\partial p} = 0 .$$

Vypočteme-li příslušnou derivaci, obdržíme rovnici

$$\frac{k}{p} - \frac{n-k}{1-p} = 0 ,$$

jejímž jediným řešením je číslo $p = k/n$. Druhá derivace funkce $\ln\left[p^k (1-p)^{n-k}\right]$ v proměnné $p$ je přitom záporná, což znamená, že tato funkce nabývá v bodě $p = k/n$ skutečně svého maxima. *Maximálně věrohodným odhadem parametru $p$ je tedy veličina* (*resp. číslo*) $\hat{p} = X/n = k/n$ *vyjadřující relativní počet vyklíčených semen*. *Tento odhad je též nestranný a konzistentní*. Vskutku

$$E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p ,$$

jde tedy o odhad nestranný. Dále

$$D\left(\frac{X}{n}\right) = \frac{D(X)}{n^2} = \frac{pq}{n} \xrightarrow{n\to\infty} 0 ,$$

odkud vyplývá konzistence. Uvědomte si nakonec, že nestrannost a konzistence odhadu je též důsledkem tvrzení 1. Parametr $p$ alternativního rozdělení je totiž střední hodnotou tohoto rozdělení, přičemž při námi zavedeném označení je

$$\frac{X}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n} .$$

**Příklad 2 (odhad parametru Poissonova rozdělení).** Úlohu odhadu parametru Poissonova rozdělení lze ekvivalentně formulovat jako úlohu odhadu intenzity, s níž dochází k výskytu jistých náhodných událostí v prostoru či čase. Prostor může mít libovolnou dimenzi, časová osa pak představuje prostor jednodimenzionální. Předpokládejme přitom, že vymezíme-li v prostoru dvě omezené oblasti nemající společný průnik, pak počet událostí, které nastanou v jedné z oblastí, nemá žádný vliv na počet událostí, které nastanou v druhé oblasti. Dále předpokládejme, že k událostem dochází v celém prostoru se stejnou intenzitou. Kvantitativním vyjádřením intenzity budiž číslo $\lambda$ vyjadřující střední počet událostí připadajících na zvolenou jednotku prostoru či času. Úkolem je odhadnout tuto intenzitu.

Za tím účelem pozorujme výskyt událostí v jisté omezené části $B$ zkoumaného prostoru. Velikost množiny $B$ (tj. její objem, obsah či délku) ve zvolených prostorových či časových jednotkách označme $m(B)$. Představujeme si, že výskyt události v množině $B$ je výsledkem nějakého náhodného procesu $\Phi$; za výše uvedených předpokladů se tento proces nazývá *procesem Poissonovým.* Označme $\Phi(B)$ počet událostí vygenerovaných procesem $\Phi$ v množině $B$. Konkrétní počet událostí zaznamenaných během našeho pozorování v množině $B$ nechť je přitom roven $k$. *Přirozeným odhadem inten-*

*zity* $\lambda$ *je veličina* $\Phi(B)/m(B) = k/m(B)$ *vyjadřující průměrný počet zaznamenaných událostí připadajících na jednotku prostoru či času.* Ukážeme, že *tato veličina je nestranným, konzistentním a maximálně věrohodným odhadem parametru* $\lambda$.

Víme, že náhodná veličina $\Phi(B)$ má za výše uvedených předpokladů Poissonovo rozdělení s parametrem $\lambda \cdot m(B)$. To mimo jiné znamená, že budeme-li velikost množiny $B$ považovat za velikost jednotkovou, stane se intenzita $\lambda$ parametrem Poissonova rozdělení a odhad tohoto parametru odhadem parametru Poissonova rozdělení. Dále odtud plyne, že

$$E\left[\frac{\Phi(B)}{m(B)}\right] = \frac{E[\Phi(B)]}{m(B)} = \frac{\lambda \cdot m(B)}{m(B)} = \lambda ,$$

což znamená, že veličina $\Phi(B)/m(B)$ je nestranným odhadem parametru $\lambda$. Podobně

$$D\left[\frac{\Phi(B)}{m(B)}\right] = \frac{D[\Phi(B)]}{m^2(B)} = \frac{\lambda \cdot m(B)}{m^2(B)} = \frac{\lambda}{m(B)} \xrightarrow{m(B)\to\infty} 0 ,$$

odkud plyne konzistence v tom smyslu, že je-li množina $B$ veliká, dopustíme se při odhadu parametru $\lambda$ veličinou $\Phi(B)/m(B)$ jen s velmi malou pravděpodobností veliké chyby. Přesněji řečeno:

*Jestliže zvolíme posloupnost množin* $B_1 \subseteq B_2 \subseteq \ldots$ *tak, aby* $m(B_n) \xrightarrow{n\to\infty} \infty$, *pak*

$$\Phi(B_n)/m(B_n) \xrightarrow{n\to\infty} \lambda$$

*podle pravděpodobnosti.*

Ukažme konečně, že veličina $\Phi(B)/m(B) = k/m(B)$ je maximálně věrohodným odhadem parametru $\lambda$. Víme, že

$$P[\Phi(B) = k] = e^{-\lambda \cdot m(B)} \cdot \frac{[\lambda \cdot m(B)]^k}{k!} .$$

Maximálně věrohodným odhadem parametru $\lambda$ je ta hodnota tohoto parametru, pro kterou je výše uvedená pravděpodobnost maximální. Věrohodnostní funkce je tedy dána předpisem

$$L(k, \lambda) = e^{-\lambda \cdot m(B)} \cdot \frac{[\lambda \cdot m(B)]^k}{k!} ,$$

odpovídající věrohodnostní rovnice je

$$\frac{\partial \ln L(k, \lambda)}{\partial \lambda} = 0 .$$

Tato rovnice je zřejmě ekvivalentní s rovnicí

$$\frac{\partial \ln\left[e^{-\lambda \cdot m(B)} \cdot \lambda^k\right]}{\partial \lambda} = 0$$

neboli s rovnicí

$$\frac{\partial[k \ln \lambda - \lambda \cdot m(B)]}{\partial \lambda} = 0 .$$

Vypočteme-li derivaci, dostaneme, že $k/\lambda - m(B) = 0$, pokud $k > 0$, resp. $\lambda = 0$, pokud $k = 0$. Odtud vyplývá, že maximálně věrohodným odhadem parametru $\lambda$ je číslo (veličina) $\hat{\lambda} = k/m(B) = \Phi(B)/m(B)$, což bylo dokázat.

**Příklad 3 (odhad parametrů normálního rozdělení).** Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr z rozdělení $N(\mu, \sigma^2)$. Chceme zkonstruovat maximálně věrohodné odhady parametrů $\mu$ a $\sigma^2$.

Označme $x_1, x_2, \ldots, x_n$ experimentálně zaznamenané hodnoty veličin $X_1, X_2, \ldots, X_n$. Věrohodnostní funkce je pak dána předpisem

$$L(x_1, x_2, \ldots, x_n, \mu, \sigma) = f(x_1, \mu, \sigma) \cdot f(x_2, \mu, \sigma) \cdot \ldots \cdot f(x_n, \mu, \sigma),$$

kde

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

Odhadujeme nyní simultánně dva neznámé parametry $\mu$ a $\sigma^2$ (resp. $\mu$ a $\sigma$). Maximálně věrohodné odhady těchto parametrů jsou přitom takové jejich hodnoty, pro něž funkce $L(x_1, x_2, \ldots, x_n, \mu, \sigma)$ dvou proměnných $\mu$ a $\sigma$ nabývá při pevně zdaných hodnotách proměnných $x_1, x_2, \ldots, x_n$ maximální hodnoty. Je tedy třeba řešit *soustavu dvou věrohodnostních rovnic*, totiž

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n, \mu, \sigma)}{\partial \mu} = 0,$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n, \mu, \sigma)}{\partial \sigma} = 0.$$

Snadno zjistíme, že

$$\ln L(x_1, x_2, \ldots, x_n, \mu, \sigma) = \sum_{i=1}^{n} \ln f(x_i, \mu, \sigma) = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2.$$

Vypočteme-li parciální derivace této logaritmované věrohodnostní funkce a včleníme je do soustavy věrohodnostních rovnic, obdržíme po drobných ekvivalentních úpravách soustavu

$$\sum_{i=1}^{n} (x_i - \mu) = 0,$$

$$\sum_{i=1}^{n} (x_i - \mu)^2 - n\sigma^2 = 0.$$

Z první rovnice plyne, že

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

z druhé pak

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

Vidíme tedy, že maximálně věrohodné odhady $\hat{\mu}$ a $\hat{\sigma}^2$ parametrů $\mu$ a $\sigma^2$ rozdělení $N(\mu, \sigma^2)$ jsou dány předpisem

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

## Princip statistického testování hypotéz

Nechť $\mathcal{L}$ je nějaký zákon rozdělení pravděpodobností, o kterém nemáme v danou chvíli buď vůbec žádnou anebo pouze částečnou znalost. Stanovíme nějakou pracovní hypotézu, týkající se rozdělení $\mathcal{L}$ či nějakých jeho speciálních vlastností a tuto hypotézu budeme dále testovat na základě její konfrontace s experimentálními daty. Testovanou hypotézu nazveme *nulovou hypotézou* a označíme ji $H_0$. Experimentální data, s nimiž budeme hypotézu $H_0$ konfrontovat, musí být získána takovým způsobem, aby byla hodnotami vzájemně nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$ rozdělených dle zákona $\mathcal{L}$. Hypotézu $H_0$ *zamítneme*, jestliže zaznamenané hodnoty $x_1, x_2, \ldots, x_n$ veličin $X_1, X_2, \ldots, X_n$ se jeví být s testovanou hypotézou v příliš velkém rozporu. V takovém případě nahradíme hypotézu $H_0$ nějakou *alternativní hypotézou* (stručněji *alternativou*) $H_1$, kterou je třeba zvolit dříve než přistoupíme k vlastnímu testování. Říkáme pak, že *hypotézu $H_0$ testujeme proti alternativě $H_1$*.

Je přitom třeba předem stanovit nějaké objektivní kritérium, dle kterého hypotézu $H_0$ proti alternativě $H_1$ buď zamítneme nebo nezamítneme. Bez ohledu na to, jak sotisfikovaně je toto kritérium zvoleno, nelze nikdy zcela zamezit tomu, že zamítneme správnou hypotézu, podobně musíme připustit, že nedojde k zamítnutí hypotézy, která správná není. Jestliže zamítneme správnou hypotézu, řekneme, že došlo k *chybě prvního druhu.* Nezamítneme-li nesprávnou hypotézu, řekneme, že došlo k *chybě druhého druhu.* Pravděpodobnost chyby prvního druhu se ve statistické teorii nazývá *hladina významnosti* testu. Není přitom rozumné požadovat, aby hladina významnosti testu, tj. pravděpodobnost chyby prvního druhu, byla příliš malá, neboť pak by naopak byla příliš veliká pravděpodobnost chyby druhého druhu. Požadujeme však, aby pravděpodobnost chyby prvního druhu nepřekročila jistou rozumnou mez (např. 0,05) a kritérium pro zamítnutí hypotézy $H_0$ se snažíme volit tak, aby při zachování tohoto požadavku byla pravděpodobnost chyby druhého druhu co možná nejmenší. (Je-li pravděpodobnost chyby druhého druhu malá, tj. s velikou pravděpodobností jsou zamítány nesprávné hypotézy, říkáme, že *test má velikou sílu*.)

Předpokládejme dále, že rozdělení $\mathcal{L}$ je závislé na neznámém parametru $\theta$ a hypotéza $H_0$ je hypotézou o hodnotě tohoto parametru. Přesněji nechť jde o hypotézu $H_0 : \theta = \theta_0$, která stanoví, že hodnota parametru $\theta$ je rovna nějakému pevně zvolenému číslu $\theta_0$. Test hypotézy $H_0$ lze pak založit na konstrukci intervalového odhadu parametru $\theta$. Konkrétně nechť $I = I(X_1, X_2, \ldots, X_n)$ je $(1-\alpha)\cdot 100\%$ interval spolehlivosti pro parametr $\theta$, tj. takový interval, jehož meze jsou funkcí veličin $X_1, X_2, \ldots, X_n$ a který pokrývá hodnotu parametru $\theta$ s pravděpodobností $1-\alpha$. (Později ukážeme, jak lze zkonstruovat takové intervaly spolehlivosti např. pro parametry normálního rozdělení, pro parametr $p$ binomického rozdělení, pro parametr rozdělení Poissonova apod.) Kritérium pro zamítnutí hypotézy $H_0$ budiž následující: hypotézu $H_0$ zamítneme, jestliže interval $I$ neobsahuje číslo $\theta_0$. Snadno nahlédneme, že hladina významnosti takového testu je rovna $\alpha$.

Testy hypotéz o neznámých parametrech nějakého rozdělení $\mathcal{L}$ se nazývají *testy parametrické*, ostatní testy o rozdělení $\mathcal{L}$ se nazývají *neparametrické.* Příkladem neparametrických testů jsou například testy dobré shody vyvinuté pro testování hypotézy, že náhodný výběr $X_1, X_2, \ldots, X_n$ pochází z určitého typu rozdělení (např. rovnoměrného, normálního, Poissonova apod.) Tak například můžeme testovat hypotézu, že odchylka hmotnosti mince od stanovené normy se řídí normálním zákonem rozdělení pravděpodobností.

Ve statistické praxi se k testování hypotéz používá tzv. *testovacích statistik*. Testovací statistika $T$ pro hypotézu $H_0$ je náhodná veličina, která je funkcí veličin $X_1, X_2, \ldots, X_n$ vybraných náhodně z rozdělení $\mathcal{L}$, k němuž se vztahuje testovaná hypotéza. Lze tedy psát $T = T(X_1, X_2, \ldots, X_n)$. Hypotéza $H_0$ se zamítá, jestliže hodnota testovací statistiky $T$ padne do tzv. *kritické oblasti* (*oblasti zamítání*). Kritická oblast je zvolena tak, aby hladina významnosti testu byla rovna požadované hod-

notě (např. 0,05) a současně aby příslušný test měl při dané hladině významnosti co největší možnou sílu. Síla testu závisí ovšem též podstatným způsobem na správné konstrukci testovací statistiky. O celé řadě běžně používaných statistických testů je známo, že mají při dané hladině významnosti největší možnou sílu (alespoň asymptoticky, tj. v případě, že velikost náhodného výběru $X_1, X_2, \ldots, X_n$ je hodně veliká).