Odhady parametrů polohy a rozptýlení pro důležitá rozdělení

Odhady parametrů polohy a rozptýlení pro často se vyskytující rozdělení dat v laboratoři se vyčíslují podle následujících vztahů:

a) **Laplaceovo rozdělení**: Laplaceovo (oboustranné exponenciální) rozdělení se vyskytuje v případech, kdy jsou náhodné veličiny měřeny za podmínek kolísání rozptylu kolem určité střední hodnoty. 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 Laplaceova rozdělení je $E(x) = \Theta$, *rozptyl* $D(x) = 2\,\Phi^2$, *šikmost* $g_1 = 0$ a *špičatost* $g_2 = 6$. Ve srovnání s normálním rozdělením je Laplaceovo rozdělení špičatější a má delší konce. Tak 1%ní kvantil Laplaceova rozdělení je roven $E(x) - 2.72\,\sqrt{D(x)}$, zatímco odpovídající kvantil normálního rozdělení je $E(x) - 2.33\,\sqrt{D(x)}$. Laplaceovo rozdělení připouští výskyt výrazněji odchýlených hodnot a využívá se jako "robustní" alternativa normálního rozdělení. Po dosazení a zlogaritmování vyjde logaritmus věrohodnostní funkce ve tvaru

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

Při známé hodnotě parametru Φ lze maximálně věrohodný odhad $\hat{\Theta}$ parametru Θ získat minimalizací výrazu $\sum_{i=1}^{n} /x_i - \Theta/$.

Maximálně věrohodný odhad θ je výběrový medián $\tilde{\theta} = \tilde{x}_{0.5}$. Odhad parametru φ se počítá dle rovnice

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

Rozptyl se vyjádří vztahem $D(\hat{\Phi}) = \frac{\Phi^2}{n}$. *Interval spolehlivosti parametru* Φ lze konstruovat tak, že je-li známa střední hodnota Θ, lze určit $100(1 - \alpha)$%ní interval spolehlivosti pro Φ podle vztahu

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

b) **Rovnoměrné rozdělení:** rovnoměrné (rektangulární) rozdělení je nejjednodušším typem rozdělení pro oboustranně omezenou náhodnou veličinu, která musí ležet v zadaném intervalu $a - h \le x \le a + h$. 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 chyby, vzniklé zaokrouhlením na k desetinných míst. Hustota pravděpodobnosti rovnoměrného rozdělení má tvar

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

Střední hodnota rovnoměrného rozdělení je $E(x) = a$, *rozptyl* $D(x) = h^2/3$, *šikmost* $g_1 = 0$ a *špičatost* $g_2 = 1.8$. Logaritmus věrohodnostní funkce má tvar

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

pro $a - h \le \min(x_1, ..., x_n) \le \max(x_1, ..., x_n) \le a + h$. Tento vztah nabývá maxima při minimální velikosti h. Je zřejmé, že $\min(x_1, ..., x_n) = x_{(1)}$ a $\max(x_1, ..., x_n) = x_{(n)}$. 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 $\hat{a}$ je totožný s polosumou $\tilde{x}_P$. 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

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

a

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

Pro 100(1 - α)%ní interval spolehlivosti libovolného parametru Θ lze užít asymptotický vztah

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

c) Jednoparametrové exponenciální rozdělení: exponenciální rozdělení je jednostranně ohraničené zdola. Využívá se ho k popisu řady reálných dějů. Exponenciální rozdělení má uplynulý čas, resp. obsazený prostor 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. Exponenciální rozdělení bývá spjato s Poissonovým rozdělením náhodných jevů. Popisuje statistické chování kladné náhodné veličiny pro $x \ge 0$. Jeho hustota pravděpodobnosti je definována vztahem

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

Střední hodnota jednoparametrového exponenciálního rozdělení je $E(x) = \Theta$, *rozptyl* $D(x) = \Theta^2$, *šikmost* $g_1 = 2$ a *špičatost* $g_2 = 9$. *Medián* je roven $\tilde{x}_{0.5} = \Theta \ln 2$. Logaritmus věrohodnostní funkce má tvar

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

Po dosazení se určí maximálně věrohodný odhad $\Theta = \dfrac{\sum_{i=1}^{n} x_i}{n}$ a vyčíslí se odpovídající rozptyl $D(\hat{\Theta}) = \dfrac{\Theta^2}{n}$. Při konstrukci intervalů spolehlivosti se využívá skutečnosti, že náhodná veličina $2\,\hat{\Theta}\,n/\Theta$ má rozdělení $\chi^2(2\,n)$. Pro *100(1-α)%ní interval spolehlivosti* pak platí

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

d) **Exponenciální rozdělení dvouparametrové:** popisuje statistické chování náhodné veličiny, která může nabývat hodnot $x \geq \mu$, tj. je zdola ohraničená. Hustota pravděpodobnosti má tvar

$$f(x) = \Theta^{-1} \exp[(-x + \mu)/\Theta].$$

Střední hodnota dvouparametrového exponenciálního rozdělení je $E(x) = \mu + \Theta$. Vztahy pro rozptyl, šikmost a špičatost jsou stejné jako u jednoparametrového exponenciálního rozdělení. *Odhad* $\hat{\mu}$ je

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

Pro maximálně věrohodný odhad $\hat{\Theta}$ parametru Θ lze napsat

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

Odhad $\hat{\mu}$ má střední hodnotu $E(\hat{\mu}) = \mu + \frac{\Theta}{n}$ a *rozptyl* $D(\hat{\mu}) = \Theta^2/n^2$.

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. Pro *nevychýlené odhady* $\hat{\Theta}_0$ a $\hat{\mu}_0$ lze odvodit vztahy

$$\hat{\mu}_0 = \frac{n\, x_{(1)} - \bar{x}}{n - 1}, \quad D(\hat{\mu}_0) = \frac{\Theta^2}{n(n-1)},$$

$$\hat{\Theta}_0 = \frac{n(\bar{x} - x_{(1)})}{n - 1}, \quad D(\hat{\Theta}_0) = \frac{\Theta^2}{n - 1}$$

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

Pro *100(1 - α)%ní oboustranný interval spolehlivosti parametru Θ* platí

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

Protože má podíl $n\,(x_{(1)} - \mu)\,/\,\hat{\Theta}_0$ *F*-rozdělení se 2 a $(2n - 2)$ stupni volnosti, je spodní mez μ_1 pro $100(1 - \alpha)$%ní interval spolehlivosti parametru μ vyjádřitelná 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

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

e) **Logaritmicko-normální rozdělení dvouparametrov**é: logaritmicko-normální roz-dělení je nejrozšířenější alternativou rozdělení normálního pro jednostranně ohraničená data. Fyzikální veličiny (teplota, tlak, objem, hmotnost, koncentrace, atd.) jsou buď kladné, nebo mají přirozeně definovaný počátek (např. absolutní nula u teploty). Pokud jsou však naměřené hodnoty v blízkosti počátku, je lépe použít např. lognormální rozdělení. Toto rozdělení se používá všude tam, kde se měří nízké koncentrace, malé hmotnosti, malé délky, atd. Typickým příkladem je v analytické chemii stopová analýza. Rovněž souhrnná chyba, která je součinem dílčích malých chyb, má lognormální rozdělení. Náhodná veličina x s dvouparametrovým lognormálním rozdělením souvisí s náhodnou veličinou u s normovaným normálním rozdělením vztahem

$$u = \frac{\ln(x - \mu)}{\sigma}$$

kde μ, σ jsou parametry. Log.-normální rozdělení má náhodná veličina, která může nabývat pouze kladných hodnot, tj. leží v intervalu $0 \le x < \infty$. S využitím vztahů pro hustotu pravděpodobnosti transformované náhodné veličiny lze odvodit hustotu pravděpodobnosti lognormálního rozdělení ve tvaru

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

Náhodná veličina x má dvouparametrové lognormální rozdělení, pokud má náhodná veličina $\ln x$ normální rozdělení $N(\mu, \sigma^2)$. *Střední hodnota* a *rozptyl* náhodné veličiny x se vyčíslí podle vztahů $E(x) = \exp(\mu + 0.5\,\sigma^2)$ a $D(x) = \exp(2\,\mu)\,\omega\,(\omega - 1)$, kde $\omega = \exp(\sigma^2)$. *Šikmost* g_1 a *špičatost* g_2 tohoto rozdělení závisí pouze na veličině ω podle rovnic

$$g_1 = \sqrt{\omega - 1}\,(\omega + 2) \text{ a } g_2 = \omega^4 + 2\,\omega^3 + 3\,\omega^2 - 3 .$$

Také *variační koeficient* δ je pro lognormální rozdělení funkcí pouze parametru ω a platí $\delta = \sqrt{\omega - 1}$. *Módus* $\hat{x}_M$ a *medián med* lze vyjádřit vztahy $\hat{x}_M = \exp(\mu - \sigma^2)$ a $med = \exp(\mu)$. Maximálně věrohodný *odhad parametru polohy* **μ** se určí vztahem

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

a maximálně věrohodný *odhad parametru rozptýlení* σ^2 se vypočte vztahem

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

který je však vychýlený. *Nevychýlený odhad* $\hat{\sigma}_0^2$ se vyčíslí analogicky jako u normálního rozdělení $\hat{\sigma}_0^2 = (n/(n - 1))\,\hat{\sigma}^2$.

V řadě případů je analýza v logaritmické transformaci nevyhovující. Je třeba stanovit odhady parametrů polohy a rozptýlení spolu s jejich intervaly spolehlivosti pro původní data. Jednoduše lze konstruovat intervaly spolehlivosti pro medián, který je exponenciální transformací parametru μ. Pro *100(1 - α)%ní oboustranný interval spolehlivosti mediánu* platí

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

Podobně lze sestrojit i intervaly spolehlivosti pro variační koeficient, šikmost a špičatost, které jsou funkcí pouze parametru σ^2. Pro *100(1 - α)%ní oboustranný interval spolehlivosti variačního koeficientu δ* platí

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

Pokud je třeba odhadnout střední hodnotu původních dat $M = E(x)$ a odpovídající rozptyl $V = D(x)$, užije se vztahů

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

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

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

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

f) **Logaritmicko-normální rozdělení tříparametrové**: toto rozdělení má náhodná veličina, která může nabývat hodnot vyšších než spodní mez Θ, t. zn. 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.