Inference v lognormálním a Paretově rozdělení

Jiří Likeš

V řadě technických, ekonomických a jiných aplikací se často vyskytují asymetrická rozdělení známých tvarů. Dvě důležitá z nich jsou lognormální a Paretovo rozdělení. Tato práce se zabývá odhady a testy některých parametrických funkcí těchto dvou rozdělení.

1. Lognormální rozdělení

Uvažujme náhodnou veličinu, která má hustotu pravděpodobnosti

$$f(y) = 1/(\sigma y\sqrt{2\pi})\ \exp\{-(\ln y - \mu)^2/(2\sigma^2)\}\ ,\quad y>0 \qquad (1.1)$$
$$\quad = 0, \qquad y\leq 0$$

s parametry $-\infty<\mu<\infty$, $\sigma^2>0$.

Má-li veličina Y lognormální rozdělení (1.1), má veličina $Z = \ln Y$ normální rozdělení $N(\mu,\sigma^2)$. Této transformace se využívá při hledání bodových a intervalových odhadů parametrů μ a σ^2 a pro testy hypotéz o těchto parametrech. Téže transformace lze využít i pro intervalové odhady a testy hypotéz pro distribuční funkci

$$F(y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right)\ ,\quad y>0$$

či pro kvantily

$$y_P = \exp(\mu + \sigma u_P)\ ,\quad 0<P<1,$$

lognormálního rozdělení. Zde $\Phi(u)$ značí distribuční funkci a u_P kvantil rozdělení N(0,1).Těmito případy se zde nebudeme zabývat.

1.1. Parametrická funkce θ

Budeme se zabývat odhady parametrické funkce

$$\theta = \exp(b\mu + c\sigma^2), \qquad (1.2)$$

kde b a c jsou daná reálná čísla. Příklady této parametrické funkce jsou:

(i) r-tý obecný moment

$$E(Y^r) = \exp(r\mu + \frac{r^2}{2}\sigma^2), \qquad (1.3)$$

speciálně střední hodnota

$$E(Y)\ \exp(\mu + \frac{1}{2}\sigma^2). \qquad (1.4)$$

(ii) Medián rozdělení

$$Me(Y) = \exp(\mu) \qquad (1.5)$$

(iii) Modus rozdělení

$$Mo(Y) = \exp(\mu - \sigma^2). \tag{1.6}$$

(iv) Parametrická funkce

$$\theta = \exp(c\sigma^2), \tag{1.7}$$

např. podíl E(Y)/Me(Y) nebo E(Y)/Mo(Y) apod.

1.2. Náhodný výběr

Mějme náhodný výběr $Y_1,\ldots,Y_n$ rozsahu $n \geq 2$ z rozdělení (1.1). Pak $Z_i = \ln Y_i$, $i = 1,\ldots,n$, je náhodný výběr z rozdělení $N(\mu, \sigma^2)$. Je známo, že

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n} \ln Y_i \quad , \quad W = \sum_{i=1}^{n} (\ln Y_i - \mu^*)^2 \tag{1.8}$$

je úplná postačující statistika pro (μ, σ^2). Přitom μ^* a W jsou nezávislé náhodné veličiny, μ^* má rozdělení $N(\mu, \sigma^2/n)$ a W/σ^2 má rozdělení $\chi^2(n-1)$.

1.3. Lognormální regrese

Uvažujme regresní model

$$Y = \exp(\mu + u) = \exp(\sum_{j=1}^{p} \alpha_j x_j + u) \tag{1.9}$$

kde $x_1,\ldots,x_p$ jsou vysvětlující proměnné a $\alpha_1,\ldots,\alpha_p$ neznámé regresní parametry. Nechť $x_1,\ldots,x_p$ jsou nestochastické proměnné a nechť u je náhodná veličina mající rozdělení $N(0,\sigma^2)$.

Pro dané $\mathbf{x} = (x_1,\ldots,x_p)'$ má veličina Y střední hodnotu

$$E(Y|\mathbf{x}) = \exp(\mu + \frac{1}{2}\sigma^2) \tag{1.10}$$

a medián

$$Me(Y|\mathbf{x}) = \exp(\mu). \tag{1.11}$$

Obě tyto funkce jsou parametrické funkce typu (1.2). Speciálním případem je parametrická funkce $\exp(\alpha_1)$, je-li $x_1 = 1$ pro $x_2 = \ldots = x_p = 0$ (tj. $\mu = \alpha_1 + \alpha_2 x_2 + \ldots + \alpha_p x_p$ a položíme $x_2 = \ldots = x_p = 0$).

Mějme množinu $n > p$ nezávislých pozorování

$$Y_i = \exp(\sum_{j=1}^{p} \alpha_j x_{ij} + u_i), \quad i = 1,\ldots,n$$

a označme

$$\mathbf{y} = (y_1,\ldots,y_n)', \quad \mathbf{u} = (u_1,\ldots,u_n)', \quad \boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_p)',$$

kde $y_i = \ln Y_i$, $i = 1,\dots,n$.

Pak platí

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\pi}\mathbf{u}, \tag{1.12}$$

kde $\mathbf{X} = (x_{ij})$ je matice typu n x p; předpokládejme, že hodnost $\mathbf{X}$ je rovna p.

Pomocí klasické metody nejmenších čtverců nalezneme odhad

$$\boldsymbol{\alpha}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{1.13}$$

vektoru $\boldsymbol{\alpha}$ a reziduální součet čtverců

$$W = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}^*)'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}^*) . \tag{1.14}$$

Je známo, že za uvedených předpokladů má $\boldsymbol{\alpha}^*$ p-rozměrné normální rozdělení, W/σ^2 má rozdělení $\chi^2(n-p)$, $\boldsymbol{\alpha}^*$ a W jsou nezávislé a jsou úplnou postačující statistikou pro $(\boldsymbol{\alpha}, \sigma^2)$ (viz např. [7]). Toho se opět využívá pro intervalové odhady parametrických funkcí $\sum_{j=1}^{p} h_j\alpha_j$ (např. pro $\sum_{j=1}^{p} \alpha_j x_j$ pro dané $\mathbf{x}$) či pro testy hypotéz o takovýchto parametrických funkcích (např. pro testy o parametrech α_j).

Uvažujme nyní statistiku (pro dané $\mathbf{x}$)

$$\mu^* = \mathbf{x}'\boldsymbol{\alpha}^* = \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} . \tag{1.15}$$

Tato statistika má rozdělení $N(\mu, \lambda^2\sigma^2)$, kde

$$\mu = \mathbf{x}'\boldsymbol{\alpha} , \quad \lambda^2 = \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} . \tag{1.16}$$

1.4. Nejlepší nestranný odhad parametrické funkce θ

Mějme statistiky μ^* a W, které:

jsou úplné postačující statistiky pro μ a σ^2 (nebo známými funkcemi úplné postačující statistiky: pro náhodný výběr je (1.8) úplnou postačující statistikou pro (μ, σ^2); pro lognormální regresi je (1.15) a (1.14) funkcí úplné postačující statistiky $(\boldsymbol{\alpha}^*, W)$);

μ^* má rozdělení $N(\mu, \lambda^2\sigma^2)$ se známým kladným λ^2;

W/σ^2 má rozdělení $\chi^2(\nu)$;

μ^* a W jsou nezávislé.

Hledejme nyní nestranný odhad parametrické funkce (1.2). Střední hodnota

$$E(b\mu^*) = \exp(b\mu + \frac{b^2}{2}\lambda^2\sigma^2),$$

takže je zapotřebí nalézt nestranný odhad parametrické funkce

$\exp\{(c - b^2\lambda^2/2)\sigma^2\}$. Je známo ([8]), že statistika

$$\varphi[aW,\nu] = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2}+k)} \left(\frac{aW}{4}\right)^k \tag{1.17}$$

je nestranným odhadem parametrické funkce $\exp(a\sigma^2/2)$ pro libovolné reálné a. Nestranným odhadem parametrické funkce (1.2) je tedy statistika

$$\theta^* = \exp(b\mu^*) \cdot \varphi[(2c - b^2\lambda^2)W, \nu] \quad . \tag{1.18}$$

Protože θ^* je funkcí pouze úplné postačující statistiky (μ^* a W v případě náhodného výběru, resp. α^* a W v případě lognormální regrese), je θ^* nejlepším nestranným odhadem θ.

Tak např. pro $\theta = \exp(\mu + \sigma^2/2)$ je $\theta^* = \exp(\mu^*)\ \varphi[(1-\lambda^2)W,\nu]$ nebo pro $\theta = \exp(\mu)$ je $\theta^* = \exp(\mu^*)\ \varphi[-\lambda^2 W,\nu]$, kde pro náhodný výběr je $\lambda^2 = 1/n$, $\nu = n-1$ a pro lognormální regresi je λ^2 výraz (1.16) a $\nu = n-p$.

D.J.Finney [3] uvažoval funkci

$$g[t,\nu] = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\nu^k}{\nu(\nu+2)\cdots(\nu+2k-2)} \left(\frac{\nu}{\nu+1}\right)^k t^k \ . \tag{1.19}$$

Platí

$$\varphi[aW,\nu] = g\left[\frac{\nu+1}{2\nu^2}\, aW, \nu\right] \quad . \tag{1.20}$$

1.5. Rozptyl odhadu θ^*

Použijeme-li vztahu (viz [6])

$$E\{\varphi^2[aW,\nu]\} = \exp(a\sigma^2)\ \varphi[a^2\sigma^4, \nu] \ , \tag{1.21}$$

platí pro rozptyl var(θ^*) statistiky (1.18)

$$\mathrm{var}(\theta^*) = \exp(2b\mu + 2b^2\lambda^2\sigma^2)\ \exp\{(2c-b^2\lambda^2)\sigma^2\} \cdot$$
$$\cdot\ \varphi[(2c-b^2\lambda^2)^2\sigma^4, \nu] - \theta^2 =$$
$$= \theta^2\{\exp(b^2\lambda^2\sigma^2)\ \varphi[(2c - b^2\lambda^2)^2\sigma^4,\nu] - 1\} \ . \tag{1.21}$$

Protože θ ani σ^2 neznáme, zajímá nás odhad rozptylu (1.21). Uvažujme $V(\theta^*) = \theta^{*2} - H(\theta^*)$, kde $H(\theta^*)$ je nestranný odhad parametrické funkce $\theta^2 = \exp(2b\mu + 2c\sigma^2)$. Použitím předchozího postupu dostáváme

$$V(\theta^*) = \theta^{*2} - \exp(2b\mu^*)\ \varphi[4(c - b^2\lambda^2)W, \nu] \ . \tag{1.22}$$

Protože $V(\theta^*)$ je funkcí pouze úplné postačující statistiky, je (1.22) nejlepší nestranný odhad $\mathrm{var}(\theta^*)$.

1.6. Parametrická funkce $\theta = \sum_{i=1}^{r} h_i \theta_i$

Uvažujme nyní parametrickou funkci

$$\theta = \sum_{i=1}^{r} h_i \theta_i = \sum_{i=1}^{r} h_i \exp(b_i \mu + c_i \sigma^2) \qquad (1.23)$$

kde h_i, b_i, c_i, $i = 1,\dots,r$, jsou daná reálná čísla. Příklady θ:

(i) $\mathrm{var}(Y) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$,

(ii) $E(Y) - Me(Y) = \exp(\mu + \sigma^2/2) - \exp(\mu)$

pro lognormální rozdělení,

(iii) $E(Y|x) - Me(Y|x)$ (pro dané x) či $E(Y|x_1) - E(Y|x_2)$ (pro dvě různá x_1 a x_2) pro lognormální regresi.

Nejlepším nestranným odhadem (1.23) je statistika

$$\theta^* = \sum_{i=1}^{r} h_i \exp(b_i \mu^*) \, \varphi\left[(2c_i - b_i^2 \lambda^2)W, \nu\right]. \qquad (1.24)$$

Její rozptyl

$$\mathrm{var}(\theta^*) = \sum_{i=1}^{r} \sum_{j=1}^{r} h_i h_j \,\mathrm{cov}(\theta_i^*, \theta_j^*) \qquad (1.25)$$

kde

$$\begin{aligned}\mathrm{cov}(\theta_i^*, \theta_j^*) &= E\left\{\exp\left[(b_i + b_j)\mu^*\right]\right\} \cdot \\ &\quad \cdot E\left\{\varphi\left[a_i W, \nu\right] \varphi\left[a_j W, \nu\right]\right\} - \theta_i \theta_j \\ &= \theta_i \theta_j \left\{\exp(b_i b_j \lambda^2 \sigma^2) \cdot \varphi\left[a_i a_j \sigma^2, \nu\right] - 1\right\},\end{aligned} \qquad (1.26)$$

přičemž

$$a_i = 2c_i - b_i^2 \lambda^2, \; i = 1,\dots,r . \qquad (1.27)$$

Zde se použilo vztahu (viz [5])

$$E\left\{\varphi\left[a_1 W, \nu\right] \varphi\left[a_2 W, \nu\right]\right\} = \exp\left\{(a_1 + a_2)\sigma^2/2\right\} \varphi\left[a_1 a_2 \sigma^4, \nu\right]. \qquad (1.28)$$

2. Paretovo rozdělení

Nechť náhodná veličina X má Paretovo rozdělení s parametry $\gamma > 0$ a $\beta > 0$, takže její hustota pravděpodobnosti

$$f(x) = \begin{cases} \beta \gamma^{\beta} / x^{\beta + 1}, & x > \gamma \\ 0, & x \le \gamma . \end{cases} \qquad (2.1)$$

Toto rozdělení označíme Pa(γ, β).

Má-li veličina X rozdělení Pa(γ, β), má veličina $T = \ln X$ rozdělení $E(\ln \gamma, 1/\beta)$, kde $E(A, \theta)$ značí exponenciální rozdělení s hustotou pravděpodobnosti

$$g(t) = \frac{1}{\theta} \exp\left(- \frac{t - A}{\theta}\right), \quad t > A \tag{2.2}$$
$$= 0, \qquad t \leq A .$$

Tohoto výsledku lze s výhodou využít pro inferenci v Paretově rozdělení, neboť pro exponenciální rozdělení je inference bohatě propracována.

2.1. γ známo

Nejprve uvažujme případ, kdy parametr γ je znám. Mějme náhodný výběr $X_1, \ldots, X_n$ z rozdělení Pa(γ, β), tj. mějme náhodný výběr $T_1 = \ln X_1, \ldots, T_n = \ln X_n$ z rozdělení $E(\ln \gamma, 1/\beta)$. Je známo, že $\sum_{i=1}^{n} T_i = \sum_{i=1}^{n} \ln X_i$ je úplnou postačující statistikou pro parametr $\theta = 1/\beta$ a že veličina

$$\frac{2}{\theta} \sum_{i=1}^{n} (T_i - A) = 2\beta \sum_{i=1}^{n} (\ln X_i - \ln \gamma) \tag{2.3}$$

má rozdělení $\chi^2(2n)$.

2.1.1. Odhady a testy pro parametr β

Protože veličina (2.3) má rozdělení $\chi^2(2n)$, je statistika

$$\beta^* = \frac{n - 1}{\sum_{i=1}^{n} (\ln X_i - \ln \gamma)}, \quad n \geq 2 \tag{2.4}$$

nestranným odhadem parametru β rozdělení Pa(γ, β) pro γ známé. Protože β^* je funkcí úplné postačující statistiky $\sum_{i=1}^{n} \ln X_i$, je (2.4) nejlepší nestranný odhad β pro γ známé. Rozptyl

$$\operatorname{var}(\beta^*) = \beta^2/(n - 2), \quad n \geq 3. \tag{2.5}$$

Maximálně věrohodný odhad parametru β je roven

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} (\ln X_i - \ln \gamma)} = \frac{n}{n-1} \beta^* ; \tag{2.6}$$

$\hat{\beta}$ tedy není nestranný odhad β, ale je asymptoticky nestranný.

Protože veličina (2.3) má rozdělení $\chi^2(2n)$, je interval

$$\frac{\chi^2_{\alpha_1}(2n)}{2 \sum_{i=1}^{n} (\ln X_i - \ln \gamma)} < \beta < \frac{\chi^2_{1-\alpha_2}(2n)}{2 \sum_{i=1}^{n} (\ln X_i - \ln \gamma)} \tag{2.7}$$

$100(1-\alpha)$procentní interval spolehlivosti pro β. Zde $0 \leq \alpha_1 \leq \alpha$, $0 \leq \alpha_2 \leq \alpha$, $\alpha_1 + \alpha_2 = \alpha$ a $\chi^2_\varepsilon(2n)$ je εtý kvantil rozdělení $\chi^2(2n)$.

Pro test hypotézy $H_0: \beta = \beta_0$ se jako testového kritéria použije veličiny $2\beta_0 \sum_{i=1}^{n} (\ln X_i - \ln \gamma)$, která má za platnosti H_0 rozdělení $\chi^2(2n)$.

2.1.2. Odhady hodnoty distribuční funkce

Pro dané $b > \gamma$ chceme odhadnout hodnotu distribuční funkce

$$F(b) = 1 - (\gamma/b)^\beta, \quad b > \gamma. \tag{2.8}$$

Avšak

$$F(b) = P(X < b) = P(\ln X < \ln b) = P(T < \ln b) = P(Z < \ln \frac{b}{\gamma}),$$

kde $Z = T - \ln \gamma$ je náhodná veličina mající rozdělení $E(0, 1/\beta)$.

Z práce [9] vyplývá, že nejlepší nestranný odhad $F^*(b)$ hodnoty $F(b)$ je dán výrazy

$$F^*(b) = \begin{cases} 1, & \sum_{i=1}^{n} \ln(X_i/\gamma) \leq \ln(b/\gamma) \\ 1 - \left(1 - \dfrac{\ln(b/\gamma)}{\sum_{i=1}^{n} \ln(X_i/\gamma)}\right)^{n-1}, & \sum_{i=1}^{n} \ln(X_i/\gamma) > \ln(b/\gamma) \end{cases} \tag{2.9}$$

Označme $\underline{\beta} < \beta < \overline{\beta}$ interval spolehlivosti (2.7) pro β. Pak

$$\underline{F} < F(b) < \overline{F}, \tag{2.10}$$

kde

$$\begin{aligned} \underline{F} &= 1 - (\gamma/b)^{\underline{\beta}} && \text{pro } 0 \leq \alpha_1 \leq \alpha \\ \overline{F} &= 1 - (\gamma/b)^{\overline{\beta}} && \text{pro } 0 \leq \alpha_1 < \alpha \\ &= 1 && \alpha_1 = \alpha \end{aligned} \tag{2.11}$$

je $100(1-\alpha)$procentní interval spolehlivosti pro $F(b)$.

2.1.3. Odhad kvantilu

Ptý kvantil rozdělení $Pa(\gamma, \beta)$ je roven

$$x_P = \gamma(1-P)^{-1/\beta} = \gamma \exp(-\frac{1}{\beta} \ln(1-P)), \quad 0 < P < 1. \tag{2.12}$$

Maximálně věrohodný odhad

$$\hat{x}_P = \gamma(1-P)^{-1/\hat{\beta}} = \gamma \exp\left\{[-\ln(1-P)] \frac{1}{n} \sum_{i=1}^{n} (\ln X_i - \ln \gamma)\right\} \tag{2.13}$$

Lze ukázat, že pro každé $0 < P < 1$ a konečné n je $E(\hat{x}_P) > x_P$, avšak $\lim_{n \to \infty} E(\hat{x}_P) = x_P$.

Označme opět $\underline{\beta} < \beta < \overline{\beta}$ interval spolehlivosti (2.7). Pak

$$\gamma \exp\left\{\left[-\ln(1-P)\right]\frac{1}{\overline{\beta}}\right\} < x_P < \gamma \exp\left\{\left[-\ln(1-P)\right]\frac{1}{\underline{\beta}}\right\} \qquad (2.14)$$

je $100(1-\alpha)$procentní interval spolehlivosti pro x_P.

2.2. γ i β neznámy

Nyní uvažujme případ obou neznámých parametrů γ a β. Mějme uspořádaný výběr $X_{(1)} < \ldots < X_{(n)}$, $n \geq 2$, z rozdělení $Pa(\gamma, \beta)$, tj. mějme uspořádaný výběr $T_{(1)} = \ln X_{(1)} < \ldots < T_{(n)} \; (= \ln X_{(n)})$ z rozdělení $E(\ln \gamma, 1/\beta)$.

Je známo [2], že

$$\left(T_{(1)} = \ln X_{(1)}, \sum_{i=2}^{n} (T_{(i)} - T_{(1)}) = \sum_{i=2}^{n} (\ln X_{(i)} - \ln X_{(1)})\right)$$

je úplná postačující statistika pro dvourozměrný parametr (A, θ) rozdělení $E(A, \theta)$ (v našem případě $A = \ln \gamma$ a $\theta = 1/\beta$). Přitom veličiny $T_{(1)}$ a $\sum_{i=2}^{n}(T_{(i)} - T_{(1)})$ jsou nezávislé, veličina $2n(T_{(1)}-A)/\theta$ má rozdělení $\chi^2(2)$ a veličina $2\sum_{i=2}^{n}(T_{(i)} - T_{(1)})/\theta$ má rozdělení $\chi^2(2n-2)$.

2.2.1. Odhady a testy pro parametr β

Obdobným postupem jako v případě známého γ zjistíme, že v případě, kdy γ není známo, je

$$\beta^* = \frac{n-2}{\sum_{i=2}^{n} (\ln X_{(i)} - \ln X_{(1)})}, \quad n \geq 3 \qquad (2.15)$$

nejlepším nestranným odhadem parametru β. Rozptyl tohoto odhadu

$$\operatorname{var}(\beta^*) = \beta^2/(n-3), \quad n \geq 4. \qquad (2.16)$$

Protože veličina $2\beta \sum_{i=2}^{n} (\ln X_{(i)} - \ln X_{(1)})$ má rozdělení $\chi^2(2n-2)$, je

$$\frac{\chi^2_{\alpha_1}(2n-2)}{2\sum_{i=2}^{n} (\ln X_{(i)} - \ln X_{(1)})} < \beta < \frac{\chi^2_{1-\alpha_2}(2n-2)}{2\sum_{i=2}^{n} (\ln X_{(i)} - \ln X_{(1)})} \qquad (2.17)$$

$100(1-\alpha)$procentní interval spolehlivosti pro β.

Pro test hypotézy H_0: $\beta = \beta_0$ se jako testového kritéria použije veličiny $2\beta_0 \sum_{i=2}^{n}(\ln X_{(i)} - \ln X_{(1)})$.

Maximálně věrohodným odhadem parametru β je statistika

$$\hat{\beta} = \frac{n}{n-2}\beta^* \qquad (2.18)$$

2.2.2. Odhady a testy pro parametr γ

Nejlepším nestranným odhadem parametru γ je statistika (viz [4])

$$\gamma^* = X_{(1)}\left[1 - \frac{1}{n(n-1)}\sum_{i=2}^{n}(\ln X_{(i)} - \ln X_{(1)})\right], \; n \geq 2,\; n\beta > 1. \qquad (2.19)$$

Její rozptyl

$$\operatorname{var}(\gamma^*) = \gamma^2/(\beta(n-1)(n\beta - 2)),\; n \geq 2,\; n\beta > 2. \qquad (2.20)$$

Maximálně věrohodný odhad

$$\hat{\gamma} = X_{(1)}\,. \qquad (2.21)$$

Veličina

$$F = \frac{n(n-1)(\ln X_{(1)} - \ln\gamma)}{\sum_{i=2}^{n}(\ln X_{(i)} - \ln X_{(1)})} \qquad (2.22)$$

má rozdělení $F(2,2n-2)$. Odtud vyplývá, že

$$X_{(1)} \exp\left\{-\frac{F_{1-\alpha}}{n(n-1)}\sum_{i=2}^{n}(\ln X_{(i)} - \ln X_{(1)})\right\} < \gamma < X_{(1)} \qquad (2.23)$$

je $100(1-\alpha)$procentní interval spolehlivosti pro γ. Zde $F_{1-\alpha}$ je $(1-\alpha)$tý kvantil rozdělení $F(2,2n-2)$.

Pro test hypotézy H_0: $\gamma = \gamma_0$ se jako testového kritéria použije veličiny (2.22) s $\gamma = \gamma_0$.

2.3. Některé jiné odhady parametrů β a γ

V práci [10] se uvažují některé jiné odhady parametrů β a γ. Uveďme zde dva z nich:

(i) Odhady z prvního momentu a z minimálního pozorování. Jelikož

$$E(\bar{X}) = \beta\gamma/(\beta - 1), \quad E(X_{(1)}) = n\beta\gamma/(n\beta - 1),$$

dostáváme odhady

$$\tilde{\beta} = (n\bar{X} - X_{(1)})/(n(\bar{X} - X_{(1)})), \quad \tilde{\gamma} = (n\tilde{\beta} - 1)X_{(1)}/(n\tilde{\beta}) = (\tilde{\beta}-1)\bar{X}/\tilde{\beta} \qquad (2.24)$$

(ii) Odhady z kvantilů (pro velké výběry). Zvolí se P_1 a P_2, $0 < P_1 < P_2 < 1$. Z výběru se naleznou odhady kvantilů $\tilde{x}_{P_1}$ a $\tilde{x}_{P_2}$ (např. tak, že se položí $\tilde{x}_{P_i} = X_{(k_i)}$, kde $k_i = [(n+1)P_i]$; $[a]$ značí celou část kladného čísla a). Pak

$$\tilde{x}_{P_i} = \gamma(1 - P_i)^{-1/\tilde{\beta}}, \quad i = 1,2$$

a odtud

$$\tilde{\beta} = \ln\{(1-P_1)/(1-P_2)\}/\ln(\tilde{x}_{P_2}/\tilde{x}_{P_1}), \quad \tilde{\gamma} = \tilde{x}_{P_i}(1-P_i)^{1/\tilde{\beta}} \qquad (2.25)$$

pro $i=1$ nebo $i=2$.

V práci [10] je ukázáno, že odhady (2.24) i (2.25) jsou konzistentní.

Literatura

[1] BRADU,D. and MUNDLAK,Y. (1970). Estimation in lognormal models. J. Amer. Stat. Assoc., 65, 198-211.

[2] EPSTEIN,B. and SOBEL,M. (1954). Some theorems relevant to life testing from an exponential distribution. Ann. Math. Statist., 25, 373-381.

[3] FINNEY,D.J. (1941). On the distribution of a variate whose logarithm is normally distributed. Suppl. to the J. Roy. Statist. Soc., 7, 155-161.

[4] LIKEŠ,J.(1969).Minimum variance unbiased estimates of power-function and Pareto's distribution. Stat. Hefte, 10, 104-110.

[5] LIKEŠ,J. (1980).Variance of the MVUE for lognormal variance. Technometrics, 22, 253-258.

[6] MEHRAN,F. (1973). Variance of the MVUE for the lognormal mean. J. Amer. Stat. Assoc., 68, 726-727.

[7] MOOD, A. M. and GRAYBILL,F.A. (1963). Introduction to the Theory of Statistics. New York: McGraw-Hill Book Co.

[8] NEYMANN,J. and SCOTT,E.L. (1960). Correction for bias introduced by a transformation of variables. Ann. Math. Statist., 31, 643-655.

[9] PUGH,E.L. (1963). The best estimate of reliability in the exponential case. Oper. Res., 11, 57-61.

[10] QUANDT,R.E. (1966). Old and new methods of estimation and the Pareto distribution. Metrika, 10, 55-82.