# Simulační studie robustnosti odhadů prahových parametrů některých rozdělení

**VLADIMÍR VÁCLAVÍK**
bos@kma.zcu.cz
Katedra matematiky, Západočeská univerzita v Plzni

**ABSTRAKT**

**Simulační studie ukazují, že pro výběry běžného rozsahu ($n \le 100$) je obtížné odlišit logaritmicko-normální, Weibullovo a gama rozdělení. Uvažujeme tříparametrické varianty těchto rozdělení s prahovým parametrem. Uvádíme intervalový odhad prahového parametru, který je pro tato rozdělení robustní vůči chybnému výběru modelu. Zabýváme se též nehomogenitou některého z parametrů, zobecněným logaritmicko-normálním a zobecněným gama rozdělením.**

## ÚVOD

Pravděpodobně nejčastěji používanými modely pro shora neomezenou náhodnou veličinu s prahovou hodnotou jsou rozdělení logaritmicko-normální, Weibullovo a rozdělení gama, obsahující tři parametry: prahový parametr, parametr tvaru a parametr měřítka. Dosud není uspokojivě vyřešena zejména otázka testů shody a rozlišování těchto rozdělení. Zřejmě neexistují testy shody nezávislé na parametrech tvaru.

V [1] jsme se zabývali odhady parametrů a testy shody pro první dvě rozdělení. Síla testů shody, je-li alternativou druhé z těchto rozdělení, je malá pro velkou oblast parametrů tvaru.

Uvažujme následující parametrizace hustot pravděpodobnosti těchto rozdělení:

- logaritmicko-normální LN $= LN(\gamma,\mu,\sigma^2)$

$$f(x) = \begin{cases} 0 & \text{pro } x \le \gamma, \\ \dfrac{1}{\sigma\sqrt{2\pi}}\dfrac{1}{x-\gamma}\exp\{-\dfrac{1}{2\sigma^2}[\ln(x-\gamma)-\mu]^2\} & \text{pro } x > \gamma. \end{cases}$$

- Weibullovo W $= W(\alpha,c,\delta)$

$$f(x) = \begin{cases} 0 & \text{pro } x \le \alpha, \\ \dfrac{c}{\delta}\left(\dfrac{x-\alpha}{\delta}\right)^{c-1}\exp\{-(\dfrac{x-\alpha}{\delta})^c\} & \text{pro } x > \alpha. \end{cases}$$

V literatuře se při numerických výpočtech a simulacích obvykle uvažují pro parametry $\sigma$, $c$ nejvýše rozmezí $0{,}1 \le \sigma \le 3$, $0{,}5 \le c \le 5$. Uveďme velmi stručně některé závěry pro sílu testů shody z [1] pro $n = 100$ a hladinu významnosti 0,10:

| Síla testu shody s | Pst chyby 1. druhu | alternativa: W | | alternativa: LN | |
|---|---|---|---|---|---|
| | | $c = 0{,}3$ | $c = 3$ | $\sigma = 0{,}1$ | $\sigma = 3$ |
| LN rozdělením | [0,06; 0,10] | 0,755 | 0,126 | — | — |
| W rozdělením | [0,06; 0,10] | — | — | 0,168 | 0,862 |

## INTERVALOVÉ ODHADY PRO PRAHOVÝ PARAMETR

Při konstrukci intervalových odhadů pro prahový parametr je vzhledem k předchozímu vhodné vycházet z předpokladu, že není jisté, zda výběr pochází z rozdělení LN, W, gama, případně uvažovat i další rozdělení.

Zaměřili jsme se na levý okraj intervalu spolehlivosti pro prahový parametr (shora je prahový parametr triviálně omezen nejmenším prvkem výběru). Při simulacích se ukázalo, že k málo spolehlivému odhadu dochází tehdy, když rozdělení LN je omylem považováno za rozdělení W.

Níže popíšeme intervalový odhad pro parametr $\gamma$ rozdělení LN, který podle našich simulací zajišťuje požadovanou spolehlivost i při případném chybném určení modelu ve "třídě" {LN, W, gama}.

## ALGORITMUS KONSTRUKCE ROBUSTNÍHO ODHADU

Nechť $p \in (0;1)$ (řekněme $p = 0{,}95$) a chceme určit $\gamma_1$ tak, aby pro parametr $\gamma$ rozdělení $LN(\gamma,\mu,\sigma^2)$ platilo $P(\gamma > \gamma_1) \ge p$. K dispozici mějme realizaci náhodného výběru rozsahu $n$ s průměrem $\overline{x}$, výběrovou směrodatnou odchylkou $s$ a nejmenším prvkem $x_{(1)}$.

1. Simulacemi náhodných výběrů $X_1,\ldots,X_n \sim LN(0;0;\sigma^2)$ zjistíme a tabelujeme 100(1-$p$)% kvantil $g_{1-p,n}$ statistiky
$$G = \frac{X_{(1)} - \overline{X}}{S}$$
v závislosti na parametru $\sigma$ (pro $p = 0{,}95$ a $n \in \{30;50;100\}$ je $g_{1-p,n} = g_{1-p,n}(\sigma)$ rostoucí funkce proměnné $\sigma$).
2. Pomocí tabulky z bodu 1. určíme hodnotu
$$\sigma_1 = g^{-1}_{1-p,n}\left(\frac{x_{(1)} - \overline{x}}{s}\right),$$
kde $g^{-1}_{1-p,n}$ je inverzní funkce k $g_{1-p,n}$.
3. Simulacemi náhodných výběrů $X_1,\ldots,X_n \sim LN(0;0;\sigma_1^2)$ zjistíme 100$p$% kvantil $h_{p,n}(\sigma_1)$ statistiky
$$H = X_{(1)} - \overline{X}.$$
4. Spočítáme
$$\mu_1 = \ln\left(\frac{x_{(1)} - \overline{x}}{h_{p,n}(\sigma_1)}\right).$$
5. Levou mez $\gamma_1$ intervalu spolehlivosti definujeme předpisem
$$\gamma_1 = x_{(1)} - \exp\{\mu_1 + \sigma_1\Phi^{-1}(1-2^{-1/n})\},$$
kde $\Phi$ je distribuční funkce normovaného normálního rozdělení.

Pro $p = 0{,}95$ poskytuje hodnota $\gamma_1$ pro výběry z rozdělení LN dolní mez přibližně 95 % intervalu spolehlivosti pro $\gamma$. Pro výběry z rozdělení W a gama je tato hodnota dolní mezí pro prahový parametr se spolehlivostí minimálně 95 %. Hodnotu $\gamma_1$ lze tedy doporučit ke stanovení přibližné dolní meze alespoň 95 % intervalu spolehlivosti pro prahový parametr rozdělení ze "třídy" {LN, W, gama}.

Simulace rovněž potvrdily robustnost meze $\gamma_1$ vůči nehomogenitě rozdělení způsobené změnami parametru měřítka nebo tvaru.

## ZOBECNĚNÁ ROZDĚLENÍ

Náhodná veličina $X$ se řídí čtyřparametrickým (zobecněným) gama rozdělením s parametry $\omega$, $p$, $m$, $\theta$ (viz např. [2]), jestliže má hustotu pravděpodobnosti

$$f(x) = \begin{cases} 0 & \text{pro } x \le \omega, \\ \dfrac{1}{\theta\,\Gamma(m)}\left(\dfrac{x-\gamma}{\theta}\right)^{mp-1}\exp\{-(\dfrac{x-\omega}{\theta})^p\} & \text{pro } x > \omega. \end{cases}$$

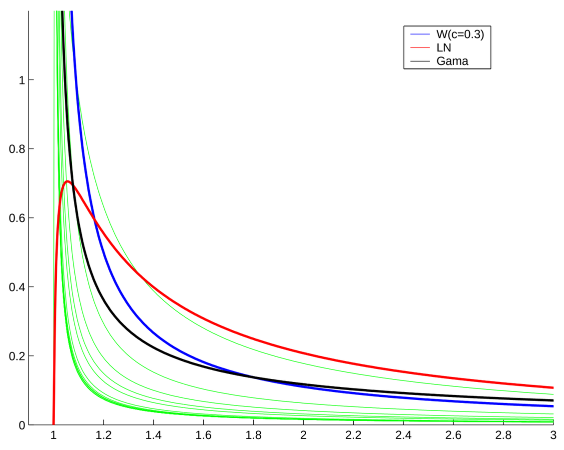
Toto rozdělení je zobecněním jak rozdělení gama, tak Weibullova rozdělení.

Náhodná veličina $X$ se řídí čtyřparametrickým (zobecněným) LN rozdělením s parametry $\gamma$, $\lambda$, $\mu$, $\sigma$ (viz např. [3]), jestliže má hustotu pravděpodobnosti

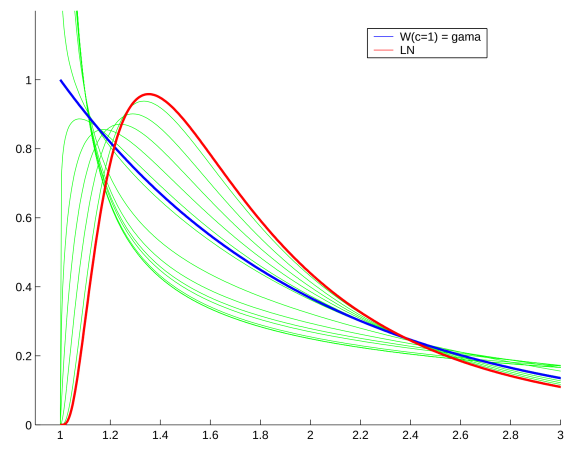
$$f(x) = \begin{cases} 0 & \text{pro } x \le \gamma, \\ \dfrac{1}{\sigma\,\Phi\left([\mu+1/\lambda]/\sigma\right)}\,\phi\left(\dfrac{x(\lambda)-\mu}{\sigma}\right)\,(x-\gamma)^{\lambda-1} & \text{pro } x > \gamma \text{ a } \lambda > 0 \\ \dfrac{1}{\sigma}\,\phi\left([\ln(x-\gamma)-\mu]/\sigma\right)\,(x-\gamma)^{-1} & \text{pro } x > \gamma \text{ a } \lambda = 0 \\ \dfrac{1}{\sigma\,\Phi\left(-[\mu+1/\lambda]/\sigma\right)}\,\phi\left(\dfrac{x(\lambda)-\mu}{\sigma}\right)\,(x-\gamma)^{\lambda-1} & \text{pro } x > \gamma \text{ a } \lambda < 0 \end{cases}$$

kde $\Phi$ a $\phi$ jsou distribuční funkce a hustota pravděpodobnosti normovaného normálního rozdělení a $x(\lambda) = ([x-\gamma]^\lambda - 1)/\lambda$.

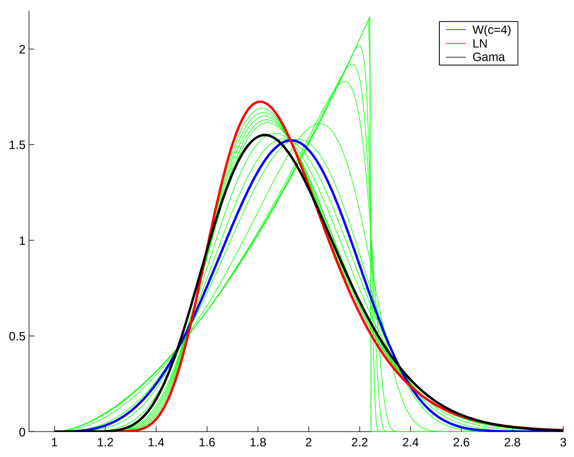
Zobecněním gama a LN rozdělení přidáním čtvrtého parametru bylo dosaženo větší variability těchto rozdělení. Na obrázcích jsou znázorněny některé tvary hustot zobecněného gama rozdělení (zelená barva) a zobecněného LN rozdělení (žlutá barva) ve srovnání s uvažovanými tříparametrickými modely. Parametry u zobrazených hustot byly voleny tak, aby se shodovala střední hodnota a rozptyl náhodných veličin, které mají hustoty zobrazeny na stejném obrázku.



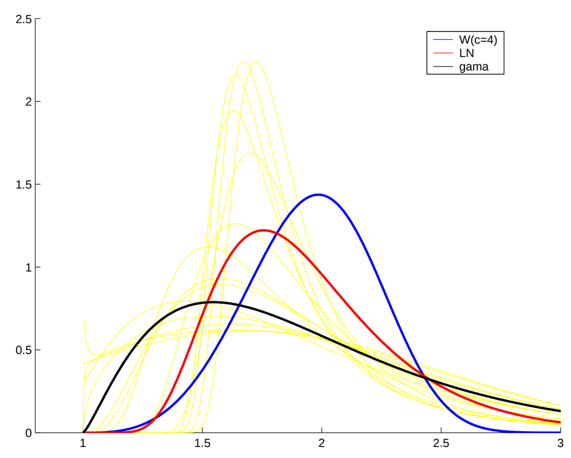
Některé hustoty zobecněného gama rozdělení



Některé hustoty zobecněného gama rozdělení

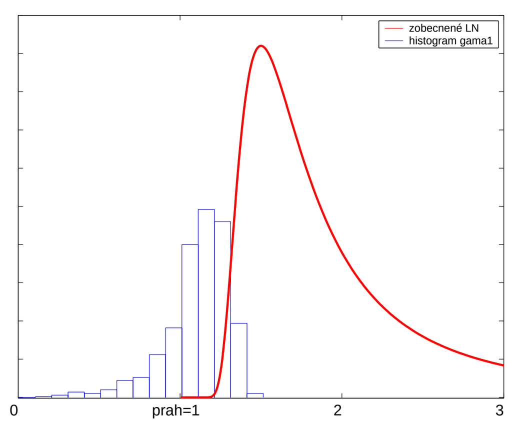


Některé hustoty zobecněného gama rozdělení



Některé hustoty zobecněného LN rozdělení

Uvedená zobecněná rozdělení jsou příkladem modelů, pro které již obecně neplatí doporučení použít $\gamma_1$ jako dolní mez pro alespoň 100$p$% interval spolehlivosti pro prahový parametr. Pro některé hodnoty parametrů (vysoké hodnoty parametrů $p$, $m$ čtyřparametrického gama rozdělení a záporné hodnoty $\lambda$ čtyřparametrického LN rozdělení) leží $\gamma_1$ pro většinu výběrů nad prahovým parametrem $\omega$, resp. $\gamma$.



Na obrázku vpravo je uveden příklad náhodné veličiny ze čtyřparametrického LN rozdělení s parametry $\omega = 1$, $\lambda = -1$, $\mu = 0$ a $\sigma = 1$, histogram zobrazuje výskyt hodnot $\gamma_1$ při 1000 simulacích výběrů z daného rozdělení pro $n = 30$. V tomto příkladu je odhad $\gamma_1$ menší než prahový parametr $\omega$ v 21 % (místo požadovaných 95 %) případů.

### Literatura

**[1]** Reif, J., Václavík, V.: Inference for the three-parameter Weibull and lognormal distributions. Vyjde ve sborníku konference Aplimat 2006 (7.-10. 2. 2006, Bratislava), 8 stran.

**[2]** Hirose, H.: Maximum likelihood parameter estimation by model augmentation with applications to the extended four-parameter generalized gamma distribution. Mathematics and Computers in Simulation 54 (2000), str. 81-97.

**[3]** Chen, G.: Generalized log-normal distributions with reliability application. Computational Statistics & Data Analysis 19 (1995), str. 309–319.

Autor byl finančně podporován z výzkumného záměru MSM 4977751301.