# Teorie extrémních hodnot

## 1.1 Parametrické modely extrémních hodnot

Potřeba parametrického modelování vychází z toho, že se zajímáme o hodnoty, kterých je dosahováno s malými pravděpodobnostmi a které se v pozorovaných datech vyskytují zřídka.

Nejprve uveďme dva základní způsoby, jakými mohou být pozorované hodnoty upraveny, aby na ně mohly být užity modely teorie extrémů:

1. **Bloková maxima**: Data $X_1, X_2, \ldots$ jsou rozdělena do $m$ bloků o velikosti $n$, z nich jsou zjištěna bloková maxima $M_{n1}, \ldots, M_{nm}$. Příkladem mohou být roční maxima průtoků vody získaná z denních měření nebo roční maximální ztráty v řadě denních změn ceny akcií.

2. **Překročení limitu**: Z dat jsou užita pouze pozorování překračující stanovenou vysokou mez (tato metoda je v současnosti preferována, neboť efektivněji využívá informace o extrémních hodnotách obsažené v datech).

### 1.1.1 Rozdělení extrémních hodnot

Parametrickými modely užívanými v analýze blokových maxim jsou *rozdělení extrémních hodnot*. Tato rozdělení jsou limitními rozděleními pro posloupnost vhodně normovaných maxim $M_n$ z $n$ nezávislých stejně rozdělených veličin (při $n \to \infty$). Hrají roli analogickou roli normálního rozdělení v konvergenci normovaných součtů nezávislých stejně rozdělených veličin.

Uveďme distribuční funkce rozdělení extrémních hodnot:

*Gumbelovo rozdělení:* $$G_0(x) = \exp(-e^{-x}),\ x \in \mathbb{R}, \tag{1}$$

*Fréchetovo rozdělení:* $$G_{1,\alpha}(x) = \exp(-x^{-\alpha}),\ x \geq 0,\ \alpha > 0, \tag{2}$$

*Weibullovo rozdělení:* $$G_{2,\alpha}(x) = \exp(-(-x)^{-\alpha}),\ x \leq 0,\ \alpha < 0. \tag{3}$$

Úplné třídy těchto rozdělení dostaneme zavedením parametru polohy $\mu$ a parametru měřítka $\sigma > 0$:

$$G_{0,\mu,\sigma}(x) = \exp\left(-e^{-(x-\mu)/\sigma}\right), \tag{4}$$

$$G_{i,\alpha,\mu,\sigma}(x) = G_{i,\alpha}\left(\frac{x-\mu}{\sigma}\right),\ i = 1, 2. \tag{5}$$

Poznamenejme, že parametr $\mu$ je levým koncovým bodem distribuční funkce Fréchetova rozdělení a pravým koncovým bodem distribuční funkce Weibullova rozdělení.

Výše uvedené modely je možné zavedením reparametrizace $\gamma = \frac{1}{\alpha}$ převést na jednotný model *zobecněného rozdělení extrémních hodnot* s distribuční funkcí

$$G_\gamma(x) = \exp\left(-(1+\gamma\,x)^{-1/\gamma}\right), \qquad \gamma \neq 0, \tag{6}$$
$$G_0(x) = \exp(-e^{-x}), \qquad \gamma = 0, \tag{7}$$

kde $1+\gamma\,x > 0$.

Parametry polohy a měřítka v původních třídách rozdělení byly zvoleny tak, aby platilo

$$\lim_{\gamma\to 0} G_\gamma(x) = G_0(x).$$

Pro $\gamma > 0$ dostáváme Fréchetovo rozdělení s levým koncovým bodem $-1/\gamma$. Pro $\gamma < 0$ dostáváme Weibullovo rozdělení s pravým koncovým bodem $1/|\gamma|$.

V modelu zobecněného rozdělení extrémních hodnot je $\gamma$ parametrem určujícícm tvar rozdělení, tříparametrickou rodinu opět dostaneme zavedením parametru polohy $\mu$ a parametru měřítka $\sigma > 0$:

$$G_{\gamma,\mu,\sigma}(x) = \exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right), \qquad \gamma \neq 0, \tag{8}$$
$$G_{0,\mu,\sigma}(x) = \exp\left(-e^{\frac{x-\mu}{\sigma}}\right), \qquad \gamma = 0, \tag{9}$$

kde $1+\gamma\,\frac{x-\mu}{\sigma} > 0$.

### 1.1.2 Fisherova-Tippettova věta

Předpokládejme, že $X_1, X_2, \ldots$ je posloupnost nezávislých stejně rozdělených náhodných veličin s distribuční funkcí $F$. Potom

$$M_n = \max(X_1, \ldots, X_n)$$

má distribuční funkci

$$\mathrm{P}(M_n \leq x) = F^n(x).$$

Limitní chování normovaných maxim charakterizuje

**Věta 1** (Fisher, Tippett)**.** *Pokud existují posloupnosti konstant $c_n > 0,\ d_n \in \mathbb{R}$ tak, že*

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n\to\infty} F^n(c_n x + d_n) = G(x) \tag{10}$$

*pro nějakou nedegenerovanou distribuční funkci $G$, potom $G$ je distribuční funkce zobecněného rozdělení extrémních hodnot.*

Pokud je splněn předpoklad Fisherovy-Tippettovy věty, říkáme, že distribuční funkce $F$ náleží do *sféry přitažlivosti (maximum domain of attraction)* rozdělení s distribuční funkcí $G$, značíme $F \in \mathrm{MDA}(G)$.

Poznámka. Při konvergenci normovaných maxim je typ limitního rozdělení (specifikovaný parametrem $\gamma$) určen jednoznačně, parametr polohy a měřítka závisí na zvolených posloupnostech normujících konstant. Ty je vždy možno zvolit tak, že limitou je standardní distribuční funkce $G_\gamma (= G_{\gamma,0,1})$.

Označme $\overline{F}(x) = 1 - F(x)$. Sféru přitažlivosti rozdělení extrémních hodnot charakterizuje následující tvrzení:

**Věta 2.** *Distribuční funkce $F$ patří do sféry přitažlivosti rozdělení extrémních hodnot s distribuční funkcí $G$ s normujícími konstantami $c_n > 0,\ d_n \in \mathbb{R}$, právě když*

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\ln G(x),\ x \in \mathbb{R}. \tag{11}$$

*(Pro $G(x) = 0$ interpretujeme limitu jako $+\infty$.)*

*Důkaz.* Nechť platí

$$\lim_{n\to\infty} \mathrm{P}\left(M_n \le c_n\, x + d_n\right) = \lim_{n\to\infty} \left(1 - \overline{F}(c_n\, x + d_n)\right)^n = G(x) \ne 0. \tag{12}$$

Potom je $\lim_{n\to\infty} \overline{F}(c_n\, x + d_n) = 0$ a tvrzení dostaneme z (12) užitím řádové rovnosti $-\log(1-x) \sim x,\ x \to 0$.

Nyní předpokládejme, že platí (11), kde $G(x) \ne 0$. Potom je

$$\mathrm{P}\left(M_n \le c_n\, x + d_n\right) = \left(1 + \frac{\log G(x)}{n} + o\left(\frac{1}{n}\right)\right)^n \to G(x),\ n \to \infty.$$

V případě $G(x) = 0$ se z předpokladu, že $\lim_{n\to\infty} n\,\overline{F}\left(c_n\, x + d_n\right) = \infty$ dokáže platnost vztahu $\lim_{n\to\infty} \mathrm{P}\left(M_n \le c_n\, x + d_n\right) = 0$ sporem: pokud by posledně uvedená limita byla nenulová, mohli bychom vybrat posloupnost $\{n_k\}$ tak, že $\lim_{k\to\infty} n_k\,\overline{F}\left(c_{n_k}\, x + d_{n_k}\right) < \infty$. Obrácená implikace se dokáže podobně. □

### 1.1.3 Fréchetovo rozdělení

U Fréchetovo rozdělení s distribuční funkcí (2) má pravděpodobnost překročení hodnoty $x$ vyjádření

$$\overline{G}_{1,\alpha}(x) = 1 - \exp\left(-x^{-\alpha}\right) \sim x^{-\alpha},\ x \to \infty.$$

Pro momenty Fréchetova rozdělení platí

$$\mathrm{E}\, X^j = \int_0^\infty x^j\, \mathrm{d}G_{1,\alpha}(x) = \Gamma(1 - j/\alpha),\ j < \alpha,$$

kde

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

Sféru přitažlivosti Fréchetova rozdělení lze charakterizovat pomocí chvostů příslušných rozdělení:

**Věta 3.**
$$F \in \mathrm{MDA}(G_{1,\alpha}) \iff \overline{F}(x) = x^{-\alpha}L(x),\ \alpha > 0,$$
*kde $L$ je kladná lebesgueovsky měřitelná funkce na* $(0,\infty)$ *splňující*
$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1,\ t > 0. \tag{SV}$$

Podmínka (SV) definuje *pomalu se měnící funkci v nekonečnu.*

Podle výše uvedeného tvrzení jsou chvosty rozdělení ze sféry přitažlivosti Fréchetova rozdělení *pravidelně se měnící funkce v nekonečnu.* Rozdělení z $\mathrm{MDA}(G_{1,\alpha})$ jsou rozdělení s „těžkými chvosty“. Lze například ukázat, že momenty $\mathrm{E}\,X^j$ rozdělení nezáporné náhodné veličiny náležející do $\mathrm{MDA}(G_{1,\alpha})$ jsou nekonečné pro $j > \alpha$.

*Příklad.* Paretovo rozdělení s distribuční funkcí
$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^{\alpha},\ \alpha > 0,\ \kappa > 0,\ x \geq 0,$$
patří do sféry přitažlivosti Fréchetova rozdělení. Vedle výše uvedené charakterizace pomocí chvostů lze toto ověřit přímo volbou normujících posloupností v (10):
$$c_n = \kappa n^{1/\alpha}/\alpha,\ d_n = \kappa n^{1/\alpha} - \kappa.$$

### 1.1.4 Gumbelovo rozdělení

Pro distribuční funkci Gumbelova rozdělení platí
$$\overline{G}_0(x) = 1 - \exp\left(-e^{-x}\right) \sim e^{-x},\ x \to \infty,$$
jeho střední hodnota je rovna Eulerově konstantě $\lambda = 0.577216\ldots$

Sféra přitažlivosti Gumbelova rozdělení obsahuje distribuční funkce s omezeným i neomezeným nosičem. Kladné náhodné veličiny z $\mathrm{MDA}(G_0)$ mají konečné momenty všech řádů.

*Příklad.* Pro exponenciální rozdělení s distribuční funkcí
$$F(x) = 1 - \exp(-\beta x),\ \beta > 0,\ x \geq 0,$$
lze ověřit podmínku (10) volbou
$$c_n = \frac{1}{\beta},\ d_n = \frac{\ln n}{\beta}.$$

Do sféry přitažlivosti Gumbelova rozdělení patří i další často používaná rozdělení, například rozdělení gama, normální či logaritmicko-normální.

### 1.1.5 Weibullovo rozdělení

Weibullovo rozdělení má zprava omezený nosič, není proto příliš významné pro modelování extrémních jevů. Pro momenty rozdělení s distribuční funkcí (3) dostaneme vyjádření

$$\mathrm{E}\, X^j = \int_{-\infty}^{0} x^j \,\mathrm{d}G_{2,\alpha}(x) = (-1)^j\, \Gamma(1 - j/\alpha).$$

Při označení

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$$

lze podobně jako u Fréchetova rozdělení charakterizovat sféru přitažlivosti Weibullova rozdělení pomocí chvostů příslušných distribučních funkcí.

**Věta 4.**

$$F \in \mathrm{MDA}(G_{2,\alpha}) \iff x_F < \infty \text{ a } \overline{F}(x_F - x^{-1}) = x^{\alpha} L(x),\ \alpha < 0,$$

*kde $L$ splňuje podmínku (SV).*

Příkladem rozdělení z $\mathrm{MDA}(G_{2,\alpha})$ je rozdělení beta definované na intervalu (0,1).

### 1.1.6 Zobecněné Paretovo rozdělení

Při analýze dat podle překročení stanovené vysoké meze hrají významnou úlohu rozdělení s distribučními funkcemi

| | | | |
|---|---|---|---|
| *exponenciální rozdělení:* | $W_0(x)$ | $= 1 - e^{-x},\ x \geq 0$ | (13) |
| *Paretovo rozdělení:* | $W_{1,\alpha}(x)$ | $= 1 - x^{-\alpha},\ x \geq 1,$ | (14) |
| *beta rozdělení:* | $W_{2,\alpha}(x)$ | $= 1 - (-x)^{-\alpha},\ -1 \leq x \leq 0.$ | (15) |

Parametr $\alpha$ je pro Paretovo rozdělení kladný a pro rozdělení beta záporný.

Uveďme vyjádření obecných momentů těchto rozdělení.

$$\begin{aligned}
W_0 : &\qquad \mathrm{E}\, X^j = j!,\\
W_{1,\alpha} : &\qquad \mathrm{E}\, X^j = \frac{\alpha}{\alpha - j},\ j < \alpha,\\
&\qquad\qquad\ \ = +\infty,\ j \geq \alpha,\\
W_{2,\alpha} : &\qquad \mathrm{E}\, X^j = (-1)^j \frac{\alpha}{\alpha - j}.
\end{aligned}$$

Podobně jako v případě rozdělení extrémních hodnot můžeme pro popis výše uvedených distribučních funkcí s vhodně zvolenými parametry polohy a měřítka zavést jednotný tvar v závislosti na parametru $\gamma$. Dostáváme pak distribuční funkci

$$W_\gamma(x) = 1 - (1 + \gamma\, x)^{-1/\gamma}, \tag{16}$$

kde Paretovu rozdělení odpovídá volba $\gamma > 0$ a nosič $x \geq 0$ a rozdělení beta odpovídá volba $\gamma < 0$ a nosič $0 \leq x \leq 1/|\gamma|$. Z (16) také vyplývá spojitost uvedeného vyjádření v parametru $\gamma$ ve smyslu konvergence k distribuční funkci exponenciálního rozdělení při $\gamma \to 0$.

Obecný tvar distribuční funkce zobecněného Paretova rozdělení pak opět dostaneme zavedením parametru polohy a měřítka:

$$W_{\gamma,\mu,\sigma}(x) = W_\gamma\left(\frac{x-\mu}{\sigma}\right), \ \mu \in \mathbb{R}, \ \sigma > 0. \tag{17}$$

Parametr $\mu$ je vždy levým koncovým bodem distribuční funkce (17).

Zaveďme nyní označení pro distribuční funkci rozdělení hodnot, které v pozorování veličin s distribuční funkcí $F$ překračují stanovenou mez $u$,

$$F^{[u]}(x) = \mathrm{P}(X \leq x | X > u) = \frac{F(x) - F(u)}{1 - F(u)}, \, x \geq u,$$

a pro distribuční funkci rozdělení excesů nad mezí $u$ (výší překročení meze),

$$F^{(u)}(x) = \mathrm{P}(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \, 0 \leq x \leq x_F - u. \tag{18}$$

### 1.1.7 POT-stabilita zobecněného Paretova rozdělení

Distribuční funkce zobecněných Paretových rozdělení jsou jediné spojité distribuční funkce takové, že pro určitou volbu konstant $b_u$ a $a_u$ platí

$$F^{[u]}(b_u + a_u\, x) = F(x). \tag{19}$$

Vlastnost (19) se nazývá *POT-stabilita* (z anglického označení pro metody založené na sledování hodnot překračujících stanovenou mez - *peaks over treshold*). Uveďme nyní některé konkrétní příklady:

1. Exponenciální rozdělení s nulovým levým koncovým bodem:

$$W^{[u]}_{0,0,\sigma} = W_{0,u,\sigma}.$$

2. Paretovo rozdělení s parametrem měřítka $\sigma \leq u$:

$$W^{[u]}_{1,\alpha,0,\sigma} = W_{1,\alpha,0,u}.$$

3. Zobecněné Paretovo rozdělení v $\gamma$-parametrizaci s parametrem polohy $\mu < u$:

$$W^{[u]}_{\gamma,\mu,\sigma} = W_{\gamma,u,\sigma+\gamma(u-\mu)}.$$

### 1.1.8 Limitní věta pro rozdělení vysokých hodnot

Parametrické modelování distribuční funkce $F^{[u]}$ pro velké hodnoty $u$ pomocí zobecněného Paretova rozdělení je založeno na následujícím tvrzení:

**Věta 5** (Balkema, de Haan)**.** *Pokud $F^{[u]}(b_u + a_u\, x)$ má spojitou limitní distribuční funkci pro $u \to x_F$, potom*

$$\lim_{u \to x_F} |F^{[u]}(x) - W_{\gamma,u,\sigma(u)}(x)| = 0, \tag{20}$$

*kde $\sigma(u) > 0$.*

### 1.1.9 Konvergence rozdělení excesů

Konvergence rozdělení excesů k zobecněnému Paretovu rozdělení lze užít k charakterizaci sféry přitažlivosti rozdělení extrémních hodnot. Uvažujme distribuční funkci zobecněného rozdělení extrémních hodnot (6).

**Věta 6.** *Distribuční funkce $F$ patří do sféry přitažlivosti* MDA$(G_\gamma)$ *právě když existuje kladná funkce $\beta(u)$ tak, že*

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F^{(u)}(x) - W_{\gamma,0,\beta(u)}(x)| = 0.$$

Výše uvedené tvrzení říká, že rozdělení, pro která konvergují normovaná maxima k zobecněnému rozdělení extrémních hodnot tvoří množinu, pro kterou příslušná rozdělení excesů konvergují s rostoucí mezí k zobecněnému Paretovu rozdělení. Navíc parametr určující tvar limitního rozdělení excesů je stejný jako parametr limitního rozdělení maxim.

### 1.1.10 Maxima veličin se zobecněným Paretovým rozdělením

Budeme-li uvažovat distribuční funkce (13), (14),resp. (15), snadno nahlédneme, že zobecněné Paretovo rozdělení vždy náleží do sféry přitažlivosti rozdělení extrémních hodnot se shodným parametrem $\alpha$, s distribuční funkcí (1), (2), resp. (3). Platí totiž

$$\lim_{n \to \infty} W_0^n(x + \ln n) = G_0(x), \tag{21}$$

$$\lim_{n \to \infty} W_{1,\alpha}^n(n^{1/\alpha}\, x) = G_{1,\alpha}(x), \tag{22}$$

$$\lim_{n \to \infty} W_{2,\alpha}^n(n^{1/\alpha}\, x) = G_{2,\alpha}(x). \tag{23}$$

## 1.2 Analýza dat v modelech extrémních hodnot

Dále se budeme zabývat vybranými postupy pro statistickou analýzu extrémních hodnot. Analyzovaná data mohou mít podobu blokových maxim nebo hodnot překračujících danou vysokou mez.

### 1.2.1 Metoda blokových maxim pro nezávislá pozorování

Předpokládejme vzájemně nezávislá pozorování pocházející z rozdělení s distribuční funkcí $F$. Nechť $F$ patří do sféry přitažlivosti zobecněného rozdělení extrémních hodnot (8). Z hlediska modelování extrémních jevů má význam zejména Fréchetovo rozdělení, tedy případ, kdy $\gamma > 0$. Data mějme rozdělená do $m$ bloků o velikosti $n$ (může jít například o roční bloky měření prováděných denně). Z těchto bloků použijeme pro analýzu maxima $M_{n1}, \ldots, M_{nm}$. Pro dostatečně velký rozsah bloků $n$ může být rozdělení blokových maxim aproximováno distribuční funkcí (8). K odhadu parametrů této distribuční funkce lze užít metodu maximální věrohodnosti. Příslušná logaritmická věrohodnostní funkce má tvar

$$
\begin{aligned}
l(\gamma, \mu, \sigma; M_{n1}, \ldots, M_{nm}) &= \sum_{i=1}^{m} \ln g_{\gamma,\mu,\sigma}(M_{ni}) \qquad (24) \\
&= -m \ln \sigma \\
&\quad - (1 + 1/\gamma) \sum_{i=1}^{m} \ln \left(1 + \gamma \frac{M_{ni} - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \gamma \frac{M_{ni} - \mu}{\sigma}\right)^{-1/\gamma},
\end{aligned}
$$

kde $g_{\gamma,\mu,\sigma}(x)$ je hustota příslušná distribuční funkci (8). Logaritmická věrohodnostní funkce se maximalizuje za podmínek $\sigma > 0,\ 1 + \gamma\,(M_{ni} - \mu)/\sigma > 0$ pro všechna $i$. Poznamenejme, že v tomto případě množina přípustných hodnot pro odhadované parametry závisí na pozorováních. Nejsou tedy splněny podmínky regularity obvykle užívané pro vyvození žádoucích vlastností odhadů. V literatuře se nicméně uvádí, že pro $\gamma > -1/2$ lze dokázat konsistenci a asymptotickou vydatnost odhadů $\hat{\gamma}, \hat{\mu}, \hat{\sigma}$ získaných maximalizací (24).

Všimněme si rozporu ve volbě optimálního počtu a rozsahu bloků. Velký počet $m$ použitých bloků znamená více dat pro metodu maximální věrohodnosti a vede tak k odhadům s menším rozptylem. Oproti tomu velký rozsah $n$ jednotlivých bloků znamená přesnější přiblížení skutečného rozdělení blokových maxim pomocí zobecněného rozdělení extrémních hodnot a vede tak k menšímu vychýlení výsledných odhadů.

### 1.2.2 Analýza extrémních jevů v modelu blokových maxim

Předpokládejme, že jsme přiblížili rozdělení blokových maxim v datech pomocí distribuční funkce zobecněného rozdělení extrémních hodnot, s parametry odhadnutými například výše zmíněnou metodou maximální věrohodnosti. Odhadnutý model pak můžeme dále využít například pro

1. stanovení velikosti extrémního jevu, který bude v průměru nastávat se zvolenou frekvencí,
2. stanovení průměrné frekvence výskytu extrémního jevu dané úrovně.

Nechť $G$ označuje skutečnou distribuční funkci maxim z bloků o rozsahu $n$. Zaveďme označení

$$r_{n,k} = q_{1-1/k}(G)$$

pro $(1-1/k)$-kvantil rozdělení s distribuční funkcí $G$, tj.

$$\mathrm{P}(M_n > r_{n,k}) = \frac{1}{k}.$$

Hodnota $r_{n,k}$ je tedy definována jako úroveň, která je v průměru překročena maximem v jednom z $k$ bloků. Nazýváme ji *úroveň návratu (return level)*. Pokud přiblížíme distribuční funkci $G$ distribuční funkcí (8) s odhadnutými parametry $\hat{\gamma}, \hat{\mu}, \hat{\sigma}$, odvodíme vyjádření odhadu úrovně návratu ve tvaru

$$\hat{r}_{n,k} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\gamma}} \left( (-\ln(1-1/k))^{-\hat{\gamma}} - 1 \right).$$

Podobně můžeme pro dané $u$ definovat pomocí distribuční funkce $G$ hodnotu

$$k_{n,u} = \frac{1}{\overline{G}(u)} = \frac{1}{1-G(u)},$$

tj.

$$\mathrm{P}(M_n > u) = \frac{1}{k_{n,u}}.$$

Lze tedy říci, že v $k_{n,u}$ blocích o rozsahu $n$ očekáváme výskyt jednoho bloku, v němž dojde k překročení úrovně $u$. Číslo $k_{n,u}$ nazýváme *doba návratu (return period)* a pro jeho odhad použijeme vztah

$$\hat{k}_{n,u} = \frac{1}{\overline{G}_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(u)}.$$

### 1.2.3 Metody založené na překročení meze

Uvažujme distribuční funkci zobecněného Paretova rozdělení (17) s nulovým parametrem polohy:

$$\begin{aligned} W_{\gamma,\sigma}(x) &= 1 - (1+\gamma\, x/\sigma)^{-1/\gamma},\ \gamma \neq 0 \\ W_{0,\sigma}(x) &= 1 - \exp(-x/\sigma), \end{aligned} \tag{25}$$

kde $\sigma > 0$. Rozdělení je definováno pro $x \geq 0$ v případě $\gamma \geq 0$ a pro $0 \leq x \leq -\sigma/\gamma$ pro $\gamma < 0$.

Zavedeme označení

$$e(u) = \mathrm{E}(X-u|X>u) \tag{26}$$

pro střední hodnotu velikosti překročení meze $u$ za podmínky, že k překročení meze dojde. Jde o střední hodnotu rozdělení s distribuční funkcí (18). Z (17) a (18) lze snadno odvodit, že pro náhodnou veličinu se zobecněným Paretovým rozdělením (25) mají excesy nad danou mezí $u$ opět rozdělení typu (18), se stejným parametrem $\gamma$ a parametrem měřítka závislým na $u$:

$$F^{(u)}(x) = W_{\gamma,\sigma(u)}(x),\ \sigma(u) = \sigma + \gamma\, u, \tag{27}$$

kde $0 \leq x < \infty$ v případě $\gamma \geq 0$ a $0 \leq x \leq -(\sigma/\gamma) - u$ v případě $\gamma < 0$. Zároveň lze z (25) odvodit vyjádření střední hodnoty zobecněného Paretova rozdělení

$$\mathrm{E}\, X = \frac{\sigma}{1-\gamma}, \tag{28}$$

která je konečná pro $\gamma < 1$. Z (27) a (28) pak plyne

$$= e(u)\frac{\sigma + \gamma\, u}{1-\gamma},$$

kde $0 \leq u < \infty$ v případě $0 \leq \gamma < 1$, $0 \leq u \leq -\beta/\gamma$ v případě $\gamma < 0$.

Linearita funkce $e(u)$ je vlastností užívanou k identifikaci zobecněného Paretova rozdělení. Z $n$ pozorování můžeme pro danou mez $u$ sestrojit empirický odhad střední hodnoty $e(u)$ ve tvaru

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)\ \mathrm{I}_{[X_i > u]}}{\sum_{i=1}^{n} \mathrm{I}_{[X_i > u]}}. \tag{29}$$

Pak můžeme graficky znázornit body

$$\{(X_{i,n},\ e_n(X_{i,n})), i = 2 \ldots n\},$$

kde $X_{n,n} \leq \cdots \leq X_{1,n}$ jsou pořadové statitiky z pozorování $X_1, \ldots, X_n$. V případě, že pozorování pocházejí ze zobecněného Paretova rozdělení, měl by tento graf vykazovat lineární trend. Místo, kde začíná být průběh grafu přibližně lineární, může případně sloužit k volbě meze $u$ pro účely modelování chvostů rozdělení pomocí zobecněného Paretova modelu. Rostoucí lineární trend bude ukazovat na model s $\gamma > 0$, směřování k horizontálnímu grafu na případ s $\gamma = 0$, klesající trend pak na $\gamma < 0$.

### 1.2.4 Modelování chvostů rozdělení

Mějme $X_1, \ldots, X_n$ pozorování nezávislých stejně rozdělených náhodných veličin. Označme $N_u$ počet pozorování překračujících zvolenou mez $u$. Z hodnot $\tilde{X}_1, \ldots, \tilde{X}_{N_u}$ překračujících mez $u$ získáme velikosti excesů $Y_j = \tilde{X}_j - u$. Pro účel proložení rozdělení (25) pozorovanými excesy lze užít metodu maximální věrohodnosti. Logaritmická věrohodnostní funkce

$$l(\gamma, \sigma; Y_1, \ldots, Y_{N_u}) = -N_u \ln \sigma - (1 + 1/\gamma) \sum_{j=1}^{N_u} \ln \left(1 + \gamma \frac{Y_j}{\sigma}\right) \tag{30}$$

se maximalizuje za podmínek $\sigma > 0$, $1 + \gamma\, Y_j/\sigma > 0$ pro všechna $j$.

Pokud pro danou distribuční funkci $F$ budeme aproximovat rozdělení excesů nad mezí $u$ zobecněným Paretovým rozdělením (25), můžeme psát pro $x \geq u$

$$\overline{F}(x) = \overline{F}(u)\, \overline{F}^{(u)}(x-u) = \overline{F}(u)\ \left(1 + \gamma\, \frac{x-u}{\sigma}\right)^{-1/\gamma}. \tag{31}$$

Inverzí odtud dostaneme vyjádření kvantilu (označovaného také jako *hodnota v riziku* $\mathrm{VaR}_\alpha$) pro $\alpha \geq F(u)$:

$$q_\alpha(F) = u + \frac{\sigma}{\gamma}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\gamma} - 1\right). \tag{32}$$

Z (31) a (32) odvodíme pro $x \geq u$ odhad

$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\gamma}\,\frac{x-u}{\hat{\sigma}}\right)^{-1/\hat{\gamma}} \tag{33}$$

a pro $\alpha > 1 - N_u/n$ odhad

$$\hat{q}_\alpha(F) = u + \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(\frac{1-\alpha}{N_u/n}\right)^{-\hat{\gamma}} - 1\right). \tag{34}$$

V (33) a (34) nahrazujeme parametry zobecněného Paretova rozdělení jejich odhady získanými například maximalizací (30). Pravděpodobnost $\overline{F}(u)$ odhadujeme poměrem $N_u/n$. Předpokládáme, že počet pozorování přesahujících mez $u$ je dostatečný pro spolehlivý odhad $\overline{F}(u)$. Oproti tomu při odhadu $\overline{F}(x)$ pro vysoká $x$ využíváme extrapolace založené na zobecněném Paretově rozdělení.

### 1.2.5 Hillova metoda

Popíšeme přístup vhodný pro odhadování chvostů rozdělení náležejících do sféry přitažlivosti Fréchetova rozdělení (2). Jde o rozdělení charakteristická „těžkými chvosty" (pomalou konvergencí $\overline{F}(x)$ k nule při $x \to \infty$). Sféru přitažlivosti Fréchetova rozdělení charakterizuje Věta 3. Parametr $\alpha$ se nazývá *index chvostu* rozdělení s distribuční funkcí $F \in \mathrm{MDA}(G_{1,\alpha})$. *Hillův odhad* je odhadem indexu $\alpha$ na základě stejně rozdělených pozorování $X_1, \ldots, X_n$. Naznačíme jedno z jeho možných odvození.

Označme

$$\begin{aligned} e^*(\ln u) &= \mathrm{E}(\ln X - \ln u |\ln X > \ln u) \\ &= \frac{1}{\overline{F}(u)} \int_u^\infty (\ln x - \ln u)\, dF(x). \end{aligned}$$

Intergrací per partes a dosazením dle Věty 3 dostaneme vyjádření

$$e^*(\ln u) = \frac{1}{\overline{F}(u)} \int_u^\infty \frac{\overline{F}(x)}{x}\, dx = \frac{1}{\overline{F}(u)} \int_u^\infty L(x)\, x^{-(\alpha+1)}\, dx. \tag{35}$$

Další zjednodušení (35) je založeno na *Karamatově větě* pro pomalu se měnící funkce v nekonečnu.

**Věta 7** (Karamata)**.** *Nechť $L$ je pomalu se měnící funkce, která je lokálně omezená v $[x_0, \infty)$ pro nějaké $x \geq 0$. Potom pro $\kappa < -1$ platí řádová rovnost*

$$\int_x^\infty t^\kappa\, L(t)\ \mathrm{d}t \sim -\frac{1}{\kappa+1}\, x^{\kappa+1}\, L(x),\ \ x \to \infty.$$

Aplikací na (35) dostáváme řádovou rovnost

$$e^*(\ln u) \sim \frac{L(u)\, u^{-\alpha}\, \alpha^{-1}}{\overline{F}(u)} = \alpha^{-1},\ \ u \to \infty. \tag{36}$$

Hillův odhad indexu $\alpha$ je převrácená hodnota odhadu $e_n^*(\ln X_{k,n})$, kde $X_{k,n}$ je $k$-tá největší hodnota v $X_1, \ldots, X_n$.

$$\hat{\alpha}_{k,n}^{(H)} = \left( \frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \right)^{-1}. \tag{37}$$

Odhad (37) závisí na volbě $k$. Doporučuje se grafické znázornění bodů (*Hillův graf*)

$$\{(k,\ \widehat{\alpha}_{k,n}^{(H)}), k = 2 \ldots n\},$$

které slouží k nalezení oblasti, ve které vycházejí podobné hodnoty odhadu počítané z různých počtů pořadových statistik.

Nyní ukážeme, jak lze s pomocí (37) odhadovat pravděpodobnosti $\overline{F}(u)$ pro distribuční funkci ze sféry přitažlivosti Fréchetova rozdělení. Předpokládejme $\overline{F}(x) = C\, x^{-\alpha}$ pro $x \geq u > 0$, kde $u$ je vysoká mez. (Pomalu se měnící funkci nahradíme pro dostatečně velké $x$ konstantou.) Pak můžeme psát $C = u^\alpha\, \overline{F}(u)$, kde index $\alpha$ nahradíme odhadem (37) a $u$ nahradíme pořadovou statistikou $X_{k,n}$. Potom $\overline{F}(u)$ je odhadnuto podílem $k/n$ a dostáváme odhad

$$\widehat{\overline{F}}(x) = \frac{k}{n} \left( \frac{x}{X_{k,n}} \right)^{-\hat{\alpha}_{k,n}^{(H)}}. \tag{38}$$

Pokud označíme $X_{k,n} = u$, $k = N_u$, $\hat{\gamma}^{(H)} = \left(\widehat{\alpha}_{k,n}^{(H)}\right)^{-1}$, můžeme (38) přepsat ve tvaru

$$\widehat{\overline{F}}(x) = \frac{N_u}{n} \left( 1 + \hat{\gamma}^{(H)}\, \frac{x-u}{\hat{\gamma}^{(H)}\, u} \right)^{-1/\hat{\gamma}^{(H)}},$$

který umožňuje srovnání s odhadem (33) založeným na maximálně věrohodných odhadech parametrů zobecněného Paretova rozdělení (při pevně zvolené mezi $u$).

# Literatura


[1] P.Embrechts, C.Klüppelberg, T.Mikosh: Modelling Extremal Events. Springer, 1997.

[2] A.J.McNeil, R.Frey, P.Embrechts: Quantitative Risk Management. Princeton University Press, 2005.

[3] R.D.Reiss, M.Thomas: Statistical Analysis of Extreme Values. Birkhäuser, 1997.