## Stabilita maxim GEV rozdělení

$X_1, \dots, X_n$ je náh. výběr z Gumbelova rozdělení s parametry $\mu$ a $\sigma$. $F(x) = e^{-e^{-\frac{x-\mu}{\sigma}}}, x \in \mathbb{R}$

označme $M_n = \max\{X_1, \dots, X_n\}$. Hledejme jeho distribuční funkci:

$$G_n(x) = [F(x)]^n = \left[e^{-e^{-\frac{x-\mu}{\sigma}}}\right]^n = e^{-n e^{-\frac{x-\mu}{\sigma}}} = e^{-e^{\log n} \cdot e^{-\frac{x-\mu}{\sigma}}} = e^{-e^{\left(\log n - \frac{x-\mu}{\sigma}\right)}} = e^{-e^{-\frac{x-(\mu+\log n)}{\sigma}}}, x \in \mathbb{R}$$

distribuční funkce Gumbelova rozdělení s parametry $\mu^* = \mu + \log n$ a $\sigma^* = \sigma$.

$\left.\begin{array}{l}\text{Podobná vlastnost platí pro náh. výběr z Fréchetova a Weibullova rozdělení, a tedy i pro GEV rozdělení.}\\ \text{A obráceně, tato jsou jediná rozdělení, pro která tato vlastnost platí}\end{array}\right\}$ max-stabilita

## Metoda blokových maxim při praktickém zpracování dat

$X_1, \dots, X_N$ je náh. výběr s distribuční funkcí $F$

Předpokládejme, že $F$ patří do oblasti přitažlivosti GEV rozdělení (tj. jsou splněny předpoklady Fisherovy-Tippettovy věty).

Data rozdělíme do $m$ bloků o velikosti $n$ $(N = m \cdot n)$ a v každém bloku určíme příslušné blokové maximum $M_i = M_i^{(n)}$.

Tedy místo původního náh. výběru $X_1, \dots, X_N$ uvažujeme $m$ blokových maxim $M_1, \dots, M_m$ (nezávislé, stejně rozdělené) a ty budeme modelovat.

Pro dostatečně velkou délku bloku $n$ můžeme rozdělení blokových maxim modelovat pomocí GEV rozdělení.

$\Rightarrow$ $M_1, \dots, M_m$ pokládáme za náh. výběr z GEV rozdělení s parametry $\gamma$, $\mu$ a $\sigma$. A ty budeme dále odhadovat.

<u>Poznámka</u>

Jak volit délku bloku?

- jsou-li data ve formě časové řady, pak se bloky volí „přirozeně" (měsíc, rok, apod.)
- <u>délka</u> bloku $\underline{n}$ musí být dostatečně velká, aby „zafungovala" Fisherova-Tippettova věta
- počet bloků $\underline{m}$ musí být také velký, abychom měli dostatek dat k odhadům parametrů GEV rozdělení

metoda maximální věrohodnosti:

model: $M_1, \dots, M_m$ náh. výběr z GEV rozdělení s distribuční funkcí $G_{\gamma,\mu,\sigma}(x) = \exp\left\{-\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}\right\}$, pro $1+\frac{\gamma(x-\mu)}{\sigma} > 0$.

hustota $g$

$$g_{\gamma,\mu,\sigma}(x) = \exp\left\{-\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}\right\}\cdot(-1)\cdot\left(-\frac{1}{\gamma}\right)\cdot\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}-1}\cdot\frac{\gamma}{\sigma} = \exp\left\{-\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}\right\}\cdot\frac{1}{\sigma}\cdot\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}-1}, \text{ pro } 1+\frac{\gamma(x-\mu)}{\sigma} > 0.$$

$$\log g_{\gamma,\mu,\sigma}(x) = -\log\sigma - \left(1+\frac{1}{\gamma}\right)\cdot\log\left(1+\frac{\gamma(x-\mu)}{\sigma}\right) - \left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}.$$

logaritmická věrohodnostní funkce. $\ell(\gamma,\mu,\sigma) = \sum_{i=1}^{m}\log g_{\gamma,\mu,\sigma}(M_i) = -m\cdot\log\sigma - \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{m}\log\left(1+\frac{\gamma(M_i-\mu)}{\sigma}\right) - \sum_{i=1}^{m}\left(1+\frac{\gamma(M_i-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}$, pro $1+\frac{\gamma(M_i-\mu)}{\sigma}>0 \quad \forall i=1,\dots,m$.

Logaritmickou věrohodnostní funkci maximalizujeme přes všechny hodnoty $\sigma>0$, $\mu$ a $\sigma$ takové, že $1+\frac{\gamma(M_1-\mu)}{\sigma}>0,\dots,1+\frac{\gamma(M_2-\mu)}{\sigma}>0$.

$\ell(\gamma,\mu,\sigma)$ není diferencovatelná, tedy je třeba provádět maximalizaci přímo $\Rightarrow$ maximálně věrohodný odhad $(\hat{\gamma},\hat{\mu},\hat{\sigma})^T$.

Navíc, nejsou splněny podmínky regularity pro odvození asymptotických vlastností.

__Poznámka__

V literatuře je pro $\gamma>-\frac{1}{2}$ dokázáno, že $\hat{\Theta}=(\hat{\gamma},\hat{\mu},\hat{\sigma})^T$ je asymptoticky nestranný, konzistentní odhad parametru $\Theta=(\gamma,\mu,\sigma)^T$.

Navíc $\hat{\Theta} \approx N_3\left(\Theta, \frac{1}{m}\cdot J^{-1}(\Theta)\right)\dots \Rightarrow$ můžeme konstruovat IS pro $\gamma,\mu$ a $\sigma$.

Je-li $-1<\gamma\le-0{,}5$, pak maximálně věrohodné odhady existují, ale nemají „hezké" vlastnosti.

Pro $\gamma\le-1$ MLE obecně neexistují.

metoda pravděpodobnostně vážených momentů (PWM)

__Definice__

Nechť $X$ je náh. veličina s distribuční funkcí $F$. Pak čísla $M_{p,r,s}=E\left[X^p[F(X)]^r\cdot[1-F(X)]^s\right]$ pro $p,r,s\in\mathbb{R}$

jsou pravděpodobnostně vážené momenty.

speciální případ: $p=1, s=0$

$$\beta_r = M_{1,r,0} = E\left[X \cdot [F(X)]^r\right] \quad \text{pro } r = 0,1,2.$$

pro GEV rozdělení platí: $\beta_r = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\left[1-(r+1)^{\gamma}\,\Gamma(1-\gamma)\right]\right\}$ pro $\gamma<1$, $\gamma\neq 0$.

odhad $\beta_r$ je $\hat{\beta}_r = \frac{1}{m}\sum_{j=1}^{m}\left(\prod_{l=1}^{r}\frac{j-l}{m-l}\right)\cdot M_{(j)}$ pro $r=1,2$ a $\hat{\beta}_0 = \frac{1}{m}\sum_{j=1}^{m} M_j$.

Odhady $(\gamma,\mu,\sigma)^T$ získáme jako řešení soustavy $\beta_r = \hat{\beta}_r$ pro $r=0,1,2$.

<u>Úlohy</u>

1.) Odhadnout $P(X_i > x)$ pro $x$ velké

$M_1 = \max\{X_1, \dots, X_n\}$

$$\left.\begin{array}{l} P(M_1 \leq x) = [P(X_i \leq x)]^n = [1-P(X_i > x)]^n \\ \quad \shortparallel \\ G_{\gamma,\mu,\sigma}(x) \end{array}\right\} \Rightarrow P(X_i > x) = 1-\left[G_{\gamma,\mu,\sigma}(x)\right]^{\frac{1}{n}}.$$

$$\widehat{P(X_i > x)} = 1 - \left[G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(x)\right]^{\frac{1}{m}} = 1 - \left[\exp\left\{-\left(1+\frac{\hat{\gamma}(x-\hat{\mu})}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}\right\}\right]^{\frac{1}{m}} = 1 - \exp\left\{-\frac{1}{m}\left(1+\frac{\hat{\gamma}(x-\hat{\mu})}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}\right\}$$

uvažujeme-li aproximaci $e^x \doteq 1+x$, pak $\widehat{P(X_i > x)} \doteq \frac{1}{m}\left(1+\frac{\hat{\gamma}(x-\hat{\mu})}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}$.

2.) Odhad extrémních kvantilů

chápeme-li data jako časovou řadu, pak chceme odhadnout velikost extrémního jevu, který bude nastávat v průměru se zvolenou frekvencí (úroveň návratu), tj. $1\times$ za $k$ časových okamžiků, tedy hledáme $(1-\frac{1}{k})$-kvantil příslušného GEV rozdělení, označme jej $q_{1-\frac{1}{k}}$.

Tedy platí $P(M_1 \leq q_{1-\frac{1}{k}}) = 1 - \frac{1}{k} \iff P(M_1 > \underbrace{q_{1-\frac{1}{k}}}_{\text{hledaná úroveň návratu}}) = \frac{1}{k}$.

pro přehlednost označme $q := q_{1-\frac{1}{k}}$.

$$1 - G_{\gamma,\mu,\sigma}(q) = \frac{1}{k}$$

$$1 - \exp\left\{-\left(1+\frac{\gamma(q-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}\right\} = \frac{1}{k}$$

$$-\left(1+\frac{\gamma(q-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}} = \log\left(1-\frac{1}{k}\right)$$

$$1+\frac{\gamma(q-\mu)}{\sigma}=\left[-\log\left(1-\frac{1}{k}\right)\right]^{-\gamma}$$

$$q=\frac{\sigma}{\gamma}\left\{\left[-\log\left(1-\frac{1}{k}\right)\right]^{-\gamma}-1\right\}+\mu$$

odhad úrovně návratu získáme dosazením odhadů parametrů: $\hat{q}=\hat{q}_{1-\frac{1}{k}}=\frac{\hat{\sigma}}{\hat{\gamma}}\left\{\left[-\log\left(1-\frac{1}{k}\right)\right]^{-\hat{\gamma}}-1\right\}+\hat{\mu}$.

3) Stanovit průměrnou frekvenci výskytu extrémního jevu dané úrovně (doba návratu)

– známe úroveň $q_{1-\frac{1}{k}}$ a hledáme příslušné $\underline{k}$.

$$P\left(M_1>q_{1-\frac{1}{k}}\right)=\frac{1}{k}\quad\Rightarrow\quad k=\frac{1}{1-G_{\gamma,\mu,\sigma}\left(q_{1-\frac{1}{k}}\right)}.$$

$$\parallel$$

$$1-G_{\gamma,\mu,\sigma}\left(q_{1-\frac{1}{k}}\right)$$

odhad doby návratu získáme opět dosazením odhadů: $\hat{k}=\frac{1}{1-G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}\left(q_{1-\frac{1}{k}}\right)}=\frac{1}{1-\exp\left\{-\left(1+\frac{\hat{\gamma}\left(q_{1-\frac{1}{k}}-\hat{\mu}\right)}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}\right\}}$.