ODHADY NÁVRATOVÝCH HODNOT KLIMATOLOGICKÝCH DAT

Jan Picek

Katedra aplikované matematiky

Fakulta přírodovědně-humanitní a pedagogická

Technická univerzita v Liberci

Robust 2018

ÚVOD

Velká pozornost v analýze extrémních dat (např. záplavy) je věnována odhadům T-leté úrovně (návratová hodnota, T-letá voda).

Představa: úroveň opakující se v průměru jednou za T let.

Úvod

Velká pozornost v analýze extrémních dat (např. záplavy) je věnována odhadům T-leté úrovně (návratová hodnota, T-letá voda).

Představa: úroveň opakující se v průměru jednou za T let.

Z pohledu statistiky: vysoký kvantil rozdělení náhodné veličiny (průtoku).

$$u(T) = F^{-1}\left(1 - \frac{1}{T}\right)$$

$$P\left(X > u(T)\right) = 1 - F(u(T)) = \frac{1}{T}$$

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9

Maximální dosažená hodnota za sledované období 36.2.

Nechť $X_1, \ X_2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F.

Nechť $M_n = \max(X_1, \ldots, X_n)$.

Předpokládejme, že existuje posloupnost reálných čísel $a_n > 0$ a b_n tak, že posloupnost $(M_n - b_n)/a_n$ konverguje v distribuci, t.j.

$$P\left((M_n - b_n)/a_n \leq x\right) = F^n(a_n x + b_n) \to G(x), \ n \to \infty,$$

pro nějakou nedegenerovanou d.f. $G(x)$

Jestliže podmínka platí, říkáme, že F je ve sféře přitažlivosti G (maximum domain of attraction), $F \in \text{MDA}(G)$.

ZÁKLADNÍ PRINCIPY

FISHEROVA-TIPPETTOVA VĚTA (1928)

Jestliže $F \in$ MDA(G) potom G je typu jedné z následujících tří d.f.

Fréchet $\quad \Phi_{1/\gamma}(x) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-1/\gamma}\right), & x > 0 \end{cases} \quad \gamma > 0$

Weibull $\quad \Psi_{1/\gamma}(x) = \begin{cases} \exp\left\{-(-x)^{1/\gamma}\right\}, & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \gamma > 0$

Gumbel $\quad \Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$

Základní principy

Fisherova-Tippettova věta (1928)

Jestliže $F \in \text{MDA}(G)$ potom G je typu jedné z následujících tří d.f.

Fréchet $\Phi_{1/\gamma}(x) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-1/\gamma}\right), & x > 0 \end{cases} \quad \gamma > 0$

Weibull $\Psi_{1/\gamma}(x) = \begin{cases} \exp\left\{-(-x)^{1/\gamma}\right\}, & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \gamma > 0$

Gumbel $\Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$

Gnėděnko (1943)

Limitní rozdělení je zobecněné rozdělení extrémních hodnot.

$$G(x) = G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & \gamma \neq 0 \\ \exp(-e^{-x}) & \gamma = 0 \end{cases},$$

kde $1 + \gamma x > 0$

G je určena jednoznačně až na parametr polohy a měřítka.

Method of block maxima

Předpokládáme, že data rozdělíme do bloků obsahující n (velké) hodnot, bereme maximum v každém bloku a využijeme limitní výsledek, t.j. GEV rozdělení.

Užití limitního rozdělení:

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G_\gamma(x).$$

$$y = a_n x + b_n \Rightarrow P(M_n \leq y) \approx G_\gamma\left(\tfrac{y-b_n}{a_n}\right) = G_{\gamma,b_n,a_n}(y).$$

Parametry odhadneme, např. metodou maximální věrohodnosti

Method of block maxima

$$L(b,a,\gamma) = -m\log a - (1+1/\gamma)\sum_{i=1}^{m}\log\left(1+\gamma\left(\frac{z_i-b}{a}\right)\right) - \sum_{i=1}^{m}\left[1+\gamma\left(\frac{z_i-b}{a}\right)\right]^{-1/\gamma}$$

pro $\gamma = 0$

$$L(b,a) = -m\log a - \sum_{i=1}^{m}\left(\frac{z_i-b}{a}\right) - \sum_{i=1}^{m}\exp\left\{\left(\frac{z_i-b}{a}\right)\right\}.$$

Neexistuje analytické řešení.

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9

Maximální dosažená hodnota za sledované období 36.2.

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9
Gumbelovo rozdělení				
metoda max. věrohodnosti	34.4	35.3	38.4	42.2

Maximální dosažená hodnota za sledované období 36.2.

L-momentová metoda

Pro malé a střední rozsahy výběrů odhady lepší vlastnosti než metoda maximální věrohodnosti *např. simulační studie (Hosking, Wallis, Wood) ukazuje, že pro všechna* k *GEV z intervalu (-0.5,0.5) a rozsah výběru do 100 mají odhady menší či srovnatelnou střední kvadratickou chybu ve srovnání s odhady maximální věrohodností*

L-MOMENTOVÁ METODA

Nechť $X_1, X_2, \dots X_n$ je náhodný výběr s distribuční funkcí $F(x)$ a kvantilovou funkcí $Q(u)$ a nechť $X_{1:n} \leq X_{2:n} \leq X_{n:n}$ jsou pořádkové statistiky.

L-momenty:

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \dots$$

$$EX_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int x \left(F(x)\right)^{j-1} \left(1 - F(x)\right)^{r-j} \mathrm{d}F(x)$$

$$\lambda_1 = EX = \int_0^1 Q(u)du$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 Q(u)(2u-1)du$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 Q(u)(6u^2 - 6u + 1)du$$

Příklady L-momentů některých rozdělení:

Rovnoměrné na (a, b) $\lambda_1 = \frac{1}{2}(a+b), \lambda_2 = \frac{1}{6}(b-a), \tau_3 = 0, \tau_4 = 0$

Normální $\mathcal{N}(\mu, \sigma^2)$ $\lambda_1 = \mu, \lambda_2 = \frac{\sigma}{\pi}, \tau_3 = 0, \tau_4 = 0.1226$

Gumbelovo rozdělení $F(x) = \exp[-\exp(-(x-\xi)/\alpha)]$

$\lambda_1 = \xi + \alpha\gamma, \lambda_2 = \alpha \log 2, \tau_3 = 0.1699,$
$\tau_4 = 0.1504, \gamma = 0.5772...$ konst.

Zobecněné rozdělení extrémních hodnot (GEV)

$F(x) = \exp[-(1-k(x-\xi)/\alpha)^{\frac{1}{k}}]$
$\lambda_1 = \xi + \alpha(1-\Gamma(1+k))/k,$
$\lambda_2 = \alpha(1-2^{-k})\Gamma(1+k)/k,$
$\tau_3 = 2(1-3^{-k})/(1-2^{-k}) - 3, \tau_4 = \dots$
$k > -1,\ \Gamma(.)$ označuje gamma funkci

Odhady: Výběrový $L-$moment:

$$l_r = \binom{n}{r}^{-1} \sum\sum\ldots\sum_{1\le i_1<i_2<\ldots<i_r\le n} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{i_{r-k}:n},$$

$r = 1, 2, \ldots, n.$

Speciálně:

$$l_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad l_2 = \frac{1}{2}\binom{n}{2}^{-1} \sum\sum_{i>j} (X_{i:n} - X_{j:n})$$

$$l_3 = \frac{1}{3}\binom{n}{3}^{-1} \sum\sum\sum_{i>j>k} (X_{i:n} - 2X_{j:n} + X_{k:n})$$

$$l_4 = \frac{1}{4}\binom{n}{4}^{-1} \sum\sum\sum\sum_{i>j>k>l} (X_{i:n} - 3X_{j:n} + 3X_{k:n} - X_{l:n})$$

Odhady paramterů – L-momentová metoda

Rovnoměrné na (a, b) $\hat{a} = l_1 - 3l_2, \hat{a} = l_1 + 3l_2$

Normální $\mathcal{N}(\mu, \sigma^2)$ $\hat{\mu} = l_1, = \hat{\sigma} = \pi^{1/2} l_2$

Gumbelovo rozdělení $F(x) = \exp[-\exp(-(x - \xi)/\alpha)]$

$\hat{\xi} = l_1 - \hat{\alpha}\gamma, \hat{\alpha} = l_2 / \log 2$
$\gamma = 0.5772...$ konst.

Zobecněné rozdělení extrémních hodnot (GEV)

$F(x) = \exp[-(1 - k(x - \xi)/\alpha)^{\frac{1}{k}}]$
$z = 2/(3 + t_3) - \log 2 / \log 3,$
$\hat{k} = 7.8590z + 2.9554z^2,$
$\hat{\alpha} = l_2 \hat{k} / [(1 - 2^{-\hat{k}})\Gamma(1 + \hat{k})],$
$\hat{\xi} = l_1 + \hat{\alpha}[\Gamma(1 + \hat{k}) - 1]/\hat{k}$

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9
L-momentová metoda	34.2	35.0	36.7	38.6

Maximální dosažená hodnota za sledované období 36.2.

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9
L-momentová metoda	34.2	35.0	36.7	38.6
Gumbelovo rozdělení				
metoda max. věrohodnosti	34.4	35.3	38.4	42.2
L-momentová metoda	34.2	35.0	37.8	41.4

Maximální dosažená hodnota za sledované období 36.2.

Nechť $X_1,\ X_2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny s ditr. funkcí F. Je "rozumné" zahrnovat všechny hodnoty překračující daný vysoký práh (threshold) u. Chování extrémních událostí je dáno podmíněnou pravděpodobností $P\left(X_i > y | X_i > u\right)$ a

$$P\left(X_i < y | X_i > u\right) \rightarrow H(y),\ u \rightarrow u_{end},$$

zobecněné Paretovo rozdělení

$$H(x) = \begin{cases} 1 - \left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\gamma} & \gamma \neq 0 \\ 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)} & \gamma = 0 \end{cases},$$

kde $1 + \gamma\left(\frac{x-\mu}{\sigma}\right) > 0$.

Uvažujeme hodnoty větší než dostatečně vysoký práh (threshold) a předpokládáme, že asymptotický výsledek je přibližně pravdivý, tj. užijeme zobecněné Paretovo rozdělení jako vhodný model.

Metoda je známa jako ‘peaks-over-threshold’ (POT).

Příklad: Roční maxima teploty vzduchu za období 1961-2007

Stanice Liberec

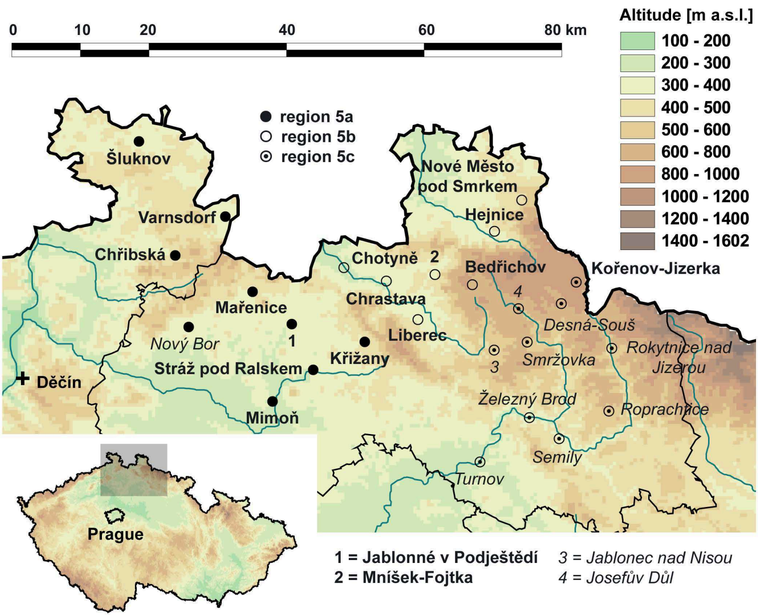
Návratová hodnota (roky)	10	20	100	1000
GEV rozdělení				
metoda max. věrohodnosti	34.1	34.9	36.3	37.9
L-momentová metoda	34.2	35.0	36.7	38.6
POT				
treshold 26.50	34.1	34.6	35.5	36.7
treshold 29.37	34.3	34.9	36.1	37.5

Maximální dosažená hodnota za sledované období 36.2.

Srážky - Liberecko

Kyselý, Jan ; Gaál, L. ; Picek, J. ; Schindler, M., 2013: Return periods of the August 2010 heavy precipitation in northern Bohemia (Czech Republic) in the present climate and under climate change, Journal of Water and Climate Change, 4, pp. 265-286

Srážky - Liberecko



1 = Jablonné v Podještědí
2 = Mníšek-Fojtka
3 = Jablonec nad Nisou
4 = Josefův Důl

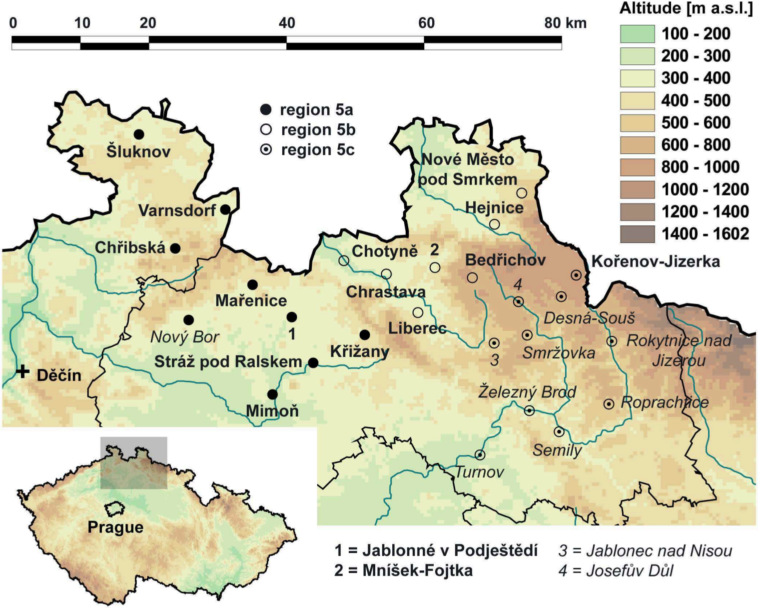
Srážky - Liberecko

Stanice	srpen 2010	roky (bez)	roky (s)
Hejnice	179.0	156.9	78.2
Mníšek	160.0	305.4	89.9
Chrastava	135.5	1337.9	66.1
Mařenice	124.2	1250.8	153.0
Bedřichov	112.0	17.3	15.9
Liberec	98.9	45.0	35.2

Srážky - Liberecko

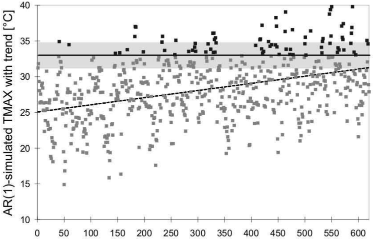
Stanice	srpen 2010	roky (bez)	roky (s)
Hejnice	179.0	156.9 (53.4 - 24 809)	78.2 (36.8 - 2 612.5)
Mníšek	160.0	305.4 (67.8 - 95 119)	89.9 (38.2 - 8 117.3)
Chrastava	135.5	1337.9 (87.1 - 1.848x10^6)	66.1 (28.9 - 4 427.8)
Mařenice	124.2	1250.8 (140.6 - 1.608x10^6)	153.0 (54.6 - 23 530)
Bedřichov	112.0	17.3 (13.1 - 42.7)	15.9 (12.3 - 36.2)
Liberec	98.9	45.0 (24.8 - 535.1)	35.2 (21.4 - 200.6)

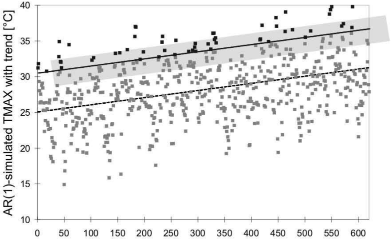
Srážky - Liberecko



1 = Jablonné v Podještědí
2 = Mníšek-Fojtka
3 = Jablonec nad Nisou
4 = Josefův Důl

Trend v datech





ZÁVĚR

?

Děkuji za pozornost.