Asymptotické pravděpodobnostní rozdělení výběrového maxima

Jana Kahounová [*]

Teorie extrémních hodnot byla nejčastěji používána pro modelování rozdělení různých přírodních jevů jako jsou dešťové srážky, větrné poryvy, záplavy, znečištění vzduchu apod. Rozdělení výběrového maxima nás zajímá např. při analýze vlivu maximální tíhy sněhu na stavební konstrukce, při studiu únavy nebo pevnosti materiálu nás může zajímat rozdělení výběrového minima. Teorie extrémních hodnot vznikla na základě praktických potřeb zejména astronomů, hydrologů a techniků. Nejprve přitahovala pozornost matematiků, zabývajících se teorií pravděpodobnosti a později i statistiků. V současné době tato teorie nachází uplatnění i v ekonomické oblasti, využívá se např. jako metoda modelování a měření extrémních rizik ve finanční sféře.

Jako rozdělení extrémních hodnot jsou obecně uvažovány tři systémy rozdělení, které jsou popsány distribuční funkcí $F(x)$:

Typ 1 – Gumbelovo rozdělení

$$F(x) = \exp\left(-e^{-\frac{x-\alpha}{\beta}}\right), \quad -\infty < x < \infty. \tag{1}$$

Typ 2 – Fréchetovo rozdělení

$$\begin{aligned} F(x) &= 0, && x < \alpha, \\ &= \exp\left[-\left(\frac{x-\alpha}{\beta}\right)^{-k}\right], && x \ge \alpha. \end{aligned} \tag{2}$$

Typ 3 – Weibullovo rozdělení

$$\begin{aligned} F(x) &= \exp\left[-\left(\frac{\alpha - x}{\beta}\right)^{k}\right], && x \le \alpha, \\ &= 1, && x > \alpha, \end{aligned} \tag{3}$$

kde α, $\beta > 0$ a $k > 0$ jsou parametry. Tato rozdělení jsou také někdy nazývána Fisherova – Tippettova rozdělení extrémních hodnot (Fisher – Tippett theorem viz např. Embrechts [1]).

V tomto článku se budeme podrobněji zabývat Gumbelovým rozdělením (normalizovaného) výběrového maxima (1). Uvažujeme posloupnost nezávislých, identicky rozdělených náhodných veličin X_i, $i = 1, 2, \ldots, n, \ldots$ s distribuční funkcí $F(x)$. Výběrové maximum je pořádková statistika $X = X_{(n)} = max(X_1, X_2, \ldots, X_n)$, $n \ge 1$. Rozdělení extrémních hodnot

* Doc Ing. Jana Kahounová, CSc.; Katedra statistiky a pravděpodobnosti, Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze, kahoun@vse.cz.

byla odvozena jako limitní rozdělení největších (nebo nejmenších) hodnot náhodného výběru o rozsahu $n \to \infty$. Za tímto účelem je třeba provést lineární transformaci s koeficienty, které závisí na výběrovém rozsahu. Tento proces je analogický s normalizací, která se používá při důkazu centrálního limitního teorému.

Nahradíme-li X ve vztazích (1) až (3) veličinou $(-X)$, jedná se o rozdělení výběrového minima, tj. pořádkové statistiky $X_{(1)} = min(X_1, X_2, \ldots, X_n)$. Toto rozdělení patří rovněž mezi rozdělení extrémních hodnot.

Gumbelovo rozdělení výběrového maxima

S přihlédnutím k (1) má hustota pravděpodobnosti Gumbelova rozdělení tvar

$$f(x) = \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}} \exp\left[-e^{-\frac{x-\alpha}{\beta}}\right], \quad -\infty < x < \infty, \qquad (4)$$

$$-\infty < \alpha < \infty,$$

$$0 < \beta < \infty.$$

Jestliže $\alpha = 0$ a $\beta = 1$ nebo ekvivalentně, provedeme-li transformaci $Y = \dfrac{X-\alpha}{\beta}$ dostáváme rozdělení (1) ve standardním tvaru s distribuční funkcí

$$F(y) = \exp(-e^{-y}), \qquad (5)$$

resp. s hustotou pravděpodobnosti

$$f(y) = \exp(-y - e^{-y}), \quad -\infty < y < \infty. \qquad (6)$$

Pak náhodná veličina

$$Z = \exp\left(-\frac{X-\alpha}{\beta}\right) = e^{-Y}$$

má zřejmě exponenciální rozdělení s hustotou pravděpodobnosti

$$f(z) = e^{-z}, \quad z \ge 0. \qquad (7)$$

Momentovou vytvořující funkci náhodné veličiny Y s pomocnou proměnnou t označíme symbolem $m_Y(t)$ a pro tuto funkci platí

$$m_Y(t) = E\left[e^{\frac{t(X-\alpha)}{\beta}}\right] = E\left[\left(e^{-\frac{X-\alpha}{\beta}}\right)^{-t}\right] = E(Z^{-t}) = \int_0^\infty z^{-t} e^{-z} dz = \Gamma(1-t), \quad t < 1. \qquad (8)$$

Momentová vytvořující funkce náhodné veličiny $X = \beta Y + \alpha$ je tedy

$$m_X(t) = E\left[e^{t(\beta Y + \alpha)}\right] = e^{t\alpha} m_Y(t\beta) = e^{t\alpha}\Gamma(1-t\beta), \quad |t|\beta < 1. \qquad (9)$$

Vytvořující funkce kumulantů náhodné veličiny X pak je

$$k_X(t) = \ln m_X(t) = t\alpha + \ln\Gamma(1-t\beta). \qquad (10)$$

Pomocí funkce (10) určíme kumulant 1. řádu, tj. střední hodnotu E(X)

$$\kappa_1(X) = E(X) = k'_X(0) = \alpha + \frac{d}{dt}\ln\Gamma(1-t\beta)\Big|_{t=0} = \alpha - \beta\frac{\Gamma'(1)}{\Gamma(1)} = \alpha - \beta\psi(1).$$

Funkce $\psi(x)$ je tzv. digamma funkce

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

a $\psi(1)$ je digamma funkce v bodě x = 1, přičemž $\psi(1) = \Gamma'(1) = -\gamma$, kde γ je Eulerova – Mascheroniho konstanta $\gamma \doteq 0{,}57722$. Takže

$$\kappa_1(X) = \alpha + \beta\gamma \doteq \alpha + 0{,}57722\beta. \tag{11}$$

Obdobně kumulant 2. řádu

$$\kappa_2(X) = \frac{d^2}{dt^2}\ln\Gamma(1-t\beta)\Big|_{t=0} = (-\beta)^2\psi_1(1),$$

kde $\psi_1(x)$ je tzv. trigamma funkce definovaná vztahem

$$\psi_1(x) = \frac{d^2}{dx^2}\ln\Gamma(x)$$

a $\psi_1(1)$ je trigamma funkce v bodě $x = 1$

$$\psi_1(1) = \frac{\pi^2}{6},$$

takže

$$\kappa_2(X) = \frac{\beta^2\pi^2}{6}. \tag{12}$$

Rozptyl $D(X)$ určíme pomocí známých vztahů

$$E(X^2) = \kappa_2(X) + [\kappa_1(X)]^2, \tag{13}$$

$$D(X) = E(X^2) - E^2(X). \tag{14}$$

Dosazením (11) a (12) do vztahu (13) dostáváme

$$E(X^2) = \frac{\beta^2\pi^2}{6} + (\alpha + \beta\gamma)^2 \tag{15}$$

a podle (14)

$$D(X) = \frac{\beta^2\pi^2}{6} + (\alpha+\beta\gamma)^2 - (\alpha+\beta\gamma)^2 = \kappa_2(X) = \frac{\beta^2\pi^2}{6} = 1{,}64493\beta^2. \tag{16}$$

Obecně pro kumulant r-tého řádu platí

$$\kappa_r(X) = (-\beta)^r\psi_{r-1}(1), \; r \geq 2, \tag{17}$$

kde $\psi_{r-1}(1)$ je polygamma funkce (r – 1). řádu (definovaná jako r-tá derivace přirozeného logaritmu gamma funkce) v bodě x = 1.

Obecné momenty, resp. momentové charakteristiky Gumbelova rozdělení můžeme stanovit i „klasickým způsobem". Pro n-tý obecný moment platí

$$E(X^n) = \frac{1}{\beta}\int_{-\infty}^{\infty} x^n e^{\frac{\alpha - x}{\beta}} \exp\left(-e^{\frac{\alpha - x}{\beta}}\right) dx. \qquad (18)$$

Uvažujeme-li v (18) substituci

$$u = \exp\left(\frac{\alpha - x}{\beta}\right)$$

dostáváme

$$E(X^n) = -\int_{\infty}^{0} (\alpha - \beta \ln u)^n e^{-u} du = \int_{0}^{\infty} (\alpha - \beta \ln u)^n e^{-u} du = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \alpha^{n-k} \beta^k \int_{0}^{\infty} (\ln u)^k e^{-u} du =$$
$$= \sum_{k=0}^{n} \binom{n}{k} (-1)^k \alpha^{n-k} \beta^k I_k, \qquad (19)$$

kde

$$I_k = (-1)^k \int_{0}^{\infty} (\ln u)^k e^{-u} du = (-1)^k \Gamma^{(k)}(1), \qquad (20)$$

$\Gamma^{(k)}(u)$ je k-tá derivace gamma funkce, jsou tzv. Eulerovy – Mascheroniho integrály.

Jestliže k = 0, pak

$$I_0 = \int_{0}^{\infty} e^{-u} du = 1,$$

k = 1

$$I_1 = \int_{0}^{\infty} \ln u e^{-u} du = \gamma,$$

k = 2

$$I_2 = \int_{0}^{\infty} (\ln u)^2 e^{-u} du = \gamma^2 + \frac{\pi^2}{6},$$

takže podle (19) dostáváme

$$E(X) = \alpha \int_{0}^{\infty} e^{-u} du - \beta \int_{0}^{\infty} \ln u e^{-u} du = \alpha + \beta\gamma,$$

$$E(X^2) = \alpha^2 \int_{0}^{\infty} e^{-u} du - 2\alpha\beta \int_{0}^{\infty} \ln u e^{-u} du + \beta^2 \int_{0}^{\infty} (\ln u)^2 e^{-u} du = \alpha^2 + 2\alpha\beta\gamma + \beta^2 \left(\gamma^2 + \frac{\pi^2}{6}\right) =$$
$$= (\alpha + \beta\gamma)^2 + \frac{\beta^2 \pi^2}{6}$$

a rozptyl $D(X)$ je (16).

Odhad kvantilu asymptotického rozdělení maximální ztráty

Z interních zdrojů jisté banky máme k dispozici $n = 252$ nezávislých pozorování ztráty v Kč, která vznikla v důsledku technických havárií a živelných pohrom. Jedná se o ztráty způsobené přírodními katastrofami, krizovými stavy, selháním transportu a externím selháním telekomunikací, dopravními nehodami apod.

Z výběrových dat byl stanoven výběrový průměr

$$\bar{x} = 63\ 091{,}6000$$

a výběrová směrodatná odchylka

$$s = 66\ 184{,}0913.$$

Pro určení odhadu kvantilu rozdělení maximální ztráty je třeba nejprve odhadnout parametry α a β rozdělení (4).

Momentové odhady α^+ a β^+ získáme řešením soustavy rovnic

$$\bar{x} = \alpha^+ + 0{,}57722\beta^+, \tag{21}$$

$$s^2 = \frac{1}{6}\pi^2 (\beta^+)^2,$$

s výsledkem

$$\beta^+ = \frac{\sqrt{6}}{\pi} s, \tag{22}$$

$$\alpha^+ = \bar{x} - 0{,}57722\frac{\sqrt{6}}{\pi} s, \tag{23}$$

tj. v uvažovaném případě

$$\beta^+ \doteq 51\ 603{,}524,$$

$$\alpha^+ \doteq 33\ 305{,}014.$$

Podle [2] je Raova-Cramérova dolní mez rozptylů nezkreslených odhadů parametru α

$$\frac{1{,}10867\beta^2}{n}$$

a parametru β

$$\frac{0{,}60793\beta^2}{n}.$$

V [4] je dokázáno, že rozptyl odhadu α^+, resp. β^+ je

$$D(\alpha^+) = \frac{1{,}1678\beta^2}{n},$$

resp.

$$D(\beta^+) = \frac{1{,}1\beta^2}{n}.$$

Relativní vydatnost odhadu α^+ je tedy 95%, zatímco u odhadu β^+ je pouze 55%. Oba odhady jsou konzistentní.

Metoda maximální věrohodnosti vede k odhadům s výhodnými vlastnostmi. Věrohodnostní funkce má tvar

$$L(\alpha,\beta) = \beta^{-n} e^{-\frac{\sum_{i=1}^{n}(x_i-\alpha)}{\beta}} \exp\left(-\sum_{i=1}^{n} e^{-\frac{x_i-\alpha}{\beta}}\right). \tag{24}$$

Maximálně věrohodné odhady $\hat{\alpha}, \hat{\beta},$ jsou řešením soustavy věrohodnostních rovnic

$$n - \sum_{i=1}^{n} e^{-\frac{x_i-\hat{\alpha}}{\hat{\beta}}} = 0,$$

$$-n\hat{\beta} + \sum_{i=1}^{n}(x_i - \hat{\alpha})\left(1 - e^{-\frac{x_i-\hat{\alpha}}{\hat{\beta}}}\right) = 0.$$

Po úpravě

$$\hat{\alpha} = \hat{\beta}\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{x_i}{\hat{\beta}}}\right), \tag{25}$$

$$\hat{\beta} = \bar{x} - \frac{\sum_{i=1}^{n} x_i e^{-\frac{x_i}{\hat{\beta}}}}{\sum_{i=1}^{n} e^{-\frac{x_i}{\hat{\beta}}}}. \tag{26}$$

Rovnici (26) je třeba řešit numericky. Použili jsme Newtonovu metodu. Jako počáteční aproximaci β_0 jsme zvolili momentový odhad β^+ a jako zastavovací podmínku $|\beta_{k+1} - \beta_k| < 10^{-12}$. Newtonova metoda konvergovala rychle a stačilo provést 5 iterací a

$$\hat{\beta} = 41\,447{,}527.$$

Dosazením hodnoty $\hat{\beta}$ do (25) získáme

$$\hat{\alpha} = 35\,448{,}540.$$

Kvantilová funkce *Q(P)* spojitých náhodných veličin je pro $0 < P < 1$ inverzní funkcí k funkci distribuční

$$Q(P) = F^{-1}(P). \tag{27}$$

Tato funkce udává přímo hodnoty kvantilů x_p pro dosazená P. Jestliže distribuční funkce má tvar (1), pak zřejmě

$$Q(P) = \alpha - \beta\ln(-\ln P) \tag{28}$$

a maximálně věrohodný odhad 100 P %-ního kvantilu

$$\hat{x}(P) = \hat{\alpha} - \hat{\beta}\ln(-\ln P), \quad 0 < P < 1. \tag{29}$$

Chceme-li nyní odhadnout hodnotu, kterou maximální ztráta banky nepřekročí s vysokou pravděpodobností, řekněme 0,99, pak podle (29)

$$\hat{x}_{0,99} = 35\ 448{,}540 - 41\ 447{,}527\ \ln(-\ln 0{,}99) \doteq 226\ 112{,}96.$$

Závěr

Teorie extrémních hodnot se zabývá dvěma základními typy modelů. Jsou to modely limitního rozdělení normalizovaného maxima (minima) a tzv. POT (peaks – over – threshold) modely, což jsou modely pro všechna velká pozorování, která překročí určitý vysoký práh (threshold).

V tomto článku se zabýváme Gumbelovým rozdělením výběrového maxima. Tento model byl aplikován na rozdělení náhodné veličiny – maximální ztráta banky, která vznikla v důsledku technických havárií a živelných pohrom. Cílem bylo odhadnout hodnotu, kterou maximální ztráta „prakticky jistě“ (P = 0,99) nepřekročí.

Literatura

EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. 1997. *Modelling Extremal Events.* Wien : Springer Verlag, 1997.

DOWNTON, F., 1966. Linear estimates of parameters in the extreme value distribution. *Technometrics*, 1966, 8, s. 3–17.

JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N., 1995. *Continuous Univariate Distributions, Second Edition.* New York : John Wiley & Sons, Inc, 1995.

TIAGO DE OLIVEIRA, J., 1963. *Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation. Trabajos de Estadística*, 1963, 14, s. 61–81.

Asymptotic Probability Distribution of Sample Maximum

Abstract

Extreme value theory is the most scientific approach to an inherently difficult problem – predicting the possibility that an extreme event will occur. Broadly speaking, there are two kinds of models for extreme values. The first group of models are models for a distribution of normalized maximum (minimum) of the sequence of independent identically distributed random variables. The second, more modern, group of models are the peaks-over-threshold (POT) models. These are models for all large observations which exceed a threshold. This paper is concentrated on the first type of model. Here, the maximum loss of a bank caused by different technological accidents and natural disasters is treated.

Keywords: extreme value theory; Gumbel distribution; method of point estimation.

JEL classification: C24