др Ленка Главаш и др Павле Младеновић

СРЕДЊИ ПОВРАТНИ ПЕРИОД ЕКСТРЕМНОГ ДОГАЂАЈА

Теорија екстремних вредности обезбеђује теоријску основу за проучавање и моделирање догађаја, који се због своје екстремалне природе, ретко или веома ретко дешавају. Интересовање за овакве догађаје све је веће, јер је очигледно да они могу имати велике последице по човечанство.

У оквиру теорије екстремних вредности развијени су различити модели и технике које се могу користити за решавање важних проблема управљања ризиком у областима као што су: хидрологија (метеоролошки и поплавни надзор и предвиђање), климатологија (климатске промене), екологија, геофизика, ветар-инжењеринг, анализа поузданости механичких система и компоненти. С друге стране, значајне су и њене примене на проблеме у економији и финансијама (висока волатилност финансијских инструмената), као и у неживотном осигурању и реосигурању (појава великих одштета, узрокованих природним катастрофама или дејством људског фактора).

Јасна је потреба за развојем вероватносних и статистичких метода примењивих у анализи екстрема, па стога не чуди чињеница да је у последњих неколико деценија теорија екстремних вредности постала врло актуелна и широко проучавана математичка дисциплина са бројним применама.

1. Расподеле екстремних вредности и генералисане Паретове расподеле

Класична теорија екстремних вредности случајних низова бави се проучавањем граничног понашања узорачких екстрема, тј. случајних величина

$$M_n := \max_{1\leq j\leq n} X_j \quad \text{и} \quad m_n := \min_{1\leq j\leq n} X_j, \quad \text{при } n\to\infty,$$

где је $(X_n)_{n\in\mathbb{N}}$ низ независних, једнако расподељених случајних величина, са задатом, недегенерисаном функцијом расподеле F. Акценат ће, у наставку, бити на резултатима за максимуме због примена у актуарској математици, а аналогни резултати се могу формулисати и за минимуме на основу једнакости

$$\min_{1\leq j\leq n} X_j = -\max_{1\leq j\leq n}(-X_j).$$

Функција расподеле случајне величине M_n дата је, за $x\in\mathbb{R}$, са

$$P\{M_n\leq x\} = P\{X_j\leq x,\ \text{за } j=1,2,\dots,n\} = \prod_{j=1}^{n} P\{X_j\leq x\} = (F(x))^n. \tag{1.1}$$

Питање које се намеће тиче се асимптотског понашања максимума M_n, при $n \to \infty$.

Из једнакости (1.1) лако се види да случајна величина M_n конвергира у вероватноћи ка тачки x_F – десном крају носача функције расподеле F, тј. $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}$, при чему је $x_F \in \mathbb{R} \cup \{+\infty\}$, што није нарочито корисна информација у погледу реда величине максимума. Далеко је интересантније размотрити конвергенцију у расподели линеарно нормираног максимума ка некој недегенерисаној функцији расподеле. Следећа теорема, позната у литератури као теорема о екстремалним типовима, решава описани екстремални гранични проблем.

Теорема 1. [Гнеденко (1943); de Haan (1976)] *Ако постоје низови константи (a_n) и (b_n), $a_n > 0$, $b_n \in \mathbb{R}$ за $n \in \mathbb{N}$, тако да важи*

$$P\Big\{\frac{M_n - b_n}{a_n} \le x\Big\} = (F(a_n x + b_n))^n \to G(x), \quad \text{при } n \to \infty,$$

за $\forall x \in C(G)$, где је $C(G)$ скуп тачака непрекидности неке недегенерисане функције расподеле G, тада је функција расподеле G истог типа као нека од функција расподеле из следеће три фамилије:

$$G_0(x) = \exp(-e^{-x}), \quad -\infty < x < +\infty \tag{1.2}$$

$$G_{1,\alpha}(x) = \begin{cases} 0, & \text{за } x \le 0 \\ \exp(-x^{-\alpha}), & \text{за } x \ge 0, \end{cases} \quad (\alpha > 0) \tag{1.3}$$

$$G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{\alpha}), & \text{за } x \le 0 \\ 1, & \text{за } x \ge 0, \end{cases} \quad (\alpha > 0). \tag{1.4}$$

Значај овог резултата је у томе што он карактерише класу граничних расподела линеарно нормираног максимума у низу независних, једнако расподељених случајних величина.

Каже се да су две функције расподеле H_1 и H_2 истог типа ако постоје константе $a > 0$ и $b \in \mathbb{R}$ такве да једнакост $H_2(x) = H_1(ax + b)$ важи за $\forall x \in \mathbb{R}$.

Заједнички назив за горенаведене три параметарске (у α-параметризацији) фамилије функција расподела је расподеле екстремних вредности, а стандардни представници ових фамилија дати формулама (1.2)–(1.4) познати су, редом, и као Гумбелова, Фрешеова и Вејбулова расподела. Оне су математички повезане једноставним трансформацијама. У применама, тј. са становишта моделирања, ове три фамилије расподела, међутим, прилично различито манифестују екстремално понашање, што одговара различитом понашању репа функције расподеле F случајне величине X_j. Фрешеова расподела, као правилно променљива расподела, једна је од представника расподела са изразито тешким репом, које се генерално користе као добри модели за велике одштете у актуарској математици (ситуација када једна појединачна одштета доминира у збиру свих исплаћених одштета).

Расподеле екстремних вредности могу се, γ-параметризацијом, објединити у једну фамилију расподела, тзв. стандардних генералисаних расподела екстремних вредности чији су чланови облика

$$G_\gamma(x) = \begin{cases} \exp(-e^{-x}), & \text{ако је } \gamma = 0 \\ \exp(-(1+\gamma x)^{-1/\gamma}), & \text{ако је } \gamma \neq 0, \end{cases} \qquad 1+\gamma x > 0. \tag{1.5}$$

Други метод издвајања максималних вредности у случајном низу заснива се на прекорачењима задатог високог нивоа (прага). Појму расподеле максимума низа независних, једнако расподељених случајних величина у овом приступу одговара функција расподеле F_u прекорачења $X-u$ при услову $X>u$, дефинисана са

$$F_u(x) = P\{X-u \le x \,|\, X>u\} = \frac{F(x+u)-F(u)}{1-F(u)}, \quad x \ge 0,$$

где је u дати ниво (фиксирана вредност унутар носача функције расподеле F случајне величине X). Функција e_F дефинисана са

$$e_F(u) = E(X-u \,|\, X>u),$$

је функција средњег прекорачења.

Значајна је примена овог концепта у реосигурању, јер постоје типови уговора о реосигурању код којих реосигуравајућа компанија преузима обавезу да исплати само евентуална прекорачења максималне суме коју примарна осигуравајућа компанија може да исплати по појединачној одштети, и располаже подацима једино о таквим одштетама. Посебно су важне и његове импликације у актуарској математици, јер пружа могућност за разликовање расподела са лаким од расподела са тешким репом случајних величина са носачима неограниченим здесна. Наиме, ако функција средњег прекорачења $e_F(u)$ тежи бесконачности при $u \to \infty$ природно је сматрати да функција расподеле F има тежак реп, а ако $e_F(u)$ тежи (коначној) константи при $u \to \infty$ природно је сматрати да функција расподеле F има лак реп.

Дакле, питање које се намеће тиче се асимптотске расподеле прекорачења високог нивоа када тај ниво тежи десном крају носача расподеле. Теоријски резултати који су добијени у вези с тим омогућавају моделирање прекорачења високог нивоа.

Теорема 2. [Balkema, de Haan (1974); Pickands (1975)] *Ако за* $\forall x \in \mathbb{R}$ *важи*

$$F_u(a_u x + b_u) \to W(x), \quad \text{при } u \to x_F, \tag{1.6}$$

где је W *непрекидна функција расподеле, а* $a_u > 0$ *и* $b_u \in \mathbb{R}$, *онда је функција расподеле* W *истог типа као нека од функција расподеле из следеће фамилије:*

$$W_\gamma(x) = \begin{cases} 1-e^{-x}, & \text{ако је } \gamma = 0 \\ 1-(1+\gamma x)^{-1/\gamma}, & \text{ако је } \gamma \neq 0, \end{cases} \tag{1.7}$$

за $x \ge 0$ *ако је* $\gamma \ge 0$, *односно за* $0 \le x \le -1/\gamma$ *ако је* $\gamma < 0$.

Значај овог резултата је у томе што он карактерише класу непрекидних граничних расподела прекорачења високог нивоа за дату функцију расподеле F.

Назив горенаведене параметарске фамилије (у γ-параметризацији) функција расподела је стандардне генералисане Паретове расподеле. Стандардни представници ове фамилије, али у α-параметризацији, су:

- експоненцијална расподела

$$W_0(x) = \begin{cases} 0, & \text{за } x \le 0 \\ 1 - e^{-x}, & \text{за } x \ge 0 \end{cases}$$

- Паретова расподела

$$W_{1,\alpha}(x) = \begin{cases} 0, & \text{за } x \le 1 \\ 1 - x^{-\alpha}, & \text{за } x \ge 1 \end{cases}$$

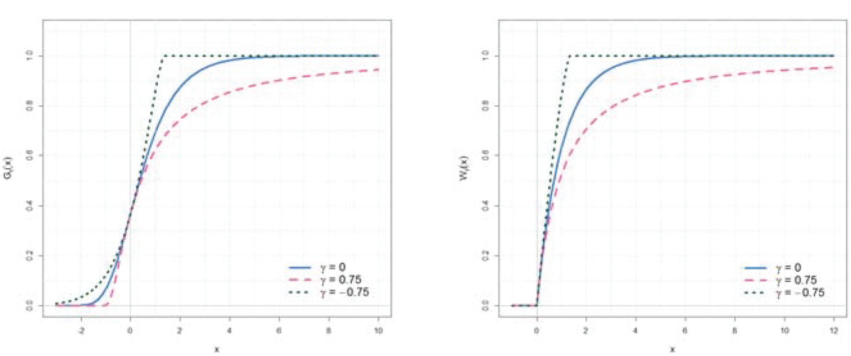
- бета расподела

$$W_{2,\alpha}(x) = \begin{cases} 0, & \text{за } x \le -1 \\ 1 - (-x)^{\alpha}, & \text{за } -1 \le x \le 0 \\ 1, & \text{за } x \ge 0. \end{cases}$$

Ако се дозволи да недегенерисана функција расподеле W, у граничној вредности датој формулом (1.6) за $\forall x \in C(W)$, није непрекидна, онда се као граничне могу појавити и неке недегенерисане дискретне расподеле.

Једнакости (1.5) и (1.7) указују на једноставну аналитичку везу која постоји између генералисаних расподела екстремних вредности и генералисаних Паретових расподела, а дата је са

$$W_\gamma(x) = 1 + \ln G_\gamma(x), \quad \text{ако је } \ln G_\gamma(x) > -1.$$



Сл. 1. Графици функција расподеле стандардних

(а) генералисаних расподела екстремних вредности (б) генералисаних Паретових расподела

2. Повратни период екстремног догађаја

Један од важних проблема, посебно са становишта примена у хидрологији и метеорологији као и управљања ризиком, чије решавање захтева познавање теорије екстремних вредности је оцењивање тзв. t-годишњег нивоа. Под t-годишњим нивоом $u(t)$ подразумева се праг $u = u(t)$ такав да је очекивани број прекорачења тог прага у временском интервалу дужине t једнак 1. При томе, подразумева се да су подаци бележени у еквидистантим временским тренуцима, односно да је за сваку годину (или ма коју другу јединицу мерења времена – дан, месец, годишње доба) доступна по једна опсервација, тј. реализована вредност случајне величине од интереса.

Нека је $(X_n)_{n\in\mathbb{N}}$ низ независних, једнако расподељених случајних величина са задатом функцијом расподеле F. Тада је ниво $u(t)$, заправо, решење једначине

$$E\Bigg(\sum_{j\le t} I_{\{X_j\le u\}}\Bigg) = 1. \tag{2.1}$$

Очигледно је да важи

$$u(t) = F^{-1}\left(1-\frac{1}{t}\right), \tag{2.2}$$

тј. $u(t)$ је, уствари, $(1-1/t)$-квантил функције расподеле F ($F^{-1}(\cdot)$ је уопштени инверз функције F). Важи и

$$P\{X_1 > u(t)\} = 1 - F(u(t)) = \frac{1}{t},$$

па, према томе, случајна величина од интереса одређене године прекорачује t-годишњи ниво са вероватноћом једнаком $1/t$.

У наставку ће бити објашњено зашто се t-годишњи ниво $u(t)$, дат са (2.2), назива и t-годишњи повратни ниво, што је још једна могућа интерпретација за $u(t)$.

Поред броја и величина прекорачења унапред одређеног нивоа значајно је уочити и тренутке у којима се дешавају та прекорачења. У вези с тим интересантно је, пре свега, размотрити питање појаве следећег прекорачења извесног вишег нивоа u у временски неограниченој будућности. Формално гледано, посматра се раније уведени случајан низ (X_n) са функцијом расподеле F; од интереса су догађаји облика $\{X_j > u\}$; тренутак првог прекорачења нивоа u, $u < x_F$, је случајна величина τ_1 дата са

$$\tau_1 := \min\{j \in \mathbb{N} : X_j > u\}.$$

На сличан начин, за фиксирани ниво u, може се дефинисати читав низ $(\tau_n)_{n\in\mathbb{N}}$ -- случајних тренутака прекорачења нивоа u, уређених монотоно растуће: $\tau_1 < \tau_2 < \dots$. Случајна величина τ_m тада се задаје рекурзивно као

$$\tau_m := \min\{j \in \mathbb{N} : j > \tau_{m-1},\, X_j > u\}, \quad m > 1.$$

Јасно је да постоји зависност случајних величина у низу (τ_n) од нивоа u, што се, по потреби, може нагласити и записом $\tau_m = \tau_m(u)$, $m \in \mathbb{N}$.

Разлике између узастопних чланова низа (τ_n), тј. прираштаји $\tau_2 - \tau_1$, $\tau_3 - \tau_2$, $\dots$, једноставно, представљају дужине временских интервала између узастопних прекорачења нивоа u, односно, у ширем контексту, свака од њих је дужина чекања између два екстремална догађаја одређеног типа. Оне су независне, једнако расподељене случајне величине са геометријском расподелом. Наиме,

$$\begin{aligned} P\{\tau_1 = k\} &= P\{X_1 \le u, X_2 \le u, \dots, X_{k-1} \le u, X_k > u\} \\ &= p(1-p)^{k-1}, \quad k \in \mathbb{N}, \end{aligned}$$

где је $p := 1 - F(u)$, при чему је коришћена независност случајних величина у низу (X_n). Према томе, τ_1 има $\mathcal{G}(p)$ расподелу, па је њено математичко очекивање једнако

$$E\tau_1 = \frac{1}{p} = \frac{1}{1 - F(u)}. \tag{2.3}$$

Даље,

$$\begin{aligned} &P\{\tau_1 = k_1, \tau_2 - \tau_1 = k_2, \dots, \tau_m - \tau_{m-1} = k_m\} \\ &= P\left\{\tau_1 = k_1, \tau_2 = k_1 + k_2, \dots, \tau_m = \sum_{j=1}^{m} k_j\right\} \\ &= P\{X_1 \le u, \dots, X_{k_1 - 1} \le u, X_{k_1} > u, X_{k_1+1} \le u, \dots, X_{k_1+k_2+\dots+k_m} > u\} \\ &= \prod_{j=1}^{m} P\{\tau_1 = k_j\}, \quad k_1, k_2, \dots, k_m \in= \mathbb{N}. \end{aligned}$$

Стога следи да тренутак m-тог прекорачења нивоа u има негативну биномну расподелу с параметрима m и p.

Математичко очекивање које се појављује у формули (2.3) назива се средњи повратни период (екстремалног) догађаја $\{X_j > u\}$.

Вероватноћа да ће временски период између два узастопна прекорачења нивоа u бити краћи од средњег повратног периода једнака је

$$P\{\tau_1 \le E\tau_1\} = P\left\{\tau_1 \le \left[\frac{1}{p}\right]\right\} = 1 - (1-p)^{[1/p]},$$

где је $[\cdot]$ цео део позитивног броја. За врло високе нивое u, тј. при $u \uparrow \infty$, а то значи при $p \downarrow 0$, добија се

$$\lim_{u \to \infty} P\{\tau_1 \le E\tau_1\} = \lim_{p \to 0} \left(1 - (1-p)^{[1/p]}\right) = 1 - e^{-1} \approx 0.63212,$$

што показује да је за врло високе нивое u математичко очекивање случајне величине τ_1, тј. средњи повратни период, веће од његове медијане, па расподела постаје асиметрична удесно.

Ниво u такав да је средњи повратни период једнак t назива се t-годишњи повратни ниво и он је решење једначине

$$E\tau_1 = \frac{1}{1-F(u)} = t.$$

Требало би приметити да ова једначина има исто решење као и једначина (2.1) и то решење $u(t)$ је управо оно дато формулом (2.2). Дисперзија случајне величине $\tau_1 = \tau_1(u(t))$ је тада једнака $(t-1)t$.

Једнакости (2.3) и (2.2) указују на значај оцењивања високих квантила расподеле. Важну улогу у решавању тог проблема имају статистике поретка. Посебне тешкоће појављују се при оцењивању високих квантила, јер је често потребно оцењивати квантиле који излазе изван опсега вредности у узорку на основу кога се врши оцењивање.

Подаци који се срећу у применама најчешће су такви да постоји зависност и/или нестационарност опсервација у узорку. У тим ситуацијама претходно изложена прича се мора модификовати са циљем проналажења адекватних процедура за оцењивање високих квантила у таквим узорцима.

3. Примери и примене

Интересантно је приметити да је код произвољне расподеле средњи повратни период који одговара нивоу једнаком медијани расподеле 2 године (или неке друге јединице мерења времена). Слично, средњи повратни период који одговара нивоу једнаком трећем квартилу расподеле је 4 године итд.

Нека је (X_n) низ независних, једнако расподељених случајних величина са функцијом расподеле F и нека је $t = t(u)$, у складу са ранијим ознакама, средњи повратни период догађаја $\{X_j > u\}$. У наставку ће бити наведене формуле за t за неке познате расподеле.

Пример 1. Експоненцијална расподела $\mathcal{E}(1)$, $F(x) = 1 - e^{-x}$, за $x \geq 0$

$$t(u) = e^u, \quad u \geq 0$$

Средњи повратни период $t(Cu)$ који одговара нивоу Cu, где је $C > 1$ константа, повезан је са $t(u)$ једноставном једнакошћу

$$t(Cu) = (t(u))^C.$$

Пример 2. Логистичка расподела $\mathcal{LG}(1)$, $F(x) = (1 + e^{-x})^{-1}$, за $x \in \mathbb{R}$

$$t(u) = 1 + e^u, \quad u \in \mathbb{R}$$

Овде је $t(Cu) = 1 + (t(u) - 1)^C$.

Пример 3. Паретова расподела $\mathcal{PAR}(a)$, $a > 0$, $F(x) = 1 - x^{-a}$, за $x \geq 1$

$$t(u) = u^a, \quad u \geq 1$$

Овде је $t(Cu) = C^a \cdot t(u)$.

Пример 4. Нека су чланови низа (X_n) груписани хронолошки у блокове дужине k, где је k велико, и нека је, затим, у сваком блоку уочена највећа вредност, чиме је добијен низ $M_{k,1}, M_{k,2}, \ldots, M_{k,l}$ блок максимума. Претпостави се да се последњи низ може моделирати неком генералисаном расподелом екстремних вредности, задатом формулом

$$G_{\gamma,\mu,\sigma}(x) = \exp\left\{ -\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\gamma}\right\}, \qquad 1+\gamma(x-\mu)/\sigma > 0, \tag{3.1}$$

где је $\gamma \in \mathbb{R}$ параметар облика, а за параметре положаја и размере важи, редом, $\mu \in \mathbb{R}$, $\sigma > 0$.

Требало би напоменути да се често бирају блокови који одговарају временском периоду од једне године, па су, у том случају, блок максимуми – једногодишњи максимуми.

Инвертовањем формуле (3.1) добија се

$$u_p = \begin{cases} \mu - \sigma \ln(-\ln(1-p)), & \text{ако је } \gamma = 0 \\ \mu - \dfrac{\sigma}{\gamma}\left(1 - (-\ln(1-p))^{-\gamma}\right), & \text{ако је } \gamma \neq 0, \end{cases} \quad \text{где је } p := 1 - G_{\gamma,\mu,\sigma}(u_p). \tag{3.2}$$

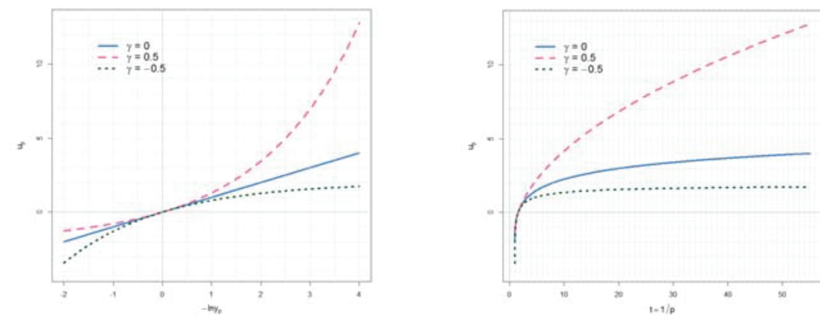
Уколико се у формуле (3.2) уврсте оцене максималне веродостојности $\widehat{\boldsymbol{\theta}^{\mathbf{0}}} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ параметара генералисане расподеле екстремних вредности добијају се оцене максималне веродостојности u_p квантила расподеле годишњег максимума. Очигледно је u_p повратни ниво који одговара средњем повратном периоду једнаком $1/p$. Према томе, једногодишњи максимум за било коју годину прекорачује ниво u_p са вероватноћом p. Такође, може се рећи да је u_p t-годишњи ниво, где је $t = 1/p$.

Ако се уведе ознака $y_p := -\ln(1-p)$ онда формула (3.2) добија облик

$$u_p = \begin{cases} \mu - \sigma \ln y_p, & \text{ако је } \gamma = 0 \\ \mu - \dfrac{\sigma}{\gamma}\left(1 - y_p^{-\gamma}\right), & \text{ако је } \gamma \neq 0, \end{cases}$$

а график квантила u_p као функције од $-\ln y_p$ је тзв. график повратног нивоа. Лако се види да је овај график – график линеарне функције за $\gamma = 0$. За вредности $\gamma < 0$ функција приказана на графику је конкавна (тј. област испод графика функције је конвексна), при чему је права $y = \mu - \sigma/\gamma$ хоризонтална асимптота при $-\ln y_p \to +\infty$ (тј. $p \to 0$). Дакле, код Вејбулових расподела екстремних вредности ($\gamma < 0$) постоји повратни ниво, и то је управо ниво $\mu - \sigma/\gamma$, чији је повратни период бесконачан. За вредности $\gamma > 0$ функција приказана на графику је конвексна (тј. област изнад графика функције је конвексна) и нема хоризонталну асимптоту при $-\ln y_p \to +\infty$, тј. неограничена је одозго. Испоставља се да је график повратног нивоа посебно погодан и за презентацију и за валидацију разматраног модела за конкретне опсервације и њихове блок максимуме.

Посебно су од интереса мале вредности p, које одговарају врло високим квантилима, односно веома ретким – екстремним догађајима. Логаритамско скалирање вредности на апсциси омогућава да се уоче необична понашања расподела



Сл. 2. Графици повратног нивоа ($\mu = 0$, $\sigma = 1$, $\gamma \in \{-0.5, 0, 0.5\}$)
(а) на апсциси су вредности $-\ln y_p$ (б) на апсциси су вредности $t = 1/p$

које се користе нпр. за моделирање блок максимума, а која указују на бесконачне повратне периоде. Ту до изражаја такође долазе и аутлајери међу опсервацијама.

ЛИТЕРАТУРА

[1] S.G. Coles, *An Introduction to Statistical Modelling of Extreme Values*, Springer, 2001.

[2] P. Emberschts, C. Klüppelberg and T. Mikosch, *Modelling Extremal Events: for Insurance and Finance*, Springer, Berlin–Heidelberg, 2013.

[3] M. Falk, J. Hüsler and R.-D. Riess, *Laws of Small Numbers: Extremes and Rare Events*, Springer, Basel, 2011.

[4] П. Младеновић, *Екстремне вредности случајних низова*, Математички факултет, Београд, 2002.

[5] П. Младеновић, *Елементи актуарске математике*, Математички факултет, Београд, 2014.

[6] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values – with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser, Basel–Boston–Berlin, 2007.

Математички факултет, Београд
E-mail: `lenka@matf.bg.ac.rs, paja@matf.bg.ac.rs`