UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET

Generalisane Paretove raspodele u teoriji ekstremnih vrednosti i primene

MAGISTARSKI RAD

kandidat: Jelena Jocković

mentor: prof. dr Pavle Mladenović

Beograd, 2009.

Sadržaj

Uvod

Često se dešava da imamo obiman skup numeričkih podataka, za koje pretpostavljamo da su dobijeni na isti, ili bar sličan način, pod istim uslovima. Dakle, možemo da ih posmatramo kao realizacije nekih slučajnih veličina sa istom raspodelom. To mogu biti, na primer:

− rezultati nekih merenja, ponovljenih mnogo puta,
− vrednosti akcija neke kompanije u poslednjih 10 godina,
− broj pogodaka u sportskim utakmicama,
− najniže dnevne temperature u poslednjih nekoliko godina...

Još češće se dešava da nam svi ti podaci nisu podjednako interesantni ili važni, već nas zanimaju samo ekstremni od njih, pa želimo nekako da predvidimo njihovu vrednost i pronadjemo neku pravilnost u njihovom pojavljivanju.

Postoje dva pristupa ovom problemu:

1. da se posmatra jedna ili nekoliko maksimalnih (minimalnih) vrednosti i modelira odgovarajućom raspodelom (pristup teorije ekstremnih vrednosti)

2. da se vrednosti veće od nekog unapred zadatog visokog praga modeliraju odgovarajućom raspodelom (POT-pristup).

Ovaj magistarski rad predstavlja pokušaj da se prikažu teorijske osnove ova dva pristupa, posebno ovog drugog, i da se istaknu neke veze izmedju njih. Podeljen je u tri poglavlja.

U prvom poglavlju ponovljeni su poznati rezultati teorije ekstremnih vrednosti. Definisane su raspodele ekstremnih vrednosti: Gumbelova, Frešeova i Vejbulova, navedene njihove osobine i osnovni rezultat klasične teorije ekstremnih vrednosti - teorema o ekstremalnim tipovima. Date su karakterizacije oblasti privlačenja za maksimume svake od raspodela ekstremnih vrednosti. Za ovo poglavlje korišćena je literatura [4],[8] i [12].

U drugom poglavlju uveden je POT-pristup. Objašnjeni su razni načini izdvajanja ekstremuma iz datog skupa podataka. Definisane su generalisane Paretove raspodele (GPD), navedene njihove osobine i veze sa raspodelama ekstremnih vrednosti. Definisana je POT-oblast privlačenja i dat je osnovni rezultat koji se na nju odnosi - teorema o graničnoj raspodeli prekoračenja visokog praga. Za svaku od GPD raspodela data je karakterizacija njene POT-oblasti privlačenja, kao i neki primeri. Literatura korišćena za ovo poglavlje je [1],[4],[8] i [12].

Treće poglavlje se odnosi na dva različita pravca primene GPD raspodela. U delu 3.1 je dat primer korišćenja rezultata navedenih u prva dva poglavlja kod nekih kombinatornih problema čekanja. Postojeći rezultati su dopunjeni novim (teoreme 3.1.2, 3.1.3, 3.1.5 i 3.1.7).
U delu 3.2 objašnjeno je modeliranje vremenskih serija GPD raspodelama. Pokazani su neki načini ocenjivanja nepoznatih parametara, kao i teškoće koje se pri tome javljaju. To je ilustrovano na primeru jedne finansijske vremenske serije. Ovde je korišćena literatura [6],[7],[8],[9] i [12].

Zahvaljujem se svom mentoru, profesoru Pavlu Mladenoviću, na velikoj pomoći tokom izrade magistarskog rada, na strpljenju i brojnim korisnim savetima, kao i na tome što me je zainteresovao za rad u ovoj oblasti.

U Beogradu, 2009. Jelena Jocković

1. poglavlje

Teorija ekstremnih vrednosti

1.1 Uvod i osnovni pojmovi

Teorija ekstremnih vrednosti proučava asimptotska svojstva slučajnih veličina oblika:

$$M_n = \max\{X_1, X_2, ..., X_n\}$$

$$m_n = \min\{X_1, X_2, ..., X_n\}$$

kad $n \to \infty$, pri čemu su $X_1, X_2, ..., X_n$ slučajne veličine sa datim raspodelama verovatnoća.
Ako za slučajnu veličinu M_n važi:

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \to G(x), \qquad n \to \infty \tag{1}$$

gde je $G(x)$ nedegenerisana funkcija raspodele, a $a_n > 0$ i b_n $(n \in N)$ realni brojevi, onda se kaže da $G(x)$ odredjuje *graničnu raspodelu linearno normiranog maksimuma* M_n, a a_n i b_n su *normirajuće konstante.*

Granična raspodela linearno normiranog minimuma se odredjuje na sličan način, korišćenjem nejednakosti:

$$\min\{X_1, X_2, ..., X_n\} = -\max\{-X_1, -X_2, ..., -X_n\}$$

Klasična teorija ekstremnih vrednosti

Klasična teorija ekstremnih vrednosti se bavi odredjivanjem granične raspodele linearno normiranog maksimuma (minimuma) u slučaju kad su članovi niza $X_1, X_2, .., X_n, ..$ nezavisne slučajne veličine sa zajedničkom funkcijom

raspodele F.
U tom slučaju jednakost (1) se svodi na:

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = P\{X_1 \leq a_n x + b_n, ..., X_n \leq a_n x + b_n\}$$
$$= (P\{X_1 \leq a_n x + b_n\})^n = (F(a_n x + b_n))^n$$

i kaže se da funkcija F pripada *oblasti privlačenja za maksimume funkcije raspodele* G ($F \in D(G)$).

Definicije M-stabilnosti i oblasti privlačenja

Definicija 1.1.1 Nedegenerisana funkcija raspodele G je *maksimum stabilna* (M-stabilna) ako za svako $n \geq 2$, $n \in N$ postoje nizovi realnih brojeva $a_n > 0$ i b_n, takvi da za svako $x \in R$ važi:

$$G^n(a_n x + b_n) = G(x).$$

Definicija 1.1.2 Funkcije raspodela G_1 i G_2 su *istog tipa* ako postoje realni brojevi $a > 0$ i b, takvi da za svako $x \in R$ važi:

$$G_2(x) = G_1(ax + b).$$

Definicija 1.1.3 Funkcija raspodele F pripada *oblasti privlačenja za maksimume* nedegenerisane funkcije raspodele G ako postoje nizovi realnih brojeva $a_n > 0$ i b_n, $n \in N$, takvi da važi:

$$F^n(a_n x + b_n) \to G(x), \qquad n \to \infty$$

za svaku tačku neprekidnosti x funkcije G ($x \in D(G)$).

Definicije pravilno promenljivih, Π-promenljivih i Γ -promenljivih funkcija

Definicija 1.1.4 (pravilna promenljivost u beskonačnosti) Merljiva funkcija $F : (0, +\infty) \to (0, +\infty)$ je *pravilno promenljiva u beskonačnosti, sa indeksom* ρ ($F \in \Pi\Pi_\rho$) ako postoji $\rho \in R$, takvo da za svaki $x \in R$ važi:

$$\lim_{t \to +\infty} \frac{F(tx)}{F(t)} = x^\rho.$$

Definicija 1.1.5 (pravilna promenljivost u konačnoj tački) Merljiva funkcija $F : (0, +\infty) \to (0, +\infty)$ je *pravilno promenljiva u tački* $x_0 > 0$ *(pri* $x \uparrow x_0$*)* ako je funkcija $F(x_0 - \frac{1}{x})$ pravilno promenljiva u beskonačnosti.

Definicija 1.1.6 (Π - promenljivost) Rastuća funkcija $F : (c, +\infty) \to R$ je Π - *promenljiva (*$F \in \Pi$*)* ako postoje pomoćne funkcije

$$a : (c, +\infty) \to (0, +\infty), \quad b : (c, +\infty) \to R,$$

takve da za svako $x > 0$ važi:

$$\lim_{t \to +\infty} \frac{F(tx) - b(t)}{a(t)} = \ln x.$$

Definicija 1.1.7 (Γ - promenljivost) Rastuća funkcija $F : (c, x_0) \to R$ je Γ - *promenljiva (*$F \in \Gamma$*)* ako važe uslovi:

1. $\lim\limits_{x \uparrow x_0} F(x) = +\infty$
2. postoji funkcija $g : (c, x_0) \to (0, +\infty)$, takva da za svako $x \in R$ važi:

$$\lim_{t \uparrow x_0} \frac{F(t + xg(t))}{F(t)} = e^x.$$

1.2 Gumbelova, Frešeova i Vejbulova raspodela

Ako je granična raspodela linearno normiranog maksimuma niza nezavisnih i jednako raspodeljenih slučajnih veličina nedegenerisana, tada je ona istog tipa kao jedna od sledeće tri funkcije raspodele:

- *Gumbelova raspodela*
  $G_0(x) = e^{-e^{-x}}, \qquad -\infty < x < +\infty$

- *Frešeova raspodela s parametrom* $\alpha > 0$:
  $G_{1,\alpha}(x) = 0, \qquad x < 0$
  $G_{1,\alpha}(x) = e^{-x^{-\alpha}}, \qquad x \geq 0$

- *Vejbulova raspodela s parametrom* $\alpha > 0$:
  $G_{2,\alpha}(x) = e^{-(-x)^{\alpha}}, \qquad x \leq 0$
  $G_{2,\alpha}(x) = 1, \qquad x > 0$

Ove tri funkcije raspodele se nazivaju *raspodele ekstremnih vrednosti.* Parametar α kod Frešeove i Vejbulove raspodele se naziva *parametar oblika*, a za funkcije raspodele se kaže da su date u α - *parametrizaciji.*

Parametri položaja i razmere

Ako slučajna veličina X ima Gumbelovu, Frešeovu ili Vejbulovu raspodelu (redom), onda je funkcija raspodele slučajne veličine $\sigma X + \mu$, $\sigma > 0$ odredjena sledećim funkcijama raspodele:

- $G_{0,0,\mu,\sigma}(x) = G_0(\frac{x-\mu}{\sigma}) = e^{-e^{-\frac{x-\mu}{\sigma}}}$

- $G_{1,\alpha,\mu,\sigma}(x) = G_1(\frac{x-\mu}{\sigma}) = e^{-(\frac{x-\mu}{\sigma})^{-\alpha}}$

- $G_{2,\alpha,\mu,\sigma}(x) = G_2(\frac{x-\mu}{\sigma}) = e^{-(-\frac{x-\mu}{\sigma})^{\alpha}}$

Raspodele odredjene funkcijama raspodele $G_{i,\alpha,\mu,\sigma}(x)$, $i \in \{0, 1, 2\}$ se nazivaju *Gumbelova, Frešeova i Vejbulova raspodela sa parametrom oblika* α*, parametrom položaja* μ *i parametrom razmere* σ.

γ-parametrizacija

Raspodele ekstremnih vrednosti, iako su na prvi pogled različitog tipa, pokazuju neke zajedničke osobine:

$$\lim_{\alpha\to\infty} G_{1,\alpha}(x) = G_0(x),$$
$$\lim_{\alpha\to\infty} G_{2,\alpha}(x) = G_0(x).$$

Ova osobina se može zapaziti posmatranjem grafika funkcije raspodele i gustine raspodele, za dovoljno veliko α.
Ako se uvede smena $\gamma = \frac{1}{\alpha}$ kod Frešeove raspodele i $\gamma = -\frac{1}{\alpha}$ kod Vejbulove raspodele, dobija se jedinstven, neprekidan model, $G_\gamma(x)$, pri čemu važi:

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

tj. sve tri funkcije raspodele ekstremnih vrednosti se mogu prikazati kao familija raspodela koja zavisi od istog parametra oblika, γ.
U γ-parametrizaciji:

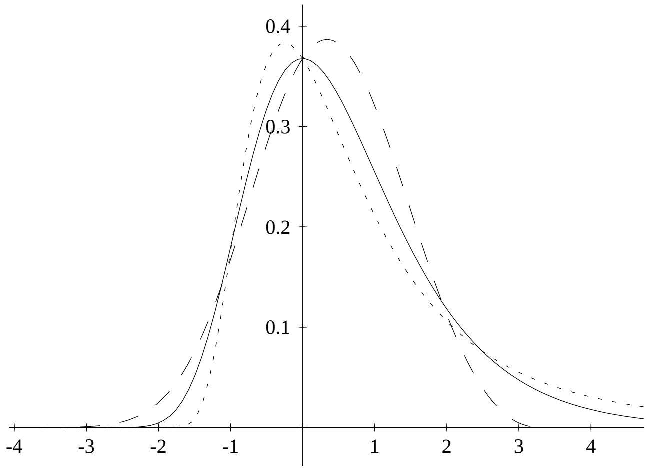
- Gumbelova raspodela:
  $G_0(x) = e^{-e^{-x}}$
- Frešeova i Vejbulova raspodela:
  $G_\gamma(x) = e^{-(1+\gamma x)^{-\frac{1}{\gamma}}}, \qquad 1+\gamma x > 0, \qquad \gamma \neq 0$

Veza izmedju α i γ - parametrizacije

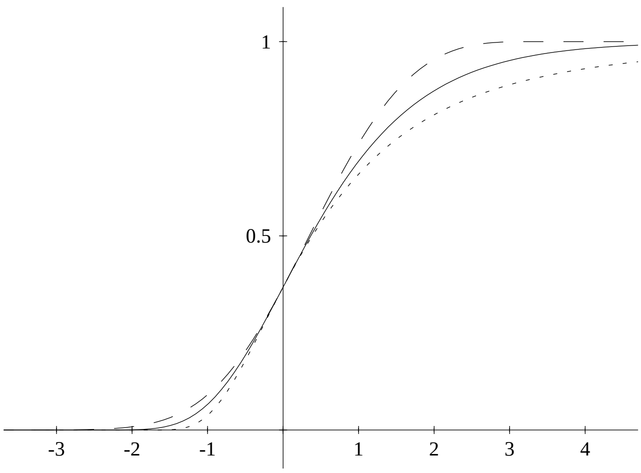
- $G_\gamma(x) = G_{1,\alpha,-\alpha,\alpha}(x), \qquad$ za $\gamma = \frac{1}{\alpha} > 0$
- $G_\gamma(x) = G_{2,\alpha,\alpha,\alpha}(x), \qquad$ za $\gamma = -\frac{1}{\alpha} < 0$

Gustine raspodela ekstremnih vrednosti u različitim parametrizacijama

raspodela	α-parametrizacija	γ-parametrizacija
Gumbelova raspodela	$e^{-x}\exp(-e^{-x})$ $-\infty < x < \infty$	$e^{-x}\exp(-e^{-x})$ $-\infty < x < \infty$
Frešeova raspodela	$\alpha x^{-(1+\alpha)} G_{1,\alpha}(x)$ $x \geq 0,\ \alpha > 0$	$(1+\gamma x)^{-(1+\frac{1}{\gamma})} G_{\gamma}(x)$ $\gamma > 0,\ 1+\gamma x > 0$
Vejbulova raspodela	$\alpha(-x)^{\alpha-1} G_{2,\alpha}(x)$ $x \leq 0,\ \alpha > 0$	$(1+\gamma x)^{-(1+\frac{1}{\gamma})} G_{\gamma}(x)$ $\gamma < 0,\ 1+\gamma x > 0$



slika 1: Gustine raspodela ekstremnih vrednosti u γ-parametrizaciji: $\gamma = 0$ (puna linija), $\gamma = 0.3$ (tačkasta linija), $\gamma = -0.3$ (isprekidana linija)



slika 2: Funkcije raspodela ekstremnih vrednosti u γ-parametrizaciji: $\gamma = 0$ (puna linija), $\gamma = 0.3$ (tačkasta linija), $\gamma = -0.3$ (isprekidana linija)

Teorema o ekstremalnim tipovima

Teorema 1.2.1 (Gnedenko (1943), de Haan (1976)) *Neka je* (X_n), $n \in N$, *niz nezavisnih slučajnih veličina sa zajedničkom funkcijom raspodele* F *i* $M_n = \max\{X_1, X_2, ..., X_n\}$. *Ako postoje nizovi realnih brojeva* $a_n > 0$ *i* b_n, $n \in N$, *takvi da:*

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

gde je G *nedegenerisana funkcija raspodele, tada je funkcija* F *istog tipa kao jedna od funkcija raspodele ekstremnih vrednosti.*

1.3 Karakterizacija oblasti privlačenja svake od raspodela ekstremnih vrednosti

Za svaku od raspodela ekstremnih vrednosti, odredjeni su potrebni i dovoljni uslovi da neka funkcija raspodele F pripada njenoj oblasti privlačenja. Poznato je da je sve funkcije iz oblasti privlačenja $G_{1,\alpha}$ desni kraj nosača raspodele jednak $+\infty$, za funkcije iz oblasti privlačenja $G_{2,\alpha}$ je desni kraj nosača raspodele konačan broj, a oblasti privlačenja G_0 pripadaju funkcije i jednog i drugog tipa.

Oblast privlačenja Frešeove raspodele

Sledeća teorema daje dovoljne uslove da apsolutno neprekidna funkcija raspodele pripada oblasti privlačenja Frešeove raspodele:

Teorema 1.3.1 (von Mises (1936)) *Neka je F apsolutno neprekidna funkcija raspodele sa gustinom raspodele f. Ako važe uslovi:*

1. $f(x) > 0$ *za* $x \geq x_0$
2. $\lim\limits_{t\to\infty} \frac{tf(t)}{1-F(t)} = \alpha > 0,$

tada $F \in D(G_{1,\alpha})$.

Sledeća teorema daje potrebne i dovoljne uslove da funkcija raspodele pripada oblasti privlačenja Frešeove raspodele:

Teorema 1.3.2 (Gnedenko (1943)) *Funkcija raspodele F pripada oblasti privlačenja Frešeove raspodele ako i samo ako važe uslovi:*

1. $x_0 = \sup\{t : F(t) < 1\} = +\infty$
2. $1 - F \in \Pi\Pi_{-\alpha}$.

Normirajuće konstante su:

$$a_n = \left(\frac{1}{1-F}\right)^{-1}(n), \qquad b_n = 0.$$

Oblast privlačenja Vejbulove raspodele

Sledeća teorema daje dovoljne uslove da apsolutno neprekidna funkcija raspodele pripada oblasti privlačenja Vejbulove raspodele:

Teorema 1.3.3 (von Mises (1936)) *Neka je F apsolutno neprekidna funkcija raspodele sa gustinom raspodele f. Ako važe uslovi:*

1. $x_0 = \sup\{t : F(t) < 1\} < +\infty$
2. $f(x) > 0$ *za* $x \in (a, x_0)$
3. $\lim\limits_{t \uparrow x_0} \frac{(x_0 - t)f(t)}{1 - F(t)} = \alpha > 0,$

tada $F \in D(G_{2,\alpha})$.

Sledeća teorema daje potrebne i dovoljne uslove da funkcija raspodele pripada oblasti privlačenja Vejbulove raspodele:

Teorema 1.3.4 (Gnedenko (1943)) *Funkcija raspodele F pripada oblasti privlačenja Vejbulove raspodele ako i samo ako važe uslovi:*

1. $x_0 = \sup\{t : F(t) < 1\} < +\infty$
2. $1 - F\left(x_0 - \frac{1}{x}\right) \in \Pi\Pi_{-\alpha}$, $x \to +\infty$.

Normirajuće konstante su:

$$a_n = x_0 - \gamma_n, \qquad b_n = x_0$$

gde je

$$\gamma_n = \left(\frac{1}{1-F}\right)^{-1}(n).$$

Oblast privlačenja Gumbelove raspodele

Sledeća teorema daje dovoljne uslove da apsolutno neprekidna funkcija raspodele pripada oblasti privlačenja Gumbelove raspodele:

Teorema 1.3.5 (von Mises (1936)) *Neka je F apsolutno neprekidna funkcija raspodele i $x_0 = \sup\{t : F(t) < 1\}$. Ako važe uslovi:*

1. $F''(x) < 0$ *za sve* $x \in (a, x_0)$, $x_0 \leq +\infty$
2. $F'(x) = 0$ *za* $x \geq x_0$
3. $\lim\limits_{t \to x_0} \frac{F''(t)(1 - F(t))}{(F'(t))^2} = -1,$

tada važi $F \in D(G_0)$.

Sledeća teorema daje potrebne i dovoljne uslove da funkcija raspodele pripada oblasti privlačenja Gumbelove raspodele:

Teorema 1.3.6 (Gnedenko (1943), Mejzler (1949), de Haan (1970)) *Neka je F funkcija raspodele, $x_0 = \sup\{t : F(t) < 1\}$ i $H(x) = \frac{1}{1-F(x)}$ za $x < x_0$. Tada su ekvivalentni sledeći uslovi:*

1. $F \in D(G_0)$
2. $H \in \Gamma$
3. $H^{-1} \in \Pi$.

Teorema 1.3.7 (de Haan (1970)) *$F \in D(G_0)$ ako i samo ako važi:*

$$\lim_{x \uparrow x_0} \frac{(1-F(x)) \int_x^{x_0} \int_y^{x_0} (1-F(t))dtdy}{(\int_x^{x_0}(1-F(t))dt)^2} = 1,$$

i svi zapisani integrali su konačni. Tada važi $\frac{1}{1-F} \in \Gamma$, a pomoćna funkcija g se može odrediti tako da važi:

$$g(t) = \frac{\int_x^{x_0} \int_y^{x_0} (1-F(t))dtdy}{\int_x^{x_0}(1-F(t))dt}$$

ili

$$g(t) = \frac{\int_x^{x_0}(1-F(t))dt}{(1-F(x))}.$$

Normirajuće konstante se mogu tako odabrati da važi:

$$a_n = F(b_n), \qquad b_n = \left(\frac{1}{1-F}\right)^{-1}(n).$$

Teorema 1.3.8 (de Haan (1970)) *$F \in D(G_0)$ ako i samo ako važi:*

$$\lim_{x \uparrow x_0} \frac{\int_x^{x_0}(1-F(t))^\alpha dt}{(1-F(x)) \int_x^{x_0}(1-F(t))^{\alpha-1}dt} = \frac{\alpha-1}{\alpha}$$

za neko $\alpha > 1$.

Dokazi teorema iz ovog poglavlja se mogu naći u knjizi [8].

Primeri raspodela iz svake od oblasti privlačenja raspodela ekstremnih vrednosti

raspodela	oblast privlačenja	normirajuće konstante
Gumbelova raspodela $F(x) = \exp(-\exp(-x))$	$D(G_0)$	$a_n = 1,\ b_n = \ln n$
Normalna raspodela	$D(G_0)$	$a_n = \frac{1}{\sqrt{2\ln n}}$, $b_n = \sqrt{2\ln n} - \frac{\ln\ln n + \ln 4\pi}{2\sqrt{2\ln n}}$
Eksponencijalna raspodela $F(x) = 1 - \exp(-\lambda x)$	$D(G_0)$	$a_n = \frac{1}{\lambda},\ b_n = \frac{1}{\lambda}\ln n$
Frešeova raspodela $F(x) = \exp(-x^{-\alpha})$	$D(G_{1,\alpha})$	$a_n = n^{\frac{1}{\alpha}},\ b_n = 0$
Paretova raspodela $F(x) = 1 - kx^{-\alpha}$	$D(G_{1,\alpha})$	$a_n = (kn)^{\frac{1}{\alpha}},\ b_n = 0$
Košijeva raspodela $F(x) = \frac{1}{2} + \frac{1}{\pi} arctgx$	$D(G_{1,1})$	$a_n = ctg\frac{\pi}{n},\ b_n = 0$
Vejbulova raspodela $F(x) = \exp(-(-x)^{\alpha})$	$D(G_{2,\alpha})$	$a_n = n^{-\frac{1}{\alpha}},\ b_n = 0$
Ravnomerna raspodela na $[0,c]$ $F(x) = \frac{x}{c}$	$D(G_{2,1})$	$a_n = \frac{c}{n},\ b_n = c$

2. poglavlje

Generalisane Paretove raspodele

2.1 Modeliranje prekoračenja datog praga

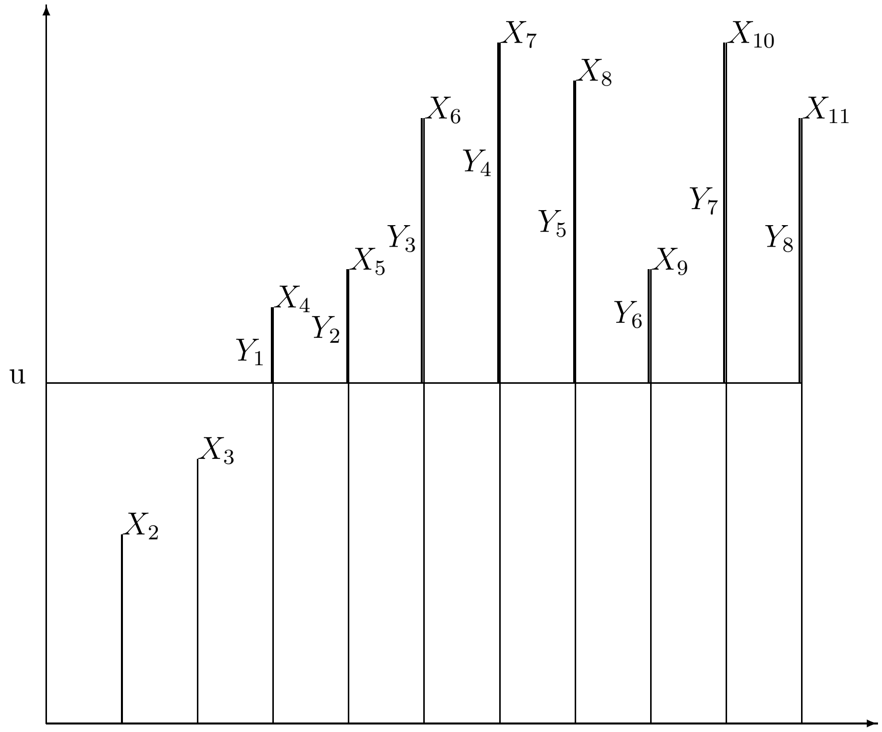
Načini dobijanja uzorka ekstremnih vrednosti

Neka je dat uzorak $(X_1, X_2, ..., X_N)$, čiji članovi predstavljaju vrednosti neke slučajne veličine X. Postoji više načina da se iz datog uzorka izdvoje ekstremne vrednosti. Neki od njih su:

- *Metod blok maksimuma.*
  Neka su k i n prirodni brojevi takvi da je $k < N$, $nk \le N < (n+1)k$ i $k \ll N$. Neka je:
  $$\begin{aligned} Y_1 &= \max\{X_1, X_2, ..., X_k\}, \\ Y_2 &= \max\{X_{k+1}, X_{k+2}, ..., X_{2k}\}, \\ &\ldots\ldots\ldots\ldots \\ Y_n &= \max\{X_{(n-1)k+1}, X_{(n-1)k+2}, ..., X_{nk}\}. \end{aligned}$$
  Tada je $(Y_1, Y_2, ..., Y_n)$ uzorak ekstremnih vrednosti obima n slučajne veličine X.

- *Metod prekoračenja datog praga (peaks-over-threshold (POT) metod).*
  Iz datog uzorka se izdvoje sve vrednosti koje su veće od nekog zadatog praga u. Ako takvih članova ima n, dobija se uzorak obima n ekstremnih vrednosti veličine X.

- *Korišćenje statistika poretka.*
  Ako su $X_{1,N}, ..., X_{N,N}$ vrednosti datog uzorka poredjane po veličini (od najmanje do najveće), onda je $(X_{N-n+1,N}, ..., X_{N,N})$ uzorak obima n ekstremnih vrednosti slučajne veličine X.



slika 3: POT-metod za izdvajanje ekstremnih vrednosti iz datog uzorka

Broj i veličina prekoračenja datog praga, srednje prekoračenje

Neka je dat slučajan uzorak $(X_1, ..., X_N)$, gde su $X_1, ..., X_N$ nezavisne slučajne veličine sa istom raspodelom, i neka je odredjen prag, u. Kad se koristi POT-metod za dobijanje uzorka ekstremnih vrednosti neke slučajne veličine, promenom praga, menjaju se i broj i veličina prekoračenja tog praga. Prekoračenja su vrednosti:

$$Y_1 = X_1 - u, \quad Y_2 = X_2 - u, ..., \quad Y_n = X_n - u,$$

gde su X_1,...,X_n elementi datog uzorka veći od u.
Za broj prekoračenja, N_u, važi:

$$N_u = \{i | i \in \{1, .., N\}, X_i > u\} = \sum_{i \leq N} I\{X_i > u\}$$

i to je jedna slučajna veličina. U slučaju da su X_i slučajne veličine sa zajedničkom funkcijom raspodele F, slučajna veličina N_u ima binomnu raspodelu sa parametrima n i p, gde je $p = 1 - F(u)$. Srednja vrednost broja prekoračenja datog praga u je onda $E(N_u) = n(1 - F(u))$.

Neka je X slučajna veličina sa funkcijom raspodele F. Prekoračenja se pojavljuju pod uslovom da je vrednost slučajne veličine X veća od praga u, pa je *funkcija raspodele prekoračenja praga* u, $F^{[u]}(x)$, jednaka:

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

za $x \geq u$,
ili, u malo drugačijem obliku:

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

za $x \geq 0$.
Matematičko očekivanje veličine prekoračenja datog praga u, za datu funkciju raspodele F, se naziva *srednje prekoračenje*:

$$e_F(u) = E(X - u | X > u).$$

Definicije POT-stabilnosti i POT-oblasti privlačenja

Definicija 2.1.1 Nedegenerisana funkcija raspodele F je *POT-stabilna* ako za zadati prag u, $u < \varpi(F)$, gde je $\varpi(F)$ desni kraj nosača raspodele, postoje realni brojevi $a_u > 0$ i b_u, takvi da za svako $x \in R$ važi:

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

Definicija 2.1.2 Funkcija raspodele F pripada *POT-oblasti privlačenja* nedegenerisane funkcije raspodele G ($F \in D_p(G)$) ako postoje funkcije $a(u) > 0$ i $b(u) \in R$, takve da važi:

$$F^{[u]}(a(u)x + b(u)) \to G(x), \qquad u \to \varpi(F).$$

2.2 Eksponencijalna, Paretova i beta raspodela

α-parametrizacija

Sve moguće neprekidne funkcije raspodele koje se mogu pojaviti kao granične raspodele prekoračenja praga u, za datu funkciju raspodele F, kad u teži desnom kraju nosača raspodele F, su istog tipa kao jedna od sledećih funkcija raspodele:

- *Eksponencijalna raspodela*:
  $W_0(x) = 1 - e^{-x}, \qquad x \geq 0$
  $W_0(x) = 0, \qquad x < 0$
- *Paretova raspodela s parametrom* $\alpha > 0$:
  $W_{1,\alpha}(x) = 1 - x^{-\alpha}, \qquad x \geq 1$
  $W_{1,\alpha}(x) = 0, \qquad x < 1$
- *Beta raspodela s parametrom* $\alpha > 0$:
  $W_{2,\alpha}(x) = 1 - (-x)^{\alpha}, \qquad -1 \leq x \leq 0$
  $W_{2,\alpha}(x) = 0, \qquad x < -1$
  $W_{2,\alpha}(x) = 1, \qquad x > 0$

Ove tri raspodele se nazivaju *generalisane Paretove raspodele*, a ovde su date u α- parametrizaciji.

Kod generalisanih Paretovih raspodela se mogu uvesti parametri položaja i razmere na isti način kao kod raspodela ekstremnih vrednosti.

Osnovni rad iz oblasti generalisanih Paretovih raspodela je:
Residual Life Time at Great Age (Balkema, de Haan (1974)).

O aktuelnosti GPD raspodela govore sledeći radovi objavljeni poslednjih godina:

- *On the Relative Approximation Error of the Generalized Pareto Approximation for a High Quantile* (Beirlant, Raoult, Worms (2003))
- *Quasi-Conjugate Bayes Estimates for GPD Parameters and Application to Heavy Tails Modelling* (Diebolt, El-Aroui, Garrido, Girard (2005))
- *Robust and Efficient Estimation for the Generalized Pareto Distribution* (Juárez, Schucany (2004))

- *Robust Estimation of the Generalized Pareto Distribution* (Peng, Welsh (2001))
- *Rate of Convergence for the Generalized Pareto Approximation of the Excesses* (Raoult, Worms (2003))

γ-parametrizacija

Ako se uvede smena $\gamma = \frac{1}{\alpha}$ kod Paretove raspodele i $\gamma = -\frac{1}{\alpha}$ kod beta raspodele, dobija se jedinstven, neprekidan model, $W_\gamma(x)$, pri čemu važi

$$\lim_{\gamma \to 0} W_\gamma(x) = W_0(x),$$

tj. sve tri funkcije raspodele se mogu prikazati kao familija raspodela koja zavisi samo od jednog parametra oblika, $\gamma > 0$.
U γ-parametrizaciji:

- Eksponencijalna raspodela:
  $W_0(x) = 1 - e^{-x}, \qquad \gamma = 0, \qquad x \geq 0$
- Paretova i beta raspodela:
  $W_\gamma(x) = 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, \qquad x \geq 0 \qquad \gamma > 0$
  $W_\gamma(x) = 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, \qquad x \in [0, -\frac{1}{\gamma}] \qquad \gamma < 0$

Veza izmedju α i γ - parametrizacije

- $W_\gamma(x) = W_{1,\alpha,-\alpha,\alpha}(x), \qquad$ za $\gamma = \frac{1}{\alpha} > 0$
- $W_\gamma(x) = W_{2,\alpha,\alpha,\alpha}(x), \qquad$ za $\gamma = -\frac{1}{\alpha} < 0$

Gustine generalisanih Paretovih raspodela u različitim parametrizacijama

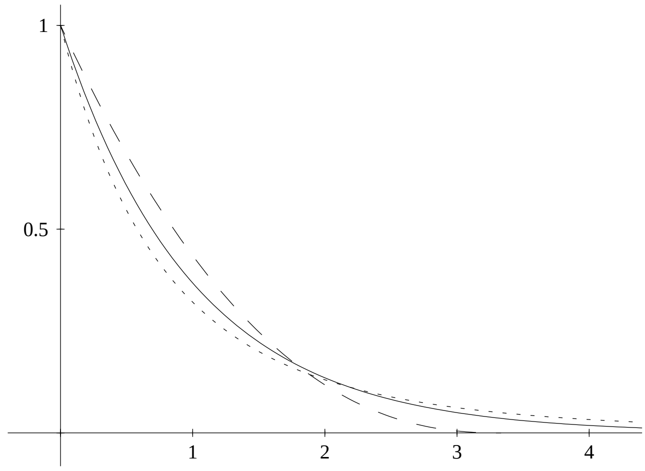
raspodela	$\alpha - parametrizacija$	$\gamma - parametrizacija$
Eksponencijalna raspodela	e^{-x} $x \geq 0$	e^{-x} $x \geq 0$
Paretova raspodela	$\alpha x^{-(1+\alpha)}$ $x \geq 0,\ \alpha > 0$	$(1+\gamma x)^{-(1+\frac{1}{\gamma})}$ $\gamma > 0,\ x \geq 0$
Beta raspodela	$\alpha(-x)^{\alpha-1}$ $-1 \leq x < 0,\ \alpha > 0$	$(1+\gamma x)^{-(1+\frac{1}{\gamma})}$ $\gamma < 0,\ 0 \leq x < -\frac{1}{\gamma}$

Veza izmedju raspodela ekstremnih vrednosti i generalisanih Paretovih raspodela

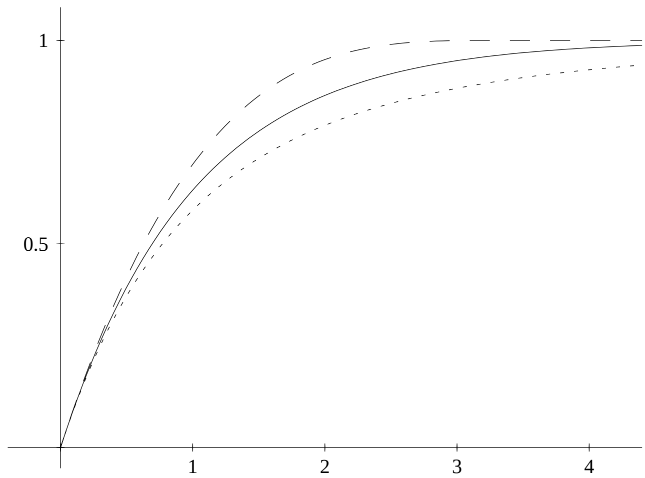
- $W_0(x) = 1 + \ln G_0(x)$, u slučaju kad je $\ln G_0(x) > -1$
- $W_{1,\alpha}(x) = 1 + \ln G_{1,\alpha}(x)$, u slučaju kad je $\ln G_{1,\alpha}(x) > -1$
- $W_{2,\alpha}(x) = 1 + \ln G_{2,\alpha}(x)$, u slučaju kad je $\ln G_{2,\alpha}(x) > -1$

Srednje prekoračenje za generalisane Paretove raspodele u različitim parametrizacijama

raspodela	$\alpha - parametrizacija$	$\gamma - parametrizacija$
Eksponencijalna raspodela	$e_{W_0}(u) = 1$ $u > 0$	$e_{W_\gamma}(u) = \frac{1+\gamma u}{1-\gamma}$ $u > 0,\ \gamma = 0$
Paretova raspodela	$e_{W_{1,\alpha}}(u) = \frac{u}{\alpha-1}$ $u > 1,\ \alpha > 1$	$e_{W_\gamma}(u) = \frac{1+\gamma u}{1-\gamma}$ $u > 0,\ 0 \leq \gamma < 1$
Beta raspodela	$e_{W_{2,\alpha}}(u) = \frac{u}{-\alpha-1}$ $-1 \leq u \leq 0,\ \alpha > 0$	$e_{W_\gamma}(u) = \frac{1+\gamma u}{1-\gamma}$ $0 < u < -\frac{1}{\gamma},\ \gamma < 0$



slika 4: Gustine generalisanih Paretovih raspodela u γ-parametrizaciji: $\gamma = 0$ (puna linija), $\gamma = 0.3$ (tačkasta linija), $\gamma = -0.3$ (isprekidana linija)



slika 5: Funkcije raspodela generalisanih Paretovih raspodela u γ-parametrizaciji: $\gamma = 0$ (puna linija), $\gamma = 0.3$ (tačkasta linija), $\gamma = -0.3$ (isprekidana linija)

2.3 Granične raspodele prekoračenja visokog praga

Neka je X slučajna promenljiva sa funkcijom raspodele F, $R = 1 - F$ i $\varpi(F)$ desni kraj nosača raspodele F . Sledeće dve teoreme daju tipove graničnih raspodela prekoračenja visokog praga za datu funkciju raspodele F. U dokazima se sledi rad [1].

Teorema 2.3.1 (Balkema, de Haan (1974)) *Neka je F funkcija raspodele. Ako postoje funkcije $a(u) > 0$ i $b(u) \in R$, takve da važi:*

$$F^{[u]}(a(u)x + b(u)) \to G(x), \qquad u \to \varpi(F)$$

za nedegenerisanu funkciju raspodele G, tada je funkcija G istog tipa kao jedna od sledećih funkcija raspodele:

$$\begin{aligned} W_0(x) &= 1 - e^{-x}, \quad x \geq 0, \\ W_{1,\alpha}(x) &= 1 - x^{-\alpha}, \quad x \geq 1, \\ T_\gamma(x) &= 1 - e^{-\gamma[1+x]}, \quad x \geq 0, \\ Q_{\gamma,\alpha}(x) &= 1 - e^{-\gamma[1+\alpha \ln(1+x)]}, \quad x \geq 0. \end{aligned}$$

za $\alpha, \gamma > 0$.

Lema 2.3.1 *Ako je $1 - S$ nedegenerisana funkcija raspodele i pri tome važi jednakost:*

$$\lim_{u \to \varpi(F)} P\left\{\frac{X - b(u)}{a(u)} > x | X > u\right\} = S(x), \tag{2}$$

za neke funkcije $a(u) > 0$ i $b(u) \in R$, onda za svaku tačku neprekidnosti y funkcije S, takvu da je $0 < S(y) < 1$, postoje realni brojevi $A(y) \geq 1$ i $B(y)$, za koje važi:

$$S(x)S(y) = S(B(y) + xA(y)) \tag{3}$$

ako je $S(x) < 1$.

Dokaz: Neka je y tačka neprekidnosti funkcije S, takva da je $0 < S(y) < 1$. Ne umanjujući opštost, neka je $y = 0$. Tada je jednakost:

$$\lim_{u \to \varpi(F)} \frac{R(b(u) + xa(u))}{R(u)} = S(x)$$

tačna za $S(x) < 1$. Kad se u izrazu na levoj strani u zameni sa $b(u)$, posle sredjivanja se dobija:

$$\frac{R(b(b(u)) + xa(b(u)))}{R(b(u))} \frac{R(b(u))}{R(u)} = \frac{R(b(u) + \left(\frac{b(b(u)) - b(u)}{a(u)} + x\frac{a(b(u))}{a(u)}\right) a(u))}{R(u)} \tag{4}$$

Ako $u \to \varpi(F)$, onda iz $\frac{R(b(u))}{R(u)} \to S(0) < 1$, za $x \to \varpi(F)$ i činjenice da je R nerastuća funkcija sledi $b(u) \to \varpi(F)$, pa leva strana slabo konvergira ka $S(x)S(0)$ za $S(x) < 1$. Neka je:

$$\lim_{u \to \varpi(F)} \frac{a(b(u))}{a(u)} = A(0) = A, \qquad \lim_{u \to \varpi(F)} \frac{b(b(u)) - b(u)}{a(u)} = B(0) = B \tag{5}$$

Tada je $0 \leq A, B \leq +\infty$, i A i B su konačni, jer bi, u suprotnom, iz (4) sledilo $S(x)S(0) = 0$ za svako x, što nije moguće. Ako je $A > 0$, iz (4) sledi:

$$S(x)S(0) = S(B + xA) \tag{6}$$

za svako x, na osnovu neprekidnosti funkcije S sa desne strane. U slučaju $x = 0$ dobija se:

$$0 < S(B) = S(0)^2 < S(0) < 1.$$

Funkcija S je u tačkama B i 0 neprekidna i uzima različite vrednosti, pa iz toga sledi, na osnovu jednakosti (6) da su brojevi A i B jedinstveno odredjeni. Još ostaje da se dokaže da je $A \geq 1$. Neka je q broj takav da važi $S(0) < q < 1$. Iz $\frac{R(b(u))}{R(u)} \to S(0)$ i činjenice da je R nerastuća funkcija sledi $\frac{R(b(u))}{R(u)} \leq q$, za $u \geq t_0$. Neka je (b_n) niz definisan sa $b_{n+1} = b(b_n)$, za $n = 0, 1, 2, ...$ Niz (b_n) je rastući. Ako bi granična vrednost tog niza, b, bila konačna, tada bi iz $\frac{R(b_{n+1})}{R(b_n)} \leq q$ sledilo $\frac{R(b-)}{R(b-)} \leq q < 1$, što nije moguće. Dakle, pretpostavka ne važi, tj. niz (b_n) neograničeno raste.
Iz (5) sledi $\lim_{n\to\infty} \frac{a(b_{n+1})}{a(b_n)} = A$. Ako bi bilo $A < 1$, tada bi niz $(a(b_n))$ eksponencijalno opadao do nule i na osnovu (5) bi niz (b_n) imao konačnu graničnu vrednost, što je nemoguće. Dakle, početna pretpostavka ne važi, tj. $A \geq 1$. Ovim je dokazano tvrdjenje leme.

Posledica $S(x) > 0$ *za svako* x.

Dokaz: Neka je niz (B_n) definisan sa $B_{n+1} = B + B_n A$ i $B_0 = 0$. Tada iz uslova: $B > 0$ i $A \geq 1$ sledi: $S(B_n) = S(0)^{n+1} > 0$ i $B_n \to \infty$ za $n \to \infty$. Pošto je S nerastuća funkcija, time je posledica dokazana.

Lema 2.3.2 *Neka je H neograničena, nerastuća funkcija na intervalu $(x_0, +\infty)$, takva da je:*

$$H(x) = S(x), \qquad \text{ako je } S(x) < 1.$$

Ako je za neki par konstanti $(A(y), B(y))$ iz leme 2.3.1 ispunjeno:

$$H(x)H(y) = H(B(y) + xA(y)), \qquad \text{za } x > x_0,$$

onda je:

$$H(x) = S(x), \text{ ako je } H(x) < 1.$$

Dokaz: Ne umanjujući opštost, neka je $y = 0$. Tada važi:

$$x_1 = \inf\{x | S(x) < 1\} < 0$$

Može se pretpostaviti da je x_1 konačan broj, jer u suprotnom bi tvrdjenje leme bilo trivijalno ispunjeno. Neka je $(x_1 - \epsilon, x_1)$ leva okolina tačke x_1. Treba dokazati da svaka takva okolina sadrži tačku x_2, takvu da je $H(x_2) \geq 1$.
Ako je $x > x_1$, tada je $H(x) < 1$ i iz uslova leme sledi da je $B(0) + xA(0) > 0$, jer je H nerastuća funkcija. Odatle sledi $B(0) + x_2A(0) > x_1$, za neko $x_2 \in (x_1 - \epsilon, x_1)$. Može se pretpostaviti da je funkcija H neprekidna u tačkama x_2 i $B(0) + x_2A(0)$. Iz uslova (5) sledi da desna strana jednakosti (4) konvergira ka $H(B(0) + x_2A(0))$, a drugi član na levoj strani konvergira ka $H(0)$. Zbog toga konvergira i prvi član na levoj strani, a granična vrednost je $H(x_2)$, na osnovu jednakosti iz uslova leme. Ako bi bilo $H(x_2) < 1$, onda bi bilo $S(x_2) < 1$, što je u suprotnosti sa izborom x_2. Dakle, $H(x_2) \geq 1$.
Ovim je dokaz leme završen.

Dokaz teoreme 2.3.1*:* Neka je X slučajna veličina sa funkcijom raspodele F, $R(x) = 1 - F(x)$, $S(x) = 1 - G(x)$, i neka je Y oznaka za skup tačaka neprekidnosti funkcije S.
Tada je jednakost:

$$S(x)S(y) = S(B(y) + xA(y))$$

tačna za $y \in Y$ i $S(x) < 1$.
Postoje dva slučaja.
1. slučaj: $A = A(y) = 1$, za svako $y \in Y$.
Neka je $g(x) = \ln S(x)$. Tada jednakost (3) postaje:

$$g(x) + g(y) = g(B(y) + x). \tag{7}$$

Neka je $y \in Y$ fiksirano. Važi $g(y) < 0$, odakle sledi $B(y) > 0$ i iz (7) se dobija:

$$g(x) = cx + g_0(x), \tag{8}$$

gde je $c = \frac{g(y)}{B(y)} < 0$, a funkcija g_0 je periodična, sa periodom $B(y)$.

slučaj 1a: Ako je funkcija g_0 konstantna, onda važi $g(x) = c(x + d)$, gde je d konstanta. Neka je funkcija H definisana sa $H(x) = e^{c(x+d)}$. Tada važi $H(x) = S(x)$, na osnovu leme 2.3.2, i funkcija G je istog tipa kao funkcija raspodele W_0.

slučaj 1b: Ako funkcija g_0 nije konstantna i ima osnovni period $p > 0$, tada iz (8) sledi $g(x+p) - g(x) = cp$. Funkcija g_0 ima period $B(y) = \frac{g(y)}{c}$, a osnovni period je p, pa mora da važi $g(y) \in \{kcp, k \in N\}$, za svako $y \in Y$. Iz toga sledi da je $g(x) = cp[\frac{x+d}{p}]$, gde je d konstanta. Neka je funkcija H definisana sa $H(x) = e^{cp[\frac{x+d}{p}]}$ za $x \in R$. Na osnovu leme 2.3.2 sledi da je $H = G$, tj. funkcija raspodele G istog tipa kao funkcija T_γ, sa parametrom $\gamma = -cp > 0$.

2. slučaj: kada je $A(y) > 1$ bar za jedno $y \in Y$.
Neka je $g(x) = \ln S(x)$, i $y, z \in Y$. Neka je:

$$\alpha(x) = B(y) + xA(y),$$
$$\beta(x) = B(z) + xA(z).$$

Iz jednakosti (3) se dobija:

$$g(\alpha(x)) = g(x) + g(y),$$
$$g(\beta(x)) = g(x) + g(z),$$

za svako $x < \varpi(F)$.

Odatle sledi da je $g(\alpha(\beta(x))) = g(\beta(\alpha(x)))$, za svako x, a pošto je funkcija g nerastuća i neograničena, mora biti $\alpha(\beta) = \beta(\alpha)$.
To znači:

$$B(y) + A(y)B(z) = B(z) + A(z)B(y)$$

ili:

$$B(y)(1 - A(z)) = B(z)(1 - A(y)).$$

Ako važi $A(y) > 1$ bar za jedno $y \in Y$, onda iz poslednje jednakosti sledi da to mora da važi za svako $y \in Y$, pa se dobija:

$$\frac{B(y)}{1 - A(y)} = \frac{B(z)}{1 - A(z)},$$

a pošto su y i z proizvoljno odabrani, sledi da izraz $m = \frac{B(y)}{1-A(y)}$ ne zavisi od y i važi:

$$B(y) + xA(y) = m + (x - m)A(y).$$

Ako je $m = 0$, zamenom poslednje jednakosti u (3) se dobija:

$$g(x) + g(y) = g(xA(y)).$$

Važi $g(y) < 0$, pa mora biti $xA(y) > x$, za svako $y \in Y$, jer je g nerastuća funkcija. Odatle sledi $x > 0$, pa se može uvesti smena $x = e^t$. Jednakost (3) tada postaje:

$$h(t) + h(s) = h(t + \alpha(s))$$

za $t > t_0$, a to je , ponovo, jednakost (7).
Postoje dva slučaja:
slučaj $2a$:

$$h(t) = c(t+d),$$

gde je d konstanta,

$$H(x) = e^{c(t+d)} = e^{ct}e^{dc} = (e^t)^c e^{dc} = (xe^d)^c,$$

za $x > 0$, pa je funkcija raspodele G tipa $W_{1,\alpha}$, sa parametrom $\alpha = -c$.

slučaj $2b$:

$$h(t) = cp\left[\frac{t+d}{p}\right],$$

gde je d konstanta,

$$H(x) = e^{cp\left[\frac{t+d}{p}\right]} = e^{cp\left[\frac{\ln e^d x}{p}\right]},$$

za $x > 0$ i funkcija raspodele G je tipa $Q_{\gamma,\alpha}$, sa parametrima $\gamma = -cp$, $\alpha = \frac{1}{p}$.

Ostaje da se dokaže da za sve moguće granične raspodele S važi jednakost:

$$\min\left\{1, \frac{S(b_u + xa_u)}{S(u)}\right\} = S(x), \qquad u > 0,$$

za neke konstante $a_u > 0$ i $b_u \in R$, tj. da su sve one POT-stabilne.

Za funkciju W_0 važi:

$$W_{0,0,\sigma}^{[u]}(x) = \frac{W_{0,0,\sigma}(x) - W_{0,0,\sigma}(u)}{1 - W_{0,0,\sigma}(u)} = \frac{e^{-\frac{u}{\sigma}} - e^{-\frac{x}{\sigma}}}{e^{-\frac{u}{\sigma}}} = 1 - e^{-\frac{x}{u}+\frac{u}{\sigma}}$$

$$= 1 - e^{-\frac{x-u}{\sigma}} = W_{0,u,\sigma}(x),$$

a normirajuće konstante su $a_u = 1$ i $b_u = u$.

Za funkciju $W_{1,\alpha}$ važi:

$$W_{1,\alpha,0,\sigma}^{[u]}(x) = \frac{W_{1,\alpha,0,\sigma}(x) - W_{1,\alpha,0,\sigma}(u)}{1 - W_{1,\alpha,0,\sigma}(u)} = \frac{\left(\frac{u}{\sigma}\right)^{-\alpha} - \left(\frac{x}{\sigma}\right)^{-\alpha}}{\left(\frac{u}{\sigma}\right)^{-\alpha}}$$

$$= 1 - \left(\frac{x}{u}\right)^{-\alpha} = W_{1,\alpha,0,u}(x),$$

za $\quad u \geq 1$, pa su normirajuće konstante: $a_u = \frac{u}{\sigma}$ i $b_u = 0$.

Za funkciju T_γ važi:

$$T_\gamma^{[u]}(x) = \frac{1 - e^{-\gamma[1+x]} - \left(1 - e^{-\gamma[1+u]}\right)}{1 - \left(1 - e^{-\gamma[1+u]}\right)} = 1 - \frac{e^{-\gamma[1+x]}}{e^{-\gamma[1+u]}}$$
$$= 1 - e^{-\gamma([1+x]-[1+u])} = 1 - e^{-\gamma[1+x-[1+u]]} = T_{\gamma,[1+u],1}(x),$$

tj. normirajuće konstante su $a_u = 1$ i $b_u = [1+u]$.

Za funkciju $Q_{\gamma,\alpha}$ važi:

$$Q_{\gamma,\alpha}^{[u]}(x) = \frac{1 - e^{-\gamma[1+\alpha\ln(1+x)]} - \left(1 - e^{-\gamma[1+\alpha\ln(1+u)]}\right)}{1 - \left(1 - e^{-\gamma[\alpha\ln(1+u)]}\right)}$$
$$= 1 - \frac{e^{-\gamma[1+\alpha\ln(1+x)]}}{e^{-\gamma[1+\alpha\ln(1+u)]}} = 1 - e^{-\gamma([1+\alpha\ln(1+x)]-[1+\alpha\ln(1+u)])}$$
$$= 1 - e^{-\gamma\left[\alpha\ln(1+x)-\alpha\ln\left(e^{\frac{[\alpha\ln(1+u)]}{\alpha}}\right)\right]} = 1 - e^{-\gamma\left[\alpha\ln\left((1+x)e^{\frac{-[\alpha\ln(1+u)]}{\alpha}}\right)\right]}$$
$$= 1 - e^{-\gamma\left[1+\alpha\ln\left((1+x)e^{\frac{-[\alpha\ln(1+u)]}{\alpha}}\right)-\alpha\ln e^{\frac{1}{\alpha}}\right]} = 1 - e^{-\gamma\left[1+\alpha\ln\frac{(1+x)e^{\frac{-[\alpha\ln(1+u)]}{\alpha}}}{e^{\frac{1}{\alpha}}}\right]}$$
$$= 1 - e^{-\gamma\left[1+\alpha\ln\frac{1+x}{e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}}\right]} = 1 - e^{-\gamma\left[1+\alpha\ln\left(1+\frac{1+x-e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}}{e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}}\right)\right]}$$
$$= 1 - e^{-\gamma\left[1+\alpha\ln\left(1+\frac{x-(e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}-1)}{e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}}\right)\right]} = Q_{\gamma,\alpha,\mu,\sigma}(x),$$

gde je:

$$\mu = e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}} - 1, \quad \sigma = e^{\frac{[\alpha\ln(1+u)]+1}{\alpha}}.$$

Ovim je dokaz teoreme završen.

Pod odredjenim uslovima, funkcija raspodele prekoračenja praga u, za datu funkciju raspodele F, može se proizvoljno tačno aproksimirati jednom od generalisanih Paretovih raspodela. Tu se kao jedna mogućnost pojavljuje i beta raspodela, iako ne spada u granične raspodele iz prethodne teoreme.

Teorema 2.3.2 (Balkema, de Haan (1974)) *Neka je F apsolutno neprekidna funkcija raspodele, a γ i δ realni brojevi. Ako važe uslovi:*

1. $f = F'$ *je pozitivna, diferencijabilna funkcija na intervalu* $[u_0, +\infty)$
2. $\gamma \leq \frac{d}{du}\left(\frac{1-F(u)}{f(u)}\right) \leq \delta$ *za* $u \geq u_0$,

tada važi:

$$W_\gamma(x) \leq F^{[u]}(a(u)x) \leq W_\delta(x),$$

gde je $a(u) = \frac{1-F(u)}{f(u)}$.

Dokaz: Za $u \geq u_0$ i $x \geq 0$, iz 2. uslova teoreme, integracijom po u, u granicama od u do $u + xa(u)$, dobija se:

$$\gamma x a(u) \leq a(u + xa(u)) - a(u) \leq \delta x a(u),$$

odakle sledi:

$$1 + \gamma x \leq \frac{a(u + xa(u))}{a(u)} \leq 1 + \delta x \tag{9}$$

Važi:

$$F^{[u]}(a(u)x) = \frac{1 - F(u + xa(u))}{1 - F(u)} = \exp\left\{\ln \frac{1 - F(u + xa(u))}{1 - F(u)}\right\}$$

$$= \exp\left\{\int_0^x \frac{\frac{\partial}{\partial s}\left(\frac{1-F(u+sa(u))}{1-F(u)}\right)}{\frac{1-F(u+sa(u))}{1-F(u)}} ds\right\}$$

$$= \exp\left\{\int_0^x \frac{(-a(u)f(u + sa(u)))}{1 - F(u + sa(u))} ds\right\} = \exp\left\{-\int_0^x \frac{a(u)}{a(u + sa(u))} ds\right\},$$

pri čemu je korišćena formula:

$$\ln g(x) = \int_0^x \frac{g'(x)}{g(x)} dx.$$

Zamenom poslednje jednakosti u (9) dobija se:

$$\exp\left\{-\int_0^x \frac{1}{1 + \gamma s} ds\right\} \leq F^{[u]}(a(u)x) \leq \exp\left\{-\int_0^x \frac{1}{1 + \delta s} ds\right\},$$

odakle sledi tvrdjenje teoreme.

Sledeće tvrdjenje je neposredna posledica teorema 2.3.1 i 2.3.2:

Teorema 2.3.3 (Balkema, de Haan (1974)) *Neka je F funkcija raspodele. Ako važi:*

$$F^{[u]}(a(u)x + b(u)) \to G(x), \qquad u \to \varpi(F),$$

za neprekidnu funkciju raspodele G i neke funkcije $a(u) > 0$ *i* $b(u) \in R$*, tada je funkcija G istog tipa kao jedna od sledećih funkcija raspodele:* W_0 *ili* W_γ*,* $\gamma \in R$*.*

2.4 POT-oblasti privlačenja graničnih raspodela prekoračenja visokog praga

Teorema 2.4.1 *Ako je D_0 skup funkcija raspodele F, takvih da $F(x) < 1$ za svako x, onda važi:*

$$D_p(W_0) = D(G_0) \cap D_0.$$

Lema 2.4.1 *Ako važi $F \in D_p(G)$, gde je G neprekidna funkcija raspodele, onda je:*

$$\lim_{u \to \varpi(F)} \frac{1 - F(u-)}{1 - F(u+)} = 1.$$

*Dokaz:*Neka je $R = 1 - F$. Za svako $q < 1$ postoji y, takvo da važi $q < 1 - G(y) < 1$. Zamenom tako odredjenog y u jednakost (2) dobija se:

$$\lim_{u \to \varpi(F)} \min \left\{ 1, \frac{R(b(u) + ya(u))}{R(u)} \right\} = 1 - G(y) < 1,$$

odakle sledi da su za $u \geq u_1$ tačne sledeće nejednakosti:

$$\begin{aligned} R(b(u) + ya(u)) &< R(u), \\ b(u) + ya(u) &> u. \end{aligned}$$

Na isti način, iz $1 - G(y) > q$ sledi:

$$R(b(u) + ya(u)) > qR(u), \text{ za } u > u_2.$$

Za $u \geq \max\{u_1, u_2\}$ važe sve navedene nejednakosti, pa se dobija:

$$1 - F(u+) > q(1 - F(u-)).$$

Pošto je $q < 1$ proizvoljno odabran broj, ovim je tvrdjenje leme dokazano.

Dokaz teoreme 2.4.1*:* Ako je $F \in D(G_0) \cap D_0$, onda važi:

$$\lim_{n \to \infty} F^n(a_n x + b_n) = \exp\left\{ -e^{-x} \right\},$$

odakle sledi:

$$\lim_{n \to \infty} n \ln F(a_n x + b_n) = -e^{-x},$$

ili, malo drugačije:

$$\lim_{n \to \infty} n \ln(1 - (1 - F(a_n x + b_n))) = -e^{-x} \tag{10}$$

Važi jednakost:
$$\lim_{n\to\infty}(1-F(a_nx+b_n))=0,$$
jer bi, u suprotnom, za svako $\delta>0$ postojao podniz (n_k) prirodnih brojeva, takav da:
$$1-F(a_{n_k}x+b_{n_k})\geq\delta,$$
pa bi bilo ispunjeno i:
$$\lim_{n\to\infty}F^n(a_{n_k}x+b_{n_k})=0,$$
što ne odgovara pretpostavci.
Iz:
$$\lim_{n\to\infty}(1-F(a_nx+b_n))=0$$
sledi:
$$\lim_{n\to\infty}n\ln(1-(1-F(a_nx+b_n)))\sim-(1-F(a_nx+b_n))$$
pa jednakost (10) postaje:
$$\lim_{n\to\infty}n(1-F(a_nx+b_n))=e^{-x}$$
za svako x. Neka je $a(u)=a_n$, $b(u)=b_n$, za svako u za koje je ispunjeno:
$$\frac{1}{n+1}\leq 1-F(u)<\frac{1}{n}.$$
Tada je tačna jednakost:
$$\lim_{u\to+\infty}\frac{1-F(b(u)+xa(u))}{1-F(u)}=e^{-x},$$
tj. $F\in D_p(W_0)$.
Ako je $F\in D_p(W_0)$, onda važi:
$$\lim_{u\to\infty}\frac{1-F(b(u)+xa(u))}{1-F(u)}=e^{-x}$$
za $x>0$. Na osnovu leme 2.4.1, mogu se odabrati normirajuće konstante a_n i b_n, takve da je ispunjeno:
$$\lim_{n\to\infty}n(1-F(b_n+xa_n))=e^{-x}, \tag{11}$$
za $x>0$. Još treba dokazati da poslednja jednakost važi za svako $x\in R$.
Neka je c najmanji realan broj za koji je jednakost (11) ispunjena na intervalu $(c,+\infty)$. Važi sledeća jednakost:
$$2n(1-F(b_{2n}+xa_{2n}))=2n\left(1-F\left(b_n+\left(\frac{b_{2n}-b_n}{a_n}+x\frac{a_{2n}}{a_n}\right)a_n\right)\right) \tag{12}$$

Kad $n \to \infty$, leva strana jednakosti konvergira ka e^{-x}, a desna strana konvergira ka $2e^{-(B+xA)}$, gde su A i B konstante. Sredjivanjem poslednje jednakosti i uporedjivanjem koeficijenata uz e^{-x} sa jedne i druge strane jednakosti, dobija se $A = 1$ i $B = \ln 2$. Po pretpostavci, desna strana jednakosti (12) konvergira za $x + \ln 2 > c$, pa onda leva strana jednakosti konvergira za $x > c - \ln 2$. Sve to važi za podniz parnih brojeva, ali zbog toga što je $2n + 1 \sim 2n$, kad $n \to \infty$, konvergencija se može proširiti na ceo niz prirodnih brojeva, sa novim normirajućim konstantama $a'_{2n+1} = a'_{2n} = a_{2n}$ i $b'_{2n+1} = b'_{2n} = b_{2n}$. Onda mora je tačno i $\frac{a'_n}{a_n} \to 1$ i $\frac{b'_n - b_n}{a_n} \to 1$, iz čega sledi da konvergencija važi i za početne normirajuće konstante. To znači da je $c = c - \ln 2$, što je ispunjeno samo za $c = -\infty$. Ovim je teorema dokazana.

Teorema 2.4.2 $D_p(W_{1,\alpha}) = D(G_{1,\alpha})$, *za svako* $\alpha > 0$.

Dokaz: Ako je $F \in D(G_{1,\alpha})$, onda važi:

$$\lim_{n\to\infty} n(1 - F(b_n + xa_n)) = x^{-\alpha}$$

za svako $x > 0$. Neka je $a(u) = a_n$, $b(u) = b_n$, za svako u za koje je ispunjeno:

$$\frac{1}{n+1} \leq 1 - F(u) < \frac{1}{n}.$$

Tada je jednakost:

$$\lim_{u\to+\infty} \frac{1 - F(b(u) + xa(u))}{1 - F(u)} = x^{-\alpha},$$

tačna za svako $x > 0$, tj. $F \in D_p(W_{1,\alpha})$.
Ako je $F \in D_p(W_{1,\alpha})$, onda važi:

$$\lim_{u\to\infty} \frac{1 - F(b(u) + xa(u))}{1 - F(u)} = x^{-\alpha}$$

za $x > 1$. Na osnovu leme 2.4.1, mogu se odabrati normirajuće konstante a_n i b_n, takve da je ispunjeno:

$$\lim_{n\to\infty} n(1 - F(b_n + xa_n)) = x^{-\alpha}, \tag{13}$$

za $x > 1$. Još treba dokazati da je poslednja jednakost tačna za svako $x > 0$. Neka je c najmanji realan broj za koji je jednakost (13) ispunjena na intervalu $(c, +\infty)$. Važi sledeća jednakost:

$$2n\left(1 - F\left(b_{2n} + xa_{2n}\right)\right) = 2n\left(1 - F\left(b_n + \left(\frac{b_{2n} - b_n}{a_n} + x\frac{a_{2n}}{a_n}\right)a_n\right)\right) \tag{14}$$

Kad $n \to \infty$, leva strana jednakosti konvergira ka $x^{-\alpha}$, a desna strana konvergira ka $2(Ax+B)^{-\alpha}$, gde su A i B konstante. Sredjivanjem poslednje jednakosti i uporedjivanjem koeficijenata uz x sa jedne i druge strane jednakosti, dobija se $A = 2^{\frac{1}{\alpha}}$ i $B = 0$. Po pretpostavci, desna strana jednakosti (14) konvergira za $2^{\frac{1}{\alpha}}x > c$, pa onda leva strana jednakosti konvergira za $x > c(\frac{1}{2})^{\frac{1}{\alpha}}$. Sve to važi za podniz parnih brojeva, ali zbog toga što je $2n+1 \sim 2n$, kad $n \to \infty$, konvergencija se može proširiti na ceo niz prirodnih brojeva, sa novim normirajućim konstantama $a'_{2n+1} = a'_{2n} = a_{2n}$ i $b'_{2n+1} = b'_{2n} = b_{2n}$. Onda je tačno i $\frac{a'_n}{a_n} \to 1$ i $\frac{b'_n - b_n}{a_n} \to 1$, iz čega sledi da konvergencija važi i za početne normirajuće konstante. To znači da je $c = c(\frac{1}{2})^{\frac{1}{\alpha}}$, što je ispunjeno samo za $c = 0$. Ovim je dokaz teoreme završen.

Primeri raspodela iz svake od POT-oblasti privlačenja generalisanih Paretovih raspodela

raspodela	oblast privlačenja	normirajuće konstante
Gumbelova raspodela $F(x) = \exp(-\exp(-x))$	$D_p(W_0)$	$a_u = 1,\ b_u = u$
Eksponencijalna raspodela $F(x) = 1 - \exp(-\frac{x}{\lambda})$	$D_p(W_0)$	$a_u = \lambda,\ b_u = u$
Frešeova raspodela $F(x) = \exp(-x^{-\alpha})$	$D_p(W_{1,\alpha})$	$a_u = u,\ b_u = u$
Košijeva raspodela $F(x) = \frac{1}{2} + \frac{1}{\pi} arctgx$	$D_p(W_{1,1})$	$a_u = u,\ b_u = -u$
Vejbulova raspodela $F(x) = \exp(-(-x)^{\alpha})$	$D_p(W_{2,\alpha})$	$a_u = u,\ b_u = -u$
Ravnomerna raspodela na $[0, c]$ $F(x) = \frac{x}{c}$	$D_p(W_{2,1})$	$a_u = c - u,\ b_u = c$

3. poglavlje

GPD raspodele i primene

3.1 GPD raspodele i veza sa problemima čekanja

U literaturi se navode i sledeća dva tipa kombinatornih problema čekanja:

1. Izvodi se niz nezavisnih eksperimenata $\xi_1, \xi_2, ...$, čiji rezultati se opisuju nizom nezavisnih slučajnih veličina $Z_1, Z_2, ...$ i posmatra se broj eksperimenata (opisan slučajnom veličinom Y_n) koji je potrebno izvesti da bi se realizovali svi dogadjaji iz nekog skupa $A = \{A_1, A_2, ..., A_n\}$. Problem se sastoji u odredjivanju tačne raspodele slučajne veličine Y_n ili njene asimptotske raspodele kad $n \to \infty$.

2. Izvodi se niz nezavisnih eksperimenata $\xi_1, \xi_2, ...$, čiji rezultati se opisuju nizom nezavisnih slučajnih veličina $Z_1, Z_2, ...$ i posmatra se broj eksperimenata koji je potrebno izvesti da bi se svi dogadjaji iz nekog skupa $A = \{A_1, A_2, ..., A_n\}$ realizovali tačno r puta, gde je $r \geq 2$. Ako je taj broj eksperimenata opisan slučajnom veličinom $Y_{n,r}$, treba odrediti tačnu ili asimptotsku raspodelu slučajne veličine $Y_{n,r}$.

Postoji više pristupa rešavanju ovih problema, a u jednom od njih se primenjuje teorija ekstremnih vrednosti. Ako se sa X_k ($X_{k,r}$) označi broj izvedenih eksperimenata do prve (r-te) realizacije dogadjaja $A_k \in A$, $k \in \{1, 2, ..., n\}$, onda slučajna veličina $M_n = \max\{X_1, X_2, ..., X_n\}$ predstavlja vreme čekanja do realizacije svih dogadjaja iz skupa A, a slučajna veličina $M_{n,r} = \max\{X_{1,r}, X_{2,r}, ..., X_{n,r}\}$ predstavlja vreme čekanja do realizacije svih dogadjaja iz istog skupa po r puta. U ovom slučaju, slučajne veličine X_i i $X_{i,r}$, $i \in \{1, 2, ..., n\}$ su zavisne, ali uz neke dodatne uslove (Lidbeterovi uslovi pomešanosti D(u_n) i D'(u_n)) može da se koristi isti metod za odredjivanje asimptotske raspodele kao za nezavisne slučajne veličine.

Ovde će biti razmatran slučaj kada slučajne veličine $Z_1, Z_2, ...$ imaju zajedničku ravnomernu raspodelu na skupu $\mathbf{N_n} = \{1, 2, ..., n\}$.

Nezavisne slučajne veličine sa geometrijskom i negativnom binomnom raspodelom

Poznato je da geometrijska raspodela i negativna binomna raspodela ne pripadaju nijednoj od oblasti privlačenja raspodela ekstremnih vrednosti. Ali, u slučaju kada parametar geometrijske , odnosno, negativne binomne raspodele zavisi od broja slučajnih veličina, n, i ako taj parametar teži nuli, kad n neograničeno raste, mogu se odrediti normirajuće konstante, takve da je granična raspodela maksimuma ovih slučajnih veličina Gumbelova raspodela.

Nezavisne slučajne veličine sa zajedničkom geometrijskom raspodelom

Neka je $X^*_{n,1}, X^*_{n,2}, ...$ niz nezavisnih slučajnih veličina sa zajedničkom geometrijskom raspodelom:

$$P\{X^*_{n,j} = k\} = \frac{1}{n}\left(1 - \frac{1}{n}\right)^{k-1}, \quad k = 1, 2, ..., \quad 1 \leq j \leq n, \quad n \in N$$

i

$$M^*_n = \max\{X^*_{n,1}, ..., X^*_{n,n}\}.$$

Teorema 3.1.1 (Mladenović (1999)) 1. *Za svaki realan broj x važi jednakost:*

$$\lim_{n\to\infty} P\{M^*_n \leq n(x + \ln n)\} = \exp\{-e^{-x}\}$$

2. *Za svaki realan broj x pri $n \to \infty$ važi jednakost:*

$$P\{M^*_n \leq n(x + \ln n)\} - \exp\{-e^{-x}\} \sim \frac{e^{-x}\exp\{-e^{-x}\}}{2}\frac{\ln n}{n}.$$

Teorema 3.1.2 *Neka je $M^*_{n,1}, ..., M^*_{n,n}, ..$ niz nezavisnih slučajnih veličina sa zajedničkom funkcijom raspodele:*

$$F(x) = P\{M^*_n \leq x\}$$

i

$$S_n^* = \max\{M_{n,1}^*, ..., M_{n,n}^*\}.$$

Za svaki realan broj x važe jednakosti:

1.

$$\lim_{n\to\infty} P\{S_n^* \le n(x + 2\ln n)\} = \exp\{-e^{-x}\}$$

2.

$$\lim_{n\to\infty} P\{M_n^* \le n(x + 2\ln n) | M_n^* > 2n\ln n\} = 1 - e^{-x}$$

Dokaz: Neka je $r_n = n(x + 2\ln n) - [n(x + 2\ln n)]$.

$$P\{X_{n,1}^* \le k\} = 1 - P\{X_{n,1}^* > k\} = 1 - \sum_{s=k+1}^{+\infty} \frac{1}{n}\left(1 - \frac{1}{n}\right)^{s-1}$$

$$= 1 - \frac{1}{n}\sum_{s=k+1}^{+\infty}\left(1 - \frac{1}{n}\right)^{s-1} = 1 - \frac{1}{n}\left(1 - \frac{1}{n}\right)^{k}\frac{1}{1 - \left(1 - \frac{1}{n}\right)}$$

$$= 1 - \left(1 - \frac{1}{n}\right)^{k}$$

1.

$$P\{S_n^* \le n(x + 2\ln n)\} = (P\{M_{n,1}^* \le n(x + 2\ln n)\})^n$$

$$= \left(\left(1 - \left(1 - \frac{1}{n}\right)^{[n(x+2\ln n)]}\right)^n\right)^n = \left(1 - \left(1 - \frac{1}{n}\right)^{[n(x+2\ln n)]}\right)^{n^2}$$

$$= \left(1 - e^{[n(x+2\ln n)]\ln\left(1-\frac{1}{n}\right)}\right)^{n^2} = \left(1 - e^{[n(x+2\ln n)]\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^{n^2}$$

$$= \left(1 - e^{(n(x+2\ln n)-r_n)\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^{n^2} = \left(1 - \frac{e^{-x}}{n^2}(1 + o(1))\right)^{n^2}$$

$$\sim \exp\{-e^{-x}\}, \qquad n \to \infty.$$

2.

$$P\{M_n^* \le n(x + 2\ln n)\} = (P\{X_{n,1}^* \le n(x + 2\ln n)\})^n$$

$$= \left(1 - \left(1 - \frac{1}{n}\right)^{[n(x+2\ln n)]}\right)^n = \left(1 - e^{[n(x+2\ln n)]\ln\left(1-\frac{1}{n}\right)}\right)^n$$

$$= \left(1 - e^{[n(x+2\ln n)]\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^n = \left(1 - e^{(n(x+2\ln n)-r_n)\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^n$$

$$= \left(1 - e^{-x-2\ln n+o\left(\frac{\ln n}{n}\right)}\right)^n = 1 - e^{-x-\ln n} + O\left(\frac{1}{n^2}\right), \quad n \to \infty.$$

$$
\begin{aligned}
P\left\{M_n^* \le n(x+2\ln n) | M_n^* > 2n\ln n\right\} &= \frac{P\left\{M_n^* \le n(x+2\ln n), M_n^* > 2n\ln n\right\}}{P\left\{M_n^* > 2n\ln n\right\}} \\
&= \frac{P\left\{M_n^* \le n(x+2\ln n)\right\} - P\left\{M_n^* \le 2n\ln n\right\}}{P\left\{M_n^* > 2n\ln n\right\}} \\
&= \frac{P\{M_n^* \le n(x+2\ln n)\} - 1 + 1 - P\{M_n^* \le 2n\ln n\}}{P\{M_n^* > 2n\ln n\}} \\
= 1 - \frac{1 - P\{M_n^* \le n(x+2\ln n)\}}{1 - P\{M_n^* \le 2n\ln n\}} &= 1 - \frac{1 - \left(1 - ne^{-x-2\ln n} + O\left(\frac{1}{n^2}\right)\right)}{1 - \left(1 - ne^{-2\ln n} + O\left(\frac{1}{n^2}\right)\right)} \\
\sim 1 - \frac{1 - \left(1 - e^{-x-\ln n} + O\left(\frac{1}{n^2}\right)\right)}{1 - \left(1 - e^{-\ln n} + O\left(\frac{1}{n^2}\right)\right)} &\sim 1 - \frac{e^{-x-\ln n} + O\left(\frac{1}{n^2}\right)}{e^{-\ln n} + O\left(\frac{1}{n^2}\right)} \\
&\sim 1 - e^{-x}, \qquad n \to \infty.
\end{aligned}
$$

Teorema 3.1.3 *Neka je data shema serija slučajnih veličina:*

$$
\begin{gathered}
X_{n,1,1}, X_{n,2,1}, ..., X_{n,n,1}; \\
X_{n,1,2}, X_{n,2,2}, ..., X_{n,n,2}; \\
\cdots\cdots\cdots \\
X_{n,1,n}, X_{n,2,n}, ..., X_{n,n,n}; \\
\cdots\cdots\cdots
\end{gathered}
$$

gde su slučajne veličine u i-toj vrsti nezavisne i jednako raspodeljene.
Neka su ispunjeni uslovi:
1. *slučajne veličine* $X_{n,1,1}, ..., X_{n,n,1}$ *imaju geometrijsku raspodelu zadatu sa:*

$$
P\{X_{n,1,1} = k\} = \frac{1}{n}\left(1 - \frac{1}{n}\right)^{k-1}, \qquad k = 1, 2, ..
$$

2. $X_{n,1,i} = \max\{X_{n,1,i-1}, X_{n,2,i-1}, ..., X_{n,n,i-1}\}, \qquad$ *za* $i = 2, 3, ...,$
a slučajne veličine $X_{n,2,i}, X_{n,3,i}, ..., X_{n,n,i}$ *imaju istu raspodelu kao* $X_{n,1,i}$ *i medjusobno su nezavisne.*
Tada za svaki realan broj x važe jednakosti:
1.
$$
\lim_{n\to\infty} P\{\max\{X_{n,1,i}, ..., X_{n,n,i}\} \le n(x + i\ln n)\} = \exp\left\{-e^{-x}\right\}
$$
2.
$$
\lim_{n\to\infty} P\{X_{n,1,i} \le n(x + i\ln n) | X_{n,1,i} > in\ln n\} = 1 - e^{-x}
$$

za $i = 1, 2, ...$

Dokaz: Funkcija raspodele slučajnih veličina u prvoj vrsti je:

$$F_1(x) = 1 - \left(1 - \frac{1}{n}\right)^x,$$

u drugoj vrsti:

$$F_2(x) = (F_1(x))^n = \left(1 - \left(1 - \frac{1}{n}\right)^x\right)^n,$$

u trecoj vrsti:

$$F_3(x) = (F_2(x))^n = \left(\left(1 - \left(1 - \frac{1}{n}\right)^x\right)^n\right)^n = \left(1 - \left(1 - \frac{1}{n}\right)^x\right)^{n^2},$$

....
u i-toj vrsti:

$$F_i(x) = (F_{i-1}(x))^n = \left(\left(1 - \left(1 - \frac{1}{n}\right)^x\right)^{n^{i-2}}\right)^n = \left(1 - \left(1 - \frac{1}{n}\right)^x\right)^{n^{i-1}}.$$

1.

$$\lim_{n\to\infty} P\{\max\{X_{n,1,i}, ..., X_{n,n,i}\} \leq n(x + i\ln n)\} = \lim_{n\to\infty} (F_i(n(x + i\ln n)))^n$$

$$= \left(1 - \left(1 - \frac{1}{n}\right)^{n(x+i\ln n)}\right)^{n^i} = \left(1 - e^{[n(x+i\ln n)]\ln\left(1-\frac{1}{n}\right)}\right)^{n^i}$$

$$= \left(1 - e^{[n(x+i\ln n)]\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^{n^i} = \left(1 - e^{(n(x+i\ln n)-r_n)\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)}\right)^{n^i}$$

$$= \left(1 - e^{-x-i\ln n+O\left(\frac{\ln n}{n}\right)}\right)^{n^i} \sim \left(1 - n^{i-1}e^{-x-i\ln n+O\left(\frac{\ln n}{n}\right)}\right)^n, \quad n \to \infty$$

$$= \left(1 - e^{-x-\ln n+O\left(\frac{\ln n}{n}\right)}\right)^{\left(-e^{x+\ln n+O\left(\frac{\ln n}{n}\right)}\right)^{-\frac{n}{e^{x+\ln n+O\left(\frac{\ln n}{n}\right)}}}} \sim \exp\{-e^{-x}\},$$

za $\qquad n \to \infty.$
2.

$$P\{X_{n,1,i} \leq n(x + i\ln n) | X_{n,1,i} > in\ln n\}$$
$$= \frac{P\{X_{n,1,i} \leq n(x + i\ln n), X_{n,1,i} > in\ln n\}}{P\{X_{n,1,i} > in\ln n\}}$$
$$= \frac{P\{X_{n,1,i} \leq n(x + i\ln n)\} - P\{X_{n,1,i} \leq in\ln n\}}{P\{X_{n,1,i} > in\ln n\}}$$

$$= \frac{P\{X_{n,1,i} \leq n(x + i\ln n)\} - 1 + 1 - P\{X_{n,1,i} \leq in\ln n\}}{P\{X_{n,1,i} > in\ln n\}}$$

$$= 1 - \frac{1 - P\{X_{n,1,i} \leq n(x + i\ln n)\}}{1 - P\{X_{n,1,i} \leq in\ln n\}} \sim 1 - \frac{1 - \left(1 - n^{i-1}e^{-x-i\ln n} + O\left(\frac{1}{n^2}\right)\right)}{1 - \left(1 - n^{i-1}e^{-i\ln n} + O\left(\frac{1}{n^2}\right)\right)}$$

$$\sim 1 - \frac{e^{-x-\ln n} + O\left(\frac{1}{n^2}\right)}{e^{-\ln n} + O\left(\frac{1}{n^2}\right)} \sim 1 - e^{-x}, \qquad n \to \infty.$$

Nezavisne slučajne veličine sa zajedničkom negativnom binomnom raspodelom

Neka je $X^*_{n,1}, X^*_{n,2}, ...$ niz nezavisnih slučajnih veličina sa zajedničkom negativnom binomnom raspodelom:

$$P\{X^*_{n,j} = k\} = \binom{k-1}{n-1}\left(1 - \frac{1}{n}\right)^{k-r}\frac{1}{n^r},$$

$$k = r, r+1, ..., \quad 1 \leq j \leq n, \quad n \in N,$$

i neka je:

$$M^*_n = \max\{X^*_{n,1}, ..., X^*_{n,n}\},$$

$$u_n = n(x + \ln n + (r-1)\ln\ln n - \ln(r-1)!)$$

Teorema 3.1.4 (Mladenović (1999)) 1. *Za svaki realan broj x važi jednakost:*

$$\lim_{n\to\infty} P\{M^*_n \leq u_n\} = \exp\{-e^{-x}\}$$

2. *Ako je $r \geq 2$, onda za svaki realan broj x pri $n \to \infty$ važi jednakost:*

$$P\{M^*_n \leq u_n\} - \exp\{-e^{-x}\} \sim -(r-1)^2 e^{-x}\exp\{-e^{-x}\}\frac{\ln\ln n}{\ln n}.$$

Teorema 3.1.5 *Neka je $M^*_{n,1}, ..., M^*_{n,n}, ..$ niz nezavisnih slučajnih veličina sa zajedničkom funkcijom raspodele:*

$$F(x) = P\{M^*_n \leq x\}$$

i

$$S^*_n = \max\{M^*_{n,1}, ..., M^*_{n,n}\}.$$

Ako je $b_n = 2\ln n + (r-1)\ln(2\ln n) - \ln((r-1)!)$*, tada za svaki realan broj* x *važe jednakosti:*
1.
$$\lim_{n\to\infty} P\{S_n^* \le n(x+b_n)\} = \exp\{-e^{-x}\}$$
2.
$$\lim_{n\to\infty} P\{M_n^* \le n(x+b_n) | M_n^* > nb_n\} = 1 - e^{-x}$$

Dokaz: U radu Mladenović (2002) odredjeno je asimptotsko ponašanje repa date negativne binomne raspodele:

$$P\{X_{n,j}^* > k\} = \frac{k^{r-1}}{n^{r-1}(r-1)!}\left(1-\frac{1}{n}\right)^{k-r+1}(1+R(k,n)),$$

pri čemu je ispunjeno $R(k,n) \to 0$ kad $n \to \infty$ i $\frac{k}{n} \to \infty$ za $n \to \infty$.

Na osnovu toga se dobija:
1.
$$P\{S_n^* \le n(x+b_n)\} = (P\{M_n^* \le n(x+b_n)\})^n$$
$$= \left(\left(1 - \frac{(n(x+b_n))^{r-1}}{n^{r-1}(r-1)!}\left(1-\frac{1}{n}\right)^{n(x+b_n)-r+1}(1+R(x,n))\right)^n\right)^n$$
$$= \left(1 - \frac{(x+b_n)^{r-1}}{(r-1)!}\left(1-\frac{1}{n}\right)^{n(x+b_n)-r+1}(1+R(x,n))\right)^{n^2}$$
$$\sim \left(1 - \frac{(x+b_n)^{r-1}}{(r-1)!}e^{-(x+b_n)+\frac{r-1}{n}}(1+R(x,n))\right)^{n^2}$$
$$\sim \left(1 - \frac{(x+b_n))^{r-1}}{(r-1)!}e^{-x}\frac{1}{n^2}\frac{(r-1)!}{(2\ln n)^(r-1)}(1+R(x,n))\right)^{n^2}$$
$$\sim \left(1 - \frac{(x+b_n))^{r-1}}{(2\ln n)^{r-1}}e^{-x}\frac{1}{n^2}(1+R(x,n))\right)^{n^2}$$
$$\sim \left(1 - \left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} e^{-x}\frac{1}{n^2}(1+R(x,n))\right)^{n^2}$$
$$\sim \left(1 - \left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} \frac{e^{-x}}{n}(1+R(x,n))\right)^{n}$$
$$\sim \left(\left(1 - \frac{e^{-x}}{n}\right)^{-\frac{n}{e^{-x}}}\right)^{-e^{-x}} \sim \exp\{-e^{-x}\}, \qquad n \to \infty$$

pri čemu važi $R(x,n) \to 0$ za $n \to \infty$.

2.

$$P\{M_n^* \le n(x+b_n) | M_n^* > nb_n\} = \frac{P\{M_n^* \le n(x+b_n), M_n^* > nb_n\}}{P\{M_n^* > nb_n\}}$$

$$= \frac{P\{M_n^* \le n(x+b_n)\} - P\{M_n^* \le nb_n\}}{P\{M_n^* > nb_n\}} = 1 - \frac{1 - P\{M_n^* \le n(x+b_n)\}}{1 - P\{M_n^* \le nb_n\}}$$

$$= 1 - \frac{1 - \left(1 - \left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} e^{-x}\frac{1}{n^2}(1+R(x,n))\right)^n}{1 - \left(1 - \left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} \frac{1}{n^2}(1+R(x,n))\right)^n}$$

$$\sim 1 - \frac{1 - 1 + n\left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} e^{-x}\frac{1}{n^2}(1+R(x,n))}{1 - 1 + n\left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} \frac{1}{n^2}(1+R(x,n))}$$

$$\sim 1 - \frac{\left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} e^{-x}\frac{1}{n}(1+R(x,n))}{\left(1 + O\left(\frac{\ln(2\ln n)}{2\ln n}\right)\right)^{r-1} \frac{1}{n}(1+R(x,n))}$$

$$\sim 1 - \frac{e^{-x}\frac{1}{n}}{\frac{1}{n}} \sim 1 - e^{-x}, \qquad n \to \infty.$$

Vreme čekanja u nizu eksperimenata

Neka je $Z_1, Z_2, ...$ niz nezavisnih slučajnih veličina sa zajedničkom ravnomernom raspodelom verovatnoća na skupu $\mathbf{N}_n = \{1, 2, ..., n\}$. Neka je X_{nj} vreme čekanja do prve pojave dogadjaja $\{Z_k = j\}$ u nizu $Z_1, Z_2, ...$ i neka je:

$$M_n = \max\{X_{n1}, X_{n2}, ..., X_{nn}\}.$$

Tada M_n predstavlja vreme čekanja do pojave svih dogadjaja $\{Z_k = j\}$, $j \in \mathbf{N}_n$ u nizu $Z_1, Z_2, ...$

Teorema 3.1.6 1.**(Erdos, Rényi (1961))** *Za svaki realan broj x važi jednakost:*

$$\lim_{n\to\infty} P\{M_n \le n(x + \ln n)\} = \exp\{-e^{-x}\}$$

2.**(Mladenović (1999))** *Za svaki realan broj x pri $n \to \infty$ važi:*

$$|P\{M_n \le n(x + \ln n)\} - \exp\{-e^{-x}\}| = O\left(\frac{\ln n}{n}\right)$$

Teorema 3.1.7 *Za svaki realan broj x važi jednakost:*

$$\lim_{n\to\infty} P\{M_n \leq n(x+2\ln n) | M_n > 2n\ln n\} = 1 - e^{-x}$$

Dokaz: Neka je $u_n = n(x+2\ln n)$, $r_n = n(x+2\ln n) - [n(x+2\ln n)]$.
U radu Mladenović (1999) odredjena je funkcija raspodele slučajne veličine M_n:

$$P\{M_n \leq k\} = \sum_{m=0}^{n} (-1)^m \binom{n}{m} \left(1 - \frac{m}{n}\right)^k, \quad k \geq n.$$

Na osnovu toga se dobija:

$$P\{M_n \leq u_n\} = \sum_{m=0}^{n} (-1)^m \binom{n}{m} \left(1 - \frac{m}{n}\right)^{[u_n]}$$

$$= \sum_{m=0}^{n} (-1)^m \binom{n}{m} e^{(u_n - r_n)\left(-\frac{m}{n} - \frac{m^2}{2n^2} + O\left(\frac{1}{n^2}\right)\right)}$$

$$\sim \sum_{m=0}^{n} (-1)^m \binom{n}{m} e^{-\frac{m}{n}(u_n - r_n) - \frac{m^2}{2n^2}(u_n - r_n)}, \qquad n \to \infty$$

$$= \sum_{m=0}^{n} (-1)^m \binom{n}{m} e^{-m(x+2\ln n) + O\left(\frac{\ln n}{n}\right)} \sim \sum_{m=0}^{n} (-1)^m \binom{n}{m} \frac{e^{-mx}}{n^{2m}}$$

$$\sim 1 - n\frac{e^{-x}}{n^2} + \frac{n(n-1)}{2}\frac{e^{-2x}}{n^4} + \frac{n(n-1)(n-2)}{6}\frac{e^{-3x}}{n^6} + \ldots$$

$$\sim 1 - \frac{e^{-x}}{n} + O\left(\frac{1}{n^2}\right), \qquad n \to \infty.$$

$$P\left\{M_n \leq [n(x+2\ln n)] | M_n > [2n\ln n]\right\}$$

$$= \frac{P\left\{M_n \leq [n(x+2\ln n)], M_n > [2n\ln n]\right\}}{P\left\{M_n > [2n\ln n]\right\}}$$

$$= \frac{P\left\{M_n \leq [n(x+2\ln n])\right\} - P\left\{M_n \leq [2n\ln n]\right\}}{P\left\{M_n > [2n\ln n]\right\}}$$

$$= \frac{P\left\{M_n \leq [n(x+2\ln n)]\right\} - 1 + 1 - P\left\{M_n \leq [2n\ln n]\right\}}{P\left\{M_n > [2n\ln n]\right\}}$$

$$= 1 - \frac{1 - P\{M_n \leq [n(x+2\ln n)]\}}{1 - P\{M_n \leq [2n\ln n]\}} \sim 1 - \frac{1 - \left(1 - \frac{e^{-x}}{n} + O\left(\frac{1}{n^2}\right)\right)}{1 - \left(1 - \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right)}$$

$$\sim 1 - e^{-x}, \qquad n \to \infty.$$

Neka je $Z_1, Z_2, ...$ niz nezavisnih slučajnih veličina sa zajedničkom ravnomernom raspodelom verovatnoća na skupu $\mathbf{N}_n = \{1, 2, ..., n\}$. Za dati prirodan broj $r \geq 2$, neka je sa X_{nj} označeno vreme čekanja dok se svaki od dogadjaja oblika $\{Z_k = j\}$ ne realizuje u nizu $Z_1, Z_2, ...$ tačno r puta. Ako je:

$$M_n = \max\{X_{n1}, X_{n2}, ..., X_{nn}\},$$

tada M_n predstavlja vreme čekanja do pojave svih dogadjaja oblika $\{Z_k = j\}$, $j \in \mathbf{N}_n$, u nizu $Z_1, Z_2, ...$ najmanje r puta.
Neka je

$$u_n = n(x + \ln n + (r-1)\ln\ln n - \ln(r-1)!).$$

Teorema 3.1.8 (Mladenović (2002)) 1. *Granična raspodela slučajne veličine M_n data je sa:*

$$\lim_{n\to\infty} P\{M_n \leq u_n\} = \exp\{-e^{-x}\}$$

2. *Za $r \geq 2$, pri $n \to \infty$ važi:*

$$P\{M_n \leq u_n\} - \exp\{-e^{-x}\} \sim -(r-1)^2 e^{-x} \exp\{-e^{-x}\} \frac{\ln\ln n}{\ln n}.$$

3.2 GPD raspodele i ocenjivanje VaR parametra

Neka je $(X_1, ..., X_N)$ uzorak čiji članovi predstavljaju vrednosti neke slučajne veličine X sa nepoznatom funkcijom raspodele, F. Da bi se funkcija raspodele prekoračenja praga u, $F^{[u]}$ aproksimirala jednom od generalisanih Paretovih raspodela, potrebno je odrediti vrednost praga u, pa na osnovu te vrednosti oceniti nepoznate parametre oblika, položaja i razmere za odgovarajuću raspodelu (odrediti koja raspodela je u pitanju).

Neki načini odabira praga

- *Korišćenje uzoračkog srednjeg prekoračenja.*
  Uzoračko srednje prekoračenje je:

  $$e_N(u) = \frac{\sum\limits_{i \leq N} (X_i - u) I(u < X_i)}{\sum\limits_{i \leq N} I(u < X_i)}$$

  i to je jedna empirijska ocena srednjeg prekoračenja za funkciju raspodele F.
  Poznato je da je za sve generalisane Paretove raspodele srednje prekoračenje linearna funkcija, pa se zato bira se ona vrednost praga u, za koju važi da je uzoračko srednje prekoračenje približno linearna funkcija, počevši od te vrednosti u.
  Ako se izabere preterano nizak prag, onda nisu ispunjeni svi uslovi teoreme 2.3.1, tj. aproksimacija ovim tipom raspodela nije odgovajuća i dobiće se pristrasne ocene nepoznatih parametara. S druge strane, ako je odabrani prag previše visok, iznad praga ostaje previše malo članova uzorka i dobijene ocene će imati veliku disperziju. Uopšte, povećanjem praga povećava se disperzija, a smanjuje pristrasnost ocene, i obrnuto.

- *Automatski odabir praga.*
  Neka su $X_{n-k+1,n}, ..., X_{n,n}$ k najvećih vrednosti u uzorku. Ako je $M_{k,n}$ medijana za deo uzorka $(X_{n-k+1,n}, ..., X_{n,n})$, onda za broj prekoračenja treba izabrati ono k koje minimizira izraz:

  $$\frac{1}{k} \sum_{i \leq k} i^{\beta} |X_{i,n} - M_{k,n}|$$

  za $\beta \in [0, \frac{1}{2}]$. Ako je poznat broj prekoračenja, time je odredjen i prag.

Neke ocene parametra oblika za generalisane Paretove raspodele

Neka je $(X_1, ..., X_N)$ uzorak čiji članovi predstavljaju vrednosti neke slučajne veličine sa funkcijom raspodele F, koja pripada jednoj od POT-oblasti privlačenja generalisanih Paretovih raspodela. Iz teorema 2.4.1 i 2.4.2 sledi da funkcija F pripada i oblasti privlačenja za maksimume one raspodele ekstremnih vrednosti koja ima isti parametar oblika. Ako su raspodele date u γ-parametrizaciji, odgovarajući parametar oblika se može oceniti na sledeće načine:

- *Pikandsova ocena*
  Neka su $X_{N,1}, ..., X_{N,N}$ članovi datog uzorka poredjani po veličini, od najmanjeg do najvećeg. Neka je $m(N)$ niz prirodnih brojeva za koji važi: $m \to \infty, \quad \frac{m}{N} \to \infty, \quad$ za $N \to \infty$.
  Pod tim uslovima, Pikandsova ocena parametra γ je data sa:

$$\gamma'_N = \frac{1}{\ln 2} \ln \frac{X_{N,N-m+1} - X_{N,N-2m+1}}{X_{N,N-2m+1} - X_{N,N-4m+1}}$$

- *Hilova ocena*
  Pod istim uslovima kao za Pikandsovu ocenu, i pod pretpostavkom da je $\gamma > 0$, Hilova ocena parametra γ je data sa:

$$H_N = \frac{1}{m} \sum_{i=1}^{m} \ln X_{N,N-i+1} - \ln X_{N,N-m}$$

- *Ocena metodom momenata*
  Ocena metodom momenata je data sa:

$$\gamma' = \gamma_1 + \gamma_2$$

  gde je γ_1 Hilova ocena,

$$\gamma_2 = 1 - \frac{1}{2}(1 - \frac{l_1^2}{l_2})^{-1}$$

$$l_j = \frac{1}{m} \sum_{i=1}^{m} (\ln X_{N,N-i+1} - \ln X_{N,N-m})^i$$

  za $j = 1, 2$.
  Ako je $\gamma \geq 0$, γ_1 služi kao ocena parametra γ, a γ_2 je blisko nuli, a ako je $\gamma \leq 0$, važi obrnuto.

Kod prethodnih ocena, pretpostavlja se da su ekstremumi odredjeni korišćenjem statistika poretka, tj. vrednost $X_{N,N-m}$ predstavlja slučajni prag. Ako je prag, u odredjen na neki drugi način, onda se vrednosti $X_{N,1},...,X_{N,N}$ zamenjuju članovima datog uzorka većim od u, vrednost $X_{N,N-m}$ se zamenjuje sa u, a sumiranje se vrši po broju prekoračenja.

- *Korišćenje uzoračkog srednjeg prekoračenja*
  Ako je odgovarajući prag, u, odabran korišćenjem uzoračkog srednjeg prekoračenja, onda je ta funkcija približno linearna, počevši od te vrednosti u. Neka je β koeficijent pravca prave koja najbolje aproksimira tu funkciju, odredjen po metodi najmanjih kvadrata. Tada je β blisko vrednosti $\frac{\gamma}{1-\gamma}$, pa je $\gamma' = \frac{\beta}{1+\beta}$ jedna moguća ocena parametra γ.

- *Korišćenje metoda blok maksimuma*
  Neka je $(Y_1, Y_2, ..., Y_n)$ uzorak ekstremnih vrednosti dobijen metodom blok maksimuma iz početnog uzorka. Ako je broj blokova, n, dovoljno veliki, raspodela maksimuma, F^n se može aproksimirati jednom od raspodela ekstremnih vrednosti. Kad se odredi koja raspodela je u pitanju, tada se i odgovarajuća raspodela prekoračenja može aproksimirati onom generalisanom Paretovom raspodelom koja ima isti parametar oblika kao ta raspodela ekstremnih vrednosti.

Neki načini ocenjivanja parametara položaja i razmere za generalisane Paretove raspodele

Neka su σ i μ parametri razmere i položaja za odabranu raspodelu. Problem njihovog ocenjivanja se svodi na ocenjivanje parametra:

$$\theta = (\sigma, \mu) \in (R \times R+).$$

Takodje, mogu se istovremeno oceniti sva tri parametra, kao:

$$\theta = (\gamma, \sigma, \mu) \in (R \times R \times R+).$$

Neki od metoda su:

- *Metod maksimalne verodostojnosti*
  Neka je H_θ raspodela čije parametere treba oceniti, h_θ odgovarajuća

gustina raspodele, $(X_1, ..., X_N)$ uzorak na osnovu koga se vrši ocenjivanje. Funkcija verodostojnosti bazirana na datom uzorku je:

$$L(\theta; X) = \prod_{i=1}^{N} h_\theta(X_i) I_{\theta \in D(\theta)},$$

gde je $D(\theta)$ skup dozvoljenih vrednosti parametra. Treba odrediti θ koje maksimizira funkciju verodostojnosti, odnosno, zbog jednostavnosti, njen logaritam. Ocene parametra se onda dobijaju kao rešenja sistema jednačina:

$$\frac{\partial(\ln L(\theta; X))}{\partial \sigma} = 0,$$

$$\frac{\partial(\ln L(\theta; X))}{\partial \mu} = 0,$$

a, ako se ocenjuje i parametar oblika, postoji i treća jednačina.
U opštem slučaju, rešenja ovog sistema se ne dobijaju u eksplicitnom obliku, već primenom numeričkih metoda (npr. Newton-Raphson metod, kad se podje od nekog početnog rešenja).

- *Metod ponderisanih momenata*
  Neka je H_θ raspodela čije parametre treba oceniti i $X_1, ..., X_N$ uzorak na osnovu koga se vrši ocenjivanje. Neka je:

  $$m_r = E(X H_\theta^r(X))$$

  tj.

  $$m_r = \int_{-\infty}^{+\infty} x H_\theta^r(x) dF(x),$$

  Kada se u poslednjoj formuli nepoznata funkcija raspodele F zameni empirijskom funkcijom raspodele $F_n(x)$, dobija se:

  $$m'_r = \int_{-\infty}^{+\infty} x H_\theta^r(x) dF_n(x),$$

  a te vrednosti se mogu izračunati na osnovu datog uzorka. Ocene parametara se dobijaju rešavanjem sistema jednačina:

  $$m_r = m'_r, \quad r = 0, 1, 2$$

  Ako je ocena parametra oblika već odredjena,ocene za σ i μ date su sa:

  $$\sigma' = \frac{(2m'_1 - m'_0)\gamma'}{\Gamma(1-\gamma')(2^{\gamma'} - 1)},$$

$$\mu' = m'_0 + \frac{\sigma'}{\gamma'}(1 - \Gamma(1-\gamma')),$$

gde je $\Gamma(x) = \int\limits_0^{+\infty} e^{-t} t^{x-1} dt$.

Ocenjivanje kvantila raspodele

Neka je $(X_1, ..., X_N)$ uzorak čiji članovi predstavljaju vrednosti slučajne veličine X sa nepoznatom funkcijom raspodele F. Neka je poznato da funkcija raspodele F pripada POT-oblasti privlačenja jedne od generalisanih Paretovih raspodela, $H_{\gamma,\mu,\sigma}$, čiji parametri su već ocenjeni sa γ', μ', σ'.
Iz jednakosti:

$$F^{(u)}(x) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x \geq 0$$

dobija se:

$$1 - F(x+u) = (1 - F^{(u)}(x))(1 - F(u)), \quad x \geq 0. \tag{15}$$

Neka je n broj članova uzorka većih od u. Tada se drugi član na desnoj strani jednakosti može oceniti sa $\frac{n}{N}$, prvi član na desnoj strani sa $1 - H_{\gamma',\mu',\sigma'}$, pa se zamenom u jednakost (15) dobija ocena repa nepoznate raspodele F:

$$(1 - F(x+u))' = \frac{n}{N}\left(1 + \gamma'\frac{x - \mu'}{\sigma'}\right)^{-\frac{1}{\gamma'}} \tag{16}$$

Ako je x_p p-kvantil raspodele F, onda važi:

$$1 - p = \frac{n}{N}\left(1 + \gamma'\frac{x_p - \mu'}{\sigma'}\right)^{-\frac{1}{\gamma'}},$$

i, sredjivanjem jednakosti, dobija se ocena kvantila x_p:

$$x_p = \mu' + \frac{\sigma'}{\gamma'}\left(\frac{N}{n}(1-p)^{-\gamma'} - 1\right). \tag{17}$$

Ocenjivanje VaR parametra

Bankama, osiguravajućim društvima i, uopste, svim učesnicima na finansijskim tržištima je potrebno da znaju približan stepen izloženosti riziku pri svojim transakcijama. Jedan od načina da se to utvrdi je ocenjivanje parametra vrednosti pri riziku, VaR. Parametar VaR predstavlja maksimalni gubitak kome se izlaže učesnik na finansijskom tržištu u nekom odredjenom vremenskom periodu, ako je verovatnoća da se pojavi takav gubitak manja od neke zadate vrednosti (obično 5% ili 1%).

Neka su I_0,I_1,...,I_t vrednosti nekog finansijskog instrumenta (na primer, cene akcija neke firme, devizni kurs...), date u diskretnim vremenskim trenucima $t = 0, 1, 2...$ Period je najčešće dan, nedelja ili mesec. Tada je prinos u periodu T (promena vrednosti tog finansijskog instrumenta u odnosu na početnu, u periodu T) jednak $\frac{I_T - I_0}{I_0}$.
Prinos u periodu $T = 1$ (dnevni, nedeljni, mesečni) je onda dat sa:

$$s_t = \frac{I_t - I_{t-1}}{I_{t-1}}$$

i naziva se *aritmetički prinos.*
Ista veličina se može izraziti u obliku:

$$r_t = \ln I_t - \ln I_{t-1}$$

što se naziva *logaritamski prinos.*
Ove dve veličine su bliske u slučaju kad je $\frac{I_t}{I_{t-1}}$ blisko 1, ali veličina r_t je pogodnija za statističko modeliranje zbog toga što je neograničena i zbog toga što neke ekonomske zakonitosti mogu pomoću nje da se prikažu u jednostavnijem obliku.
Ako je I_0 vrednost nekog finansijskog instrumenta u trenutku $t = 0$, i $r_1, r_2, ..., r_T$ prinosi tokom perioda T, onda je vrednost I_T u trenutku T jednaka:

$$I_T = I_0 \exp \sum_{t \leq T} r_t. \tag{18}$$

Gubitak kome je učesnik na tržistu izložen tokom perioda od T dana (nedelja, meseci) je:

$$G_T = \sum_{t \leq T} (-r_t).$$

Neka je β zadata verovatnoća (obično 95% ili 99%).
Tada je $VaR(T; \beta)$, tj. maksimalni gubitak u periodu T, za dati nivo poverenja β, odredjen sa:

$$F_T = P\{G_T \leq VaR(T; \beta)\} = \beta$$

Znači, $VaR(T;\beta)$ je β-kvantil funkcije raspodele gubitka, F_T.
Iz jednakosti (18) se dobija:

$$\begin{aligned} VaR(T;\beta)' &= -(I_T - I_0) = I_0(1 - \exp\sum_{t\le T} r_t) \\ &= I_0(1 - \exp\{-F_T^{-1}(\beta)\}) \approx I_0 F_T^{-1}(\beta). \end{aligned}$$

Ova ocena parametra $VaR(T;\beta)$ se može koristiti u slučaju kad je $F_T^{-1}(\beta)$ dovoljno malo. Dalje, treba naći ocenu veličine $F_T^{-1}(\beta)$, na osnovu datog uzorka I_0,I_1,...,I_T. Neki od načina da se to uradi su:

- *Korišćenje empirijske funkcije raspodele gubitaka*
  Na osnovu prethodnih vrednosti finansijskog instrumenta, odnosno prinosa i gubitaka, odredi se empirijska funkcija raspodele gubitaka i tada je njen β-kvantil ocena za $F_T^{-1}(\beta)$. Mane ovog metoda su što nije moguće oceniti ekstremne kvantile na osnovu uzorka i što je ova ocena osetljiva na pojavu ekstremnih podataka u uzorku.

- *Modeliranje GPD raspodelama*
  Funkcija raspodele gubitaka se modelira jednom od GPD raspodela, $W_{\gamma,\mu,\sigma}$. Tada se za ocenu β-kvantila raspodele $W_{\gamma,\mu,\sigma}$ može koristiti formula (17) i dobija se:

$$VaR(T;\beta) = \mu' + \frac{\sigma'}{\gamma'}\left(\frac{N}{n}(1-\beta)^{-\gamma'} - 1\right). \tag{19}$$

Primer-ocenjivanje VaR parametra

Kao primer za ocenu VaR parametra korišćeni su podaci o dnevnom kretanju cena akcija kompanije General Electric (daily closing price, January 1998-October 2008, preuzeto sa www.finance.yahoo.com). Ukupan broj podataka je 2640. Polazni podaci su logaritmovani i diferencirani, da bi se dobile dnevne stope prinosa, i pomnoženi sa 100. Vrednost visokog praga i odgovarajući skup ekstremnih podataka su odredjeni na dva načina.

1. Posmatrano je srednje prekoračenje funkcije raspodele dnevnih prinosa. Ta funkcija je približno linearna počevši od vrednosti $u = 5$, pa je ta vrednost uzeta za prag.
vrednost praga: 5, broj prekoračenja: 33

ocena	γ'	μ'	σ'	$VaR(1,0.95)'$	$VaR(1,0.99)'$
MLE	0.803	5.137	1.272	4.07%	5.45%
Moment	0.637	5.137	1.713	3.56%	5.55%
Dress-Pickands	1.244	5.137	1.318	4.27%	5.54%
Slope	0.876	5.137	1.552	3.89%	5.52%

2*a*. Pri ocenjivanju parametra $VaR(1,0.95)$, vrednost praga odredjena je kao 95% - kvantil funkcije raspodele gubitaka (funkcija raspodele dnevnih prinosa pomnožena sa -1).
vrednost praga: 3, broj prekoračenja: 124

ocena	γ'	μ'	σ'	$VaR(1,0.95)'$
MLE	0.401	3.012	1.212	2.94%
Moment	0.472	3.012	1.215	2.94%
Dress-Pickands	0.402	3.012	1.252	2.94%
Slope	0.816	3.012	1.044	2.95%

2*b*. Pri ocenjivanju parametra $VaR(1,0.99)$, vrednost praga odredjena je kao 99% - kvantil funkcije raspodele gubitaka (funkcija raspodele dnevnih prinosa pomnožena sa -1).
vrednost praga: 4, broj prekoračenja: 63

ocena	γ'	μ'	σ'	$VaR(1,0.99)'$
MLE	0.586	4	1.206	5.36%
Moment	0.57	4	1.2	5.35%
Dress-Pickands	0.559	4	1.206	5.35%
Slope	0.846	4	1.062	5.37%

Za obradu podataka je korišćen programski paket XTREMES (Reiss, Thomas, 2001.).

Literatura

[1] **Balkema, A. and L. de Haan** *Residual life time at great age*, Annals of Probability, 2,(1974) 792-804

[2] **Beirlant, J., J.P.Raoult, R.Worms** *On the Relative Approximation Error of the Generalized Pareto Approximation for a High Quantile*, Extremes 6, (2003) 335-360

[3] **Diebolt, J., M.-A. El-Aroui, M. Garrido, S. Girard** *Quasi-Conjugate Bayes Estimates for GPD Parameters and Application to Heavy Tails Modelling*, Extremes 8, (2005) 57-78

[4] **Embrechts, P., Kluppelberg, C. and T. Mikosch** *Modelling Extremal Events*, Springer-Verlag, Berlin, 4th ed., (2003)

[5] **Juárez, S., W. Schucany** *Robust and Efficient Estimation for the Generalized Pareto Distribution*, Extremes 7 (2004), 237-251

[6] **Mladenović, P.** *Limit theorems for the maximum terms of a sequence of random variables with marginal geometric distributions*, Extremes, 2 (1999), 405-419

[7] **Mladenović, P.** *A generalization of the Mejzler-de Haan theorem*, Theory Probab. Appl., 50 (2002) 141-153

[8] **Mladenović, P.** *Ekstremne vrednosti slučajnih nizova*, Matematički fakultet, Beograd, (2002)

[9] **Mladenović, Z. i P. Mladenović** *Praktični problemi ocene rizika u analizi dnevnih vremenskih serija*, SYM-OP-IS 2007, 117-120

[10] **Peng, L., A.H. Welsh** *Robust Estimation of the Generalized Pareto Distribution*, Extremes 4:1 (2001), 53-65

[11] **Raoult, J.-P., R. Worms** *Rate of Convergence for the Generalized Pareto Approximation of the Excesses* Adv. Appl. Prob. 35 (2003), 1007-1027

[12] **Reiss, R.-D., M. Thomas** *Statistical Analysis of Extreme Values*, Birkhauser Verlag, (1997)