Sveučilište J.J. Strossmayera u
Osijeku
Odjel za matematiku

Martina Peranić

Distribucije ekstremnih vrijednosti i primjene

Diplomski rad

Osijek, 2021.

Sveučilište J.J. Strossmayera u
Osijeku
Odjel za matematiku

Martina Peranić

Distribucije ekstremnih vrijednosti i primjene

Diplomski rad

Mentor: doc. dr. sc. Danijel Grahovac

Osijek, 2021.

Sadržaj

1 Uvod

U klasičnoj analizi podataka ekstremni događaji su često pogrešno identificirani kao stršeće vrijednosti. Ukoliko nas zanimaju samo svakodnevni događaji onda je to u redu, međutim često će nas zanimati upravo takvi rubni događaji kao što su poplave, veliki pad cijena dionica, svjetski rekordi u sportu i slično.
Ekstremni događaji se pojavljuju s relativno malom vjerojatnošću ali imaju značajan učinak na ponašanje cijele distribucije tj. modela.
Teorija ekstremnih vrijednosti bavi se stohastičkim ponašanjem minimuma i maksimuma nezavisnih i jednako distribuiranih slučajnih varijabli. Odgovarajući rezultati za minimalne vrijednosti mogu se lako dobiti iz onih za maksimalne korištenjem identiteta

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n),$$

stoga ćemo u nastavku promatrati samo maksimalne vrijednosti.
Tijekom studiranja upoznali smo se sa standardnom graničnom teorijom koja proučava ponašanje parcijalnih suma niza nezavisnih i jednako distribuiranih slučajnih varijabli.
Proučavanjem maksimuma takvog niza uočavamo sličnosti. Kod parcijalnih suma se kao granične distribucije, uz prikladnu normalizaciju i centriranje, pojavljuju stabilne distribucije, dok se kod maksimuma može pojaviti jedna od tri distribucije ekstremnih vrijednosti: Fréchetova, Weibullova ili Gumbelova distribucija.
U prvoj cjelini ćemo uvesti distribuciju parcijalnih maksimuma te opisati granične rezultate.
U drugoj cjelini bavit ćemo se maksimalnom domenom atrakcije, odnosno odrediti ćemo uvjete pod kojima niz parcijalnih maksimuma konvergira nekoj od distribucija ekstremnih vrijednosti. Spomenute tri distribucije ekstremnih vrijednosti mogu se objediniti u jedan zapis o čemu ćemo govoriti u sljedećoj cjelini. U drugom dijelu rada dotaknuti ćemo se problema procjene parametara ovih distribucija. Na kraju ćemo primijeniti opisanu teoriju na prikupljene podatke o osobnim rekordima atletičarki i atletičara u trčanju na 100 metara. Podatke ćemo obrađivati u statističkom programu R.

2 Teorija ekstremnih vrijednosti

Kao što smo već u uvodu spomenuli, baviti ćemo se samo maksimalnim vrijednostima. Neka je $X_1, X_2, \dots$ niz nezavisnih i jednako distribuiranih (n.j.d.) slučajnih varijabli sa zajedničkom funkcijom distribucije F. Označimo sa

$$M_1 = X_1, \quad M_n = \max(X_1, \cdots, X_n)$$

parcijalni maksimum. Funkciju distribucije maksimuma M_n zapisujemo kao

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{t=1}^{n} P(X_t \leq x) = F^n(x), \quad x \in \mathbb{R}, n \in \mathbb{N}.$$

Kako se ekstremi događaju u blizini gornjeg kraja distribucije, intuitivno asimptotsko ponašanje od M_n mora biti povezano s funkcijom distribucije F u desnom repu blizu desne krajnje točke. Neka je x_F krajnja desna točka funkcije distribucije F, tj. $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Tada imamo da za sve $x < x_F$,

$$P(M_n \leq x) = F^n(x) \to 0, \quad n \to \infty,$$

i u slučaju kada je $x_F < \infty$, za $x \geq x_F$

$$P(M_n \leq x) = F^n(x) = 1,$$

odnosno možemo pisati

$$F^n(x) \to \begin{cases} 0, & x < x_F \\ 1, & x \geq x_F. \end{cases}$$

Uočimo da smo kao graničnu distribuciju dobili degeneriranu funkciju distribucije. Kako bi dobili nedegeneriranu funkciju, potrebna je normalizacija, odnosno promatramo sljedeće:

$$P\Big(\frac{M_n - d_n}{c_n} \leq x\Big) \to H(x), \quad n \to \infty, \tag{1}$$

za neki niz konstanti $c_n > 0$ i $d_n \in \mathbb{R}$, pri čemu je $H(x)$ nedegenerirana funkcija distribucije.

Možemo pisati

$$P(M_n \leq u_n),$$

pri čemu je $u_n = c_n x + d_n$.

Prirodno nam se nameću dva pitanja:

(i) Koje su moguće nedegenerirane distribucije H koje se mogu pojaviti kao limes u (1)?

(ii) Kakve moraju biti distribucije F, za koje postoje nizovi c_n i d_n, takvi da (1) vrijedi?

U nastavku ćemo se baviti sa (i), dok ćemo odgovor na pitanje (ii) dati u poglavlju 3.

2.1 Granični rezultati

Granične rezultate maksimuma distribucija razmatramo analogno kao centralni granični teorem. Dok za prikladno centrirane i normalizirane parcijalne sume n.j.d. niza slučajnih varijabli imamo da su granične distribucije stabilne distribucije, kod parcijalnih maksimuma su to max-stabilne distribucije.

Definicija 1 (Max-stabilne distribucije). *Nedegenerirana slučajna varijabla X naziva se max-stabilna ako zadovoljava*

$$\max(X_1, \dots, X_n) =^d c_n X + d_n \tag{2}$$

za niz $X_1, \dots, X_n$ i pripadajuće konstante $c_n > 0, d_n \in \mathbb{R}$ i sve $n \geq 2$.

Ako je (X_n) n.j.d. niz max-stabilnih slučajnih varijabli, tada (2) možemo pisati kao

$$\frac{M_n - d_n}{c_n} =^d X. \tag{3}$$

Vrijedi: svaka max-stabilna distribucija je granična distribucija za maksimum nezavisnih i jednako distribuiranih slučajnih varijabli. Štoviše, max-stabilne distribucije su jedine moguće granične distribucije za normalizirani maksimum.

O tome nam svjedoči sljedeći teorem.

Teorem 1. *Klasa max-stabilnih distribucija podudara se sa klasom svih mogućih nedegeneriranih graničnih distribucija za prikladno normalnizirane maksimume nezavisnih i jednako distribuiranih slučajnih varijabli.*

Dokaz. Ostaje nam dokazati da je granična distribucija afine transformacije maksimuma max-stabilna distribucija. Pretpostavimo da je za odgovarajuće normirajuće konstante

$$\lim_{n\to\infty} F^n(c_n x + d_n) = H(x), \quad x \in \mathbb{R},$$

za neku nedegeneriranu funkciju distribucije H. Očekujemo da su moguće granične funkcije distribucije H neprekidne funkcije na cijelom $\mathbb{R}$.

Tada za svaki $k \in \mathbb{N}$

$$\lim_{n\to\infty} F^{nk}(c_n x + d_n) = \left(\lim_{n\to\infty} F^n(c_n x + d_n)\right)^k = H^k(x), \quad x \in \mathbb{R}.$$

Osim toga,

$$\lim_{n\to\infty} F^{nk}(c_{nk} x + d_{nk}) = H(x), \quad x \in \mathbb{R}.$$

Prema Teoremu A1.5[1] u [7] postoje konstante $\tilde{c}_k > 0$ i $\tilde{d}_k \in \mathbb{R}$ takve da

$$\lim_{n\to\infty} \frac{c_{nk}}{c_n} = \tilde{c}_k \quad \textbf{i} \quad \lim_{n\to\infty} \frac{d_{nk} - d_n}{c_n} = \tilde{d}_k,$$

i za nezavisne i jednako distribuirane slučajne varijable $Y_1, \dots, Y_k$ sa funkcijom distribucije H,

$$\max(Y_1, \dots, Y_k) =^d \tilde{c}_k Y_1 + \tilde{d}_k.$$

□

[1] Neka su $A, B, A_1, A_2, \dots$ slučajne varijable i $b_n > 0, \beta_n > 0$ i $a_n, \alpha_n \in \mathbb{R}$ konstante. Pretpostavimo da

$$b_n^{-1}(A_n - a_n) \xrightarrow{d} A.$$

Tada relacija

$$\beta_n^{-1}(A_n - \alpha_n) \xrightarrow{d} B \qquad \textbf{(A.1)}$$

vrijedi ako i samo ako

$$\lim_{n\to\infty} b_n/\beta_n = b \in [0, \infty\rangle, \quad \lim_{n\to\infty} (a_n - \alpha_n)/\beta_n = a \in \mathbb{R}.$$

Ako (A.1) vrijedi, tada je $B =^d bA + a$ i a, b su jedinstvene konstante za koje ovo vrijedi.
Kada (A.1) vrijedi, A je nedegenerirana ako i samo ako je $b > 0$, i tada su A i B istog tipa.

Sljedeći teorem predstavlja temeljni rezultat za teoriju ekstremnih vrijednosti te daje odgovor na pitanje (i).

Teorem 2 (Fisher-Tippett teorem)**.** *Neka je (X_n) niz nezavisnih jednako distribuiranih slučajnih varijabli. Ako postoje konstante $c_n > 0$, $d_n \in \mathbb{R}$ i neka nedegenerirana slučajna varijabla M s funkcijom distribucije $H(x)$ takva da vrijedi*

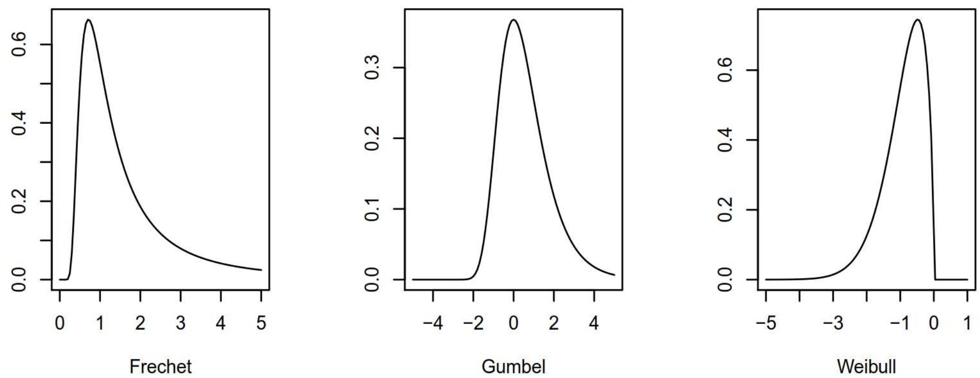
$$\frac{M_n - d_n}{c_n} \xrightarrow{d} M \Leftrightarrow \lim_{n\to\infty} F^n(c_n x + d_n) = H(x), \tag{4}$$

tada $H(x)$ pripada jednom od sljedeća tri tipa distribucija:

$$\textit{Fréchet}: \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ exp\{-x^{-\alpha}\}, & x > 0 \end{cases}, \quad \alpha > 0.$$

$$\textit{Weibull}: \Psi_\alpha(x) = \begin{cases} exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases}, \quad \alpha > 0.$$

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



Slika 1: Grafovi Fréchetove, Weibullove i Gumbelove funkcije gustoće

Definicija 2. *Funkcije distribucija $\Phi_\alpha(x), \Psi_\alpha(x)$ i $\Lambda(x)$ nazivaju se standardne distribucije ekstremnih vrijednosti. Realan broj α naziva se indeks ekstremnih vrijednosti.*

Prema Teoremu 1, distribucije ekstremnih vrijednosti su upravo max-stabilne distribucije. Ako je X slučajna varijabla s nekom od distribucija ekstremnih vrijednosti, onda X zadovoljava (3). Pri tome vrijedi:

$$\begin{aligned} \text{Fréchet} : & M_n =^d n^{1/\alpha} X \\ \text{Weibull} : & M_n =^d n^{-1/\alpha} X \\ \text{Gumbel} : & M_n =^d X + \ln n \end{aligned}$$

3 Maksimalna domena atrakcije i konstante normiranja

U ovom poglavlju nastojimo odrediti koji uvjeti na funkciju distribucije F vode do toga da normalizirani maksimum M_n konvergira po distribuciji ka danoj distribuciji ekstremnih vrijednosti, odnosno da vrijedi (4). Zanima nas i kako možemo izabrati normirajuće konstante $c_n > 0$ i $d_n \in \mathbb{R}$ tako da vrijedi (4).

Kod suma $S_n = X_1 + \ldots + X_n$ nezavisnih i jednako distribuiranih slučajnih varijabli javljaju se slični problemi. U okviru suma postavlja se pitanje: "Za danu α-stabilnu distribuciju G_α, koji uvjeti rezultiraju tome da normalizirana i centrirana suma S_n slabo konvergira ka G_α?" Za rješavanje tog problema koristi se sljedeći pristup: sve funkcije distribucije 'skupimo' u klasu za koju normalizirana suma ima istu graničnu stabilnu distribuciju. Takva se klasa naziva domena atrakcije.

Kao što smo u uvodu spomenuli, sličan pristup koristimo i za maksimum M_n.

Definicija 3 (Maksimalna domena atrakcije). *Kažemo da slučajna varijabla $X_t, t = 1, \ldots, n$ pripada maksimalnoj domeni atrakcije distribucije ekstremnih vrijednosti H ako postoje konstante $c_n > 0$, $d_n \in \mathbb{R}$ takve da vrijedi (4).*
Pišemo: $X_t \in MDA(H)$.

Sljedeća propozicija nam govori kada funkcija distribucije pripada maksimalnoj domeni atrakcije. No prije propozicije uvedimo definiciju repa distribucije.

Definicija 4 (Rep distribucije). *Za danu funkciju distribucije $F(x)$, kažemo da je $\overline{F}(x)$ repna funkcija distribucije (ili kraće "rep distribucije") od $F(x)$ ako vrijedi*

$$\overline{F}(x) = 1 - F(x) = P(X > x),$$

za velike vrijednosti x.

Propozicija 1 (Karakterizacija $MDA(H)$). *Funkcija distribucije F pripada maksimalnoj domeni atrakcije distribucije ekstremnih vrijednosti H s normirajućim konstantama $c_n > 0$, $d_n \in \mathbb{R}$ ako i samo ako*

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\ln H(x), \quad x \in \mathbb{R}.$$

Kada je $H(x) = 0$, limes se interpretira kao ∞.

Dokaz. Vidi u [3] dokaz teorema 1.1.2. str. 6. □

Kako bi odredili konstantu centriranja koristiti ćemo se funkcijom kvantila.

Definicija 5 (Funkcija kvantila). *Generalizirani inverz funkcije distribucije F*

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1,$$

zovemo funkcija kvantila funkcije distribucije F. Vrijednost $x_t = F^{\leftarrow}(t)$ definira t-kvantil od F. Ako je F strogo monotono rastuća i neprekidna, tada je $F^{\leftarrow}$ inverz od F.

Uzorački maksimum M_n je empirijska verzija $(1 - n^{-1})$-kvantila funkcije distribucije F. Zbog toga je ovaj kvantil prikladna centrirajuća konstanta.

Uvedimo još i pojam koji definira relaciju ekvivalencije na skupu svih funkcija distribucije.

Definicija 6 (Repna ekvivalencija). *Dvije funkcije distribucija F i G su repno ekvivalentne ako imaju istu desnu krajnju točku $x_F = x_G$ i*

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c, \quad \textit{za} \quad 0 < c < \infty.$$

U sljedećim potpoglavljima baviti ćemo se karakterizacijom maksimalnih domena atrakcije za Φ_α, Ψ_α i Λ. Prije toga uvodimo neke definicije potrebne za te karakterizacije.

Definicija 7 (Sporo varirajuća funkcija). *Pozitivna, Lebesgue izmjeriva funkcija L na $\langle 0, \infty \rangle$ je sporo varirajuća u ∞ ako*

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0. \tag{5}$$

Oznaka: $L \in R_0$.

Definicija 8 (Regularno varirajuća funkcija). *Pozitivna, Lebesgue izmjeriva funkcija h na $\langle 0, \infty \rangle$ je regularno varirajuća u ∞ s indeksom $\alpha \in \mathbb{R}$ ako*

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = t^\alpha, \quad t > 0. \tag{6}$$

Oznaka: $h \in R_\alpha$.

Definicija 9 (Rapidno varirajuća funkcija). *Pozitivna, Lebesgue izmjeriva funkcija h na $\langle 0, \infty \rangle$ je rapidno varirajuća s indeksom $-\infty$ ako*

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = \begin{cases} 0, & t > 1 \\ \infty, & 0 < t < 1. \end{cases}$$

Oznaka: $h \in \mathbb{R}_{-\infty}$.

Primjetimo da vrijedi sljedeće: svaka se regularno varirajuća funkcija h s indeksom α može zapisati kao $h(t) = t^\alpha L(t)$, pri čemu je L sporo varirajuća funkcija.

3.1 Maksimalna domena atrakcije Fréchetove distribucije

Promatramo distribucije oblika

$$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$$

za $\alpha > 0$.

Prema Taylorovom razvoju je

$$1 - \Phi_\alpha(x) = 1 - \exp\{-x^{-\alpha}\} \sim x^{-\alpha}, \quad x \to \infty,$$

pa zaključujemo da rep od Φ_α opada polinomijalno.

Za nizove kojima centriramo i normaliziramo maksimum možemo uzeti: $d_n = 0$ i c_n kao očekivana vrijednost funkcije kvantila, odnosno

$$\begin{aligned} c_n = F^{\leftarrow}(1 - \frac{1}{n}) &= \inf\left\{x \in \mathbb{R} : F(x) \geq 1 - \frac{1}{n}\right\} \\ &= \inf\left\{x \in \mathbb{R} : \left(\frac{1}{\overline{F}}\right)(x) \geq n\right\} \\ &= \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n) \end{aligned} \tag{7}$$

Teorem 3 (Maksimalna domena atrakcije od Φ_α). *Funkcija distribucije F pripada $MDA(\Phi_\alpha)$, $\alpha > 0$ ako i samo ako je $\overline{F}(x) = x^{-\alpha}L(x)$ za neku sporo varirajuću funkciju L.*
Ako je $F \in MDA(\Phi_\alpha)$, tada

$$\frac{M_n}{c_n} \xrightarrow{d} \Phi_\alpha,$$

gdje c_n možemo izabrati kao u (7).

Dokaz. Vidi u [7] dokaz teorema 3.3.7. str. 131. □

Slijedi da svaka $F \in MDA(\Phi_\alpha)$ ima beskonačnu krajnju desnu točku, tj. $x_F = \infty$. Štoviše, konstanta c_n formira regularno varirajući niz $c_n = n^{\frac{1}{\alpha}} L_1(n)$ za neku sporo varirajuću funkciju L_1. Dakle, zaključujemo da vrijedi

$$F \in MDA(\Phi_\alpha) \Leftrightarrow \overline{F} \in R_{-\alpha}$$

čime dobivamo jednostavnu karakterizaciju za $MDA(\Phi_\alpha)$.

Austrijski znanstvenik i matematičar Richard von Mises pronašao je uvjete na funkciju gustoće određene distribucije pod kojima ta funkcija pripada nekoj maksimalnoj domeni atrakcije.

Korolar 1 (Von Mises uvjet za $MDA(\Phi_\alpha)$). *Neka je F apsolutno neprekidna funkcija distribucije sa gustoćom f koja zadovoljava*

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \alpha > 0. \tag{8}$$

Tada je $F \in MDA(\Phi_\alpha)$.

Klasa funkcija distribucije F s regularno varirajućim repom $\overline{F}$ je zatvorena obzirom na repnu ekvivalenciju. Uvid u strukturu $MDA(\Phi_\alpha)$ daje nam sljedeća propozicija.

Propozicija 2 (Svojstvo zatvorenosti $(MDA\Phi_\alpha)$). *Neka su F i G funkcije distribucije i pretpostavimo da je $F \in MDA(\Phi_\alpha)$ s normirajućim konstantama c_n, takva da*

$$\lim_{n\to\infty} F^n(c_n x) = \Phi_\alpha(x), \quad x > 0.$$

Tada je

$$\lim_{n\to\infty} G^n(c_n x) = \Phi_\alpha(cx), \quad x > 0,$$

za neki $c > 0$ ako i samo ako su F i G repno ekvivalentne tako da je $\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c^\alpha$.

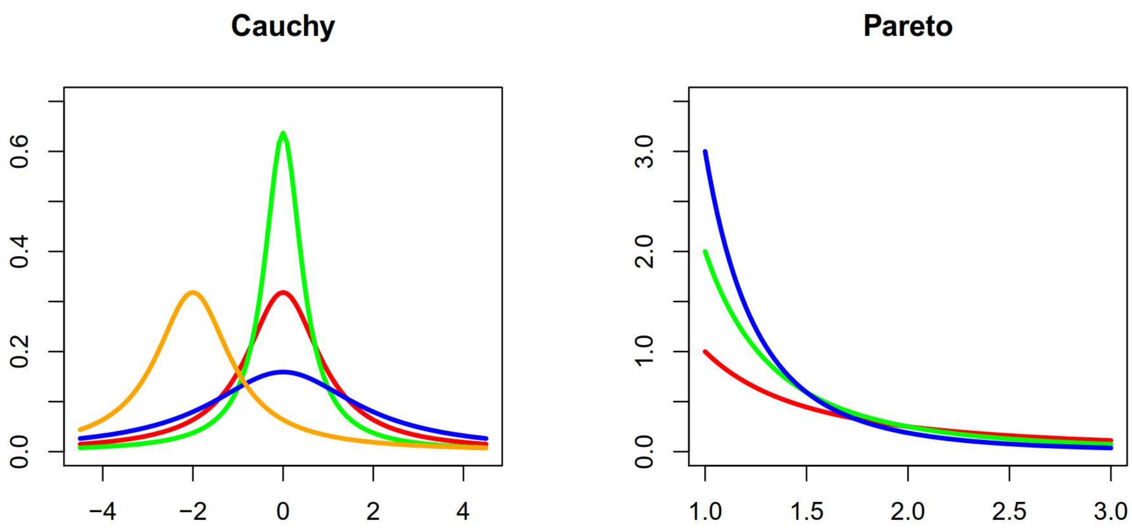
Dokaz. Vidi u [7] dokaz propozicije 3.3.9. str. 132. □

Vrijedi: $MDA(\Phi_\alpha)$ sastoji se od funkcija distribucija koje zadovoljavaju von Mises uvjet (8) i njenih repno ekvivalentnih funkcija distribucija.

Neke od distribucija koje pripadaju $MDA(\Phi_\alpha)$ su Cauchy, Pareto, i Loggama. Istaknimo dvije funkcije gustoće:

Cauchy: $f(x) = \frac{1}{\Pi(1+x^2)}, \quad x \in \mathbb{R}$

Pareto: $f(x) = \begin{cases} \alpha K^\alpha x^{-1-\alpha}, & x \in [K, \infty\rangle \\ 0, & x \in \langle -\infty, K\rangle \end{cases}$



Slika 2: Cauchyeva i Pareto funkcije gustoće

Ove distribucije imaju desni rep oblika

$$\overline{F}(x) \sim Kx^{-\alpha}, \quad x \to \infty$$

za neke $K, \alpha > 0$. Očito je da je rep distribucije $\overline{F}$ regularno varirajuća funkcija, odnosno $\overline{F} \in R_{-\alpha}$ iz čega slijedi $F \in MDA(\Phi_\alpha)$.

3.2 Maksimalna domena atrakcije Weibullove distribucije

Bavimo se distribucijama sljedećeg oblika

$$\Psi_\alpha(x) = exp\{-(-x)^\alpha\},$$

za $\alpha > 0$.

Vrijedi da sve funkcije distribucije $F \in MDA(\Psi_\alpha)$ imaju konačnu desnu krajnju točku x_F. Također vrijedi da su Ψ_α i Φ_α usko povezane, tj. imamo da je

$$\Psi_\alpha(-x^{-1}) = \Phi_\alpha(x), \quad x > 0$$

pa pretpostavljamo da su i $MDA(\Psi_\alpha)$ i $MDA(\Phi_\alpha)$ usko povezane. Sljedeći teorem to i potvrđuje.

Teorem 4 (Maksimalna domena atrakcije od Ψ_α)**.** *Funkcija distribucije F pripada $MDA(\Psi_\alpha)$, $\alpha > 0$ ako i samo ako $x_F < \infty$ i $\overline{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ za neku sporo varirajuću funkciju L. Ako je $F \in MDA(\Psi_\alpha)$, onda*

$$\frac{M_n - x_F}{c_n} \xrightarrow{d} \Psi_\alpha,$$

gdje c_n može biti izabrana kao $c_n = x_F - F^{\leftarrow}(1 - \frac{1}{n})$ i $d_n = x_F$.

Dokaz. Vidi u [7] dokaz teorema 3.3.12. str. 135. □

Zaključujemo da vrijedi:

$$F \in MDA(\Psi_\alpha) \Leftrightarrow x_F < \infty, \quad \overline{F}(x_F - x^{-1}) \in R_{-\alpha}.$$

Dakle, $MDA(\Psi_\alpha)$ se sastoji od funkcija distribucija koje su ograničene zdesna.

Slično kao u prethodnom potpoglavlju, imamo zgodnu karakterizaciju funkcije gustoće distribucije, tako da ona pripada $MDA(\Psi_\alpha)$.

Korolar 2 (Von Mises uvjet). *Neka je F apsolutno neprekidna funkcija distribucije s gustoćom f koja je pozitivna na nekom konačnom intervalu $\langle z, x_F\rangle$. Ako je*

$$\lim_{x \uparrow x_F} \frac{(x_F - x) f(x)}{\overline{F}(x)} = \alpha > 0, \tag{9}$$

tada je $F \in MDA(\Psi_\alpha)$.

Također, propoziciju 2. možemo preformulirati na sljedeći način.

Propozicija 3 (Svojstvo zatvorenosti od $MDA(\Psi_\alpha)$). *Neka su F i G funkcije distribucije s krajnjom desnom točkom $x_F = x_G < \infty$ i pretpostavimo da je $F \in MDA(\Psi_\alpha)$ s normirajućom konstantom $c_n > 0$ tako da je*

$$\lim_{n\to\infty} F^n(c_n x + x_F) = \Psi_\alpha(x), \quad x < 0.$$

Tada je

$$\lim_{n\to\infty} G^n(c_n x + x_F) = \Psi_\alpha(cx), \quad x < 0,$$

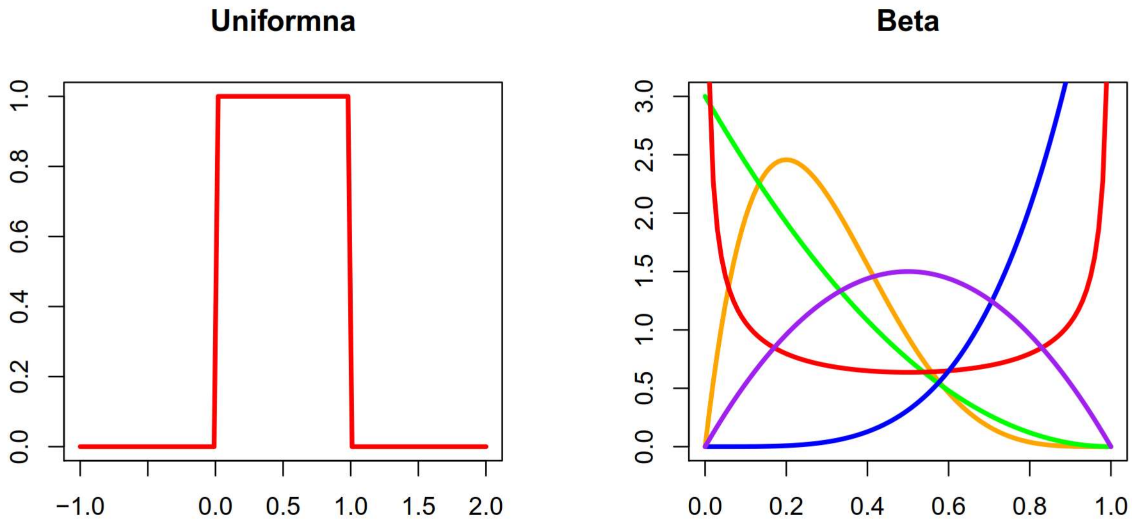
za neki $c > 0$ ako i samo ako su F i G repno ekvivalentne tako da je $\lim_{x\uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c^{-\alpha}$.

Vrijedi: $MDA(\Psi_\alpha)$ sastoji se od funkcija distribucije koje zadovoljavaju von Mises uvjet (9) i njenih repno ekvivalentnih funkcija distribucije.

Primjeri distribucija koji pripadaju $MDA(\Psi_\alpha)$ su:

Uniformna: $f(x) = 1, \quad x \in \langle 0, 1\rangle$

Beta: $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \quad a, b > 0$



Slika 3: Uniformna i Beta funkcije gustoće

3.3 Maksimalna domena atrakcije Gumbelove distribucije

Gumbelova distribucija ima oblik

$$\Lambda(x) = exp\{-exp\{-x\}\}.$$

Maksimalna domena atrakcije Gumbelove distribucije Λ pokriva širok raspon funkcija distribucije F. Iako nema direktne veze s regularno varirajućim funkcijama kao kod Fréchetove i Weibullove distribucije, naći ćemo proširenje regularne varijacije koja vrijedi za kompletnu karakterizaciju $MDA(\Lambda)$. Prema Taylorovom razvoju imamo da je

$$1 - \Lambda(x) \sim e^{-x}, \quad x \to \infty,$$

dakle $\overline{\Lambda}(x)$ pada u nulu eksponencijalno.

Vidjeti ćemo u nastavku kako klasa $MDA(\Lambda)$ sadrži funkcije distribucije s prilično različitim repovima, od onih umjereno teških pa do onih s lakim repovima kao na primjer kod normalne distribucije. Također, i oba slučaja $x_F < \infty$ i $x_F = \infty$ su moguća.

Za početak ćemo uvesti apsolutno neprekidnu distribuciju koja predstavlja važan dio maksimalne domene atrakcije.

Definicija 10 (Von Mises funkcija)**.** *Neka je F funkcija distribucije s krajnjom desnom točkom $x_F \leq \infty$. Pretpostavimo da postoji neki $z < x_F$ takav da F ima reprezentaciju*

$$\overline{F}(x) = c \cdot exp\{-\int_z^x \frac{1}{a(t)} dt\}, \quad z < x < x_F, \tag{10}$$

gdje je c neka pozitivna konstanta, $a(\cdot)$ je pozitivna i apsolutno neprekidna funkcija (Lebesgue izmjeriva) za čiju funkciju gustoće a' vrijedi $\lim_{x \uparrow x_F} a'(x) = 0$. Tada F nazivamo von Mises funkcija, a funkciju $a(\cdot)$ pomoćna funkcija od F.

Propozicija 4 (Svojstva von Mises funkcija)**.** *Svaka von Mises funkcija distribucije F je apsolutno neprekidna na $\langle z, x_F\rangle$, a funkcija gustoće f je pozitivna na tom intervalu. Pomoćna funkcija može biti izabrana kao*

$$a(x) = \frac{\overline{F}(x)}{f(x)}.$$

Vrijede sljedeća svojstva:

(a) Ako je $x_F = \infty$, tada je $\overline{F} \in R_{-\infty}$ i $\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \infty$

(b) Ako je $x_F < \infty$, tada je $\overline{F}(x_F - \frac{1}{x}) \in R_{-\infty}$ i $\lim_{x\uparrow x_F} \frac{(x_F-x)f(x)}{\overline{F}(x)} = \infty$.

Dokaz. Vidi u [7] dokaz propozicije 3.3.24. str 140. □

Sada možemo pokazati da von Mises funkcije pripadaju maksimalnoj domeni atrakcije Gumbelove distribucije. Štoviše, poseban oblik $\overline{F}$ omogućava da izračunamo normirajuće konstante c_n iz pomoćne funkcije.

Propozicija 5 (Von Mises funkcije i $MDA(\Lambda)$)**.** *Pretpostavimo da je F von Mises funkcija. Tada je $F \in MDA(\Lambda)$.*
Mogući odabir normirajućih konstanti je

$$d_n = F^{\leftarrow}\Big(1 - \frac{1}{n}\Big) \quad i \quad c_n = a(d_n),$$

gdje je a pomoćna funkcija od F.

Dokaz. Za $t \in \mathbb{R}$ i x dovoljno blizu x_F iz (10) slijedi

$$\frac{\overline{F}(x + ta(x))}{\overline{F}(x)} = exp\Big\{ - \int_x^{x+ta(x)} \frac{1}{a(u)} du\Big\}.$$

Neka je $v = \frac{u-x}{a(x)}$, tada dobivamo

$$\frac{\overline{F}(x + ta(x))}{\overline{F}(x)} = exp\Big\{ - \int_0^{t} \frac{a(x)}{a(x+va(x))} dv\Big\}. \tag{11}$$

Pokazati ćemo da podintegralna funkcija lokalno uniformno konvergira ka 1. Za dani $\varepsilon > 0$ i $x \geq x_0(\varepsilon)$

$$|a(x + va(x)) - a(x)| = \Big| \int_x^{x+va(x)} a'(s)ds \Big| \leq \varepsilon|v|a(x) \leq \varepsilon|t|a(x),$$

zato jer $a'(x) \to 0$ kad $x \uparrow x_F$. Slijedi za $x \geq x_0(\varepsilon)$ da je

$$\Big|\frac{a(x+va(x))}{a(x)} - 1\Big| \leq \varepsilon|t|.$$

Desna strana nejednakosti može biti proizvoljno mala, dakle

$$\lim_{x\uparrow x_F} \frac{a(x)}{a(x+va(x))} = 1$$

uniformno po v na ograničenim intervalima. Iz ovoga i iz (11) slijedi

$$\lim_{x\uparrow x_F} \frac{\overline{F}(x+ta(x))}{\overline{F}(x)} = e^{-t} \tag{12}$$

uniformno po t na ograničenim intervalima. Uzmimo sada da su normirajuće konstante

$$d_n = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n) \quad i \quad c_n = a(d_n).$$

Tada iz (12) slijedi

$$\lim_{n\to\infty} n\overline{F}(d_n + tc_n) = e^{-t} = -\ln\Lambda(t), \quad t\in\mathbb{R}.$$

Slijedi prema Propoziciji 1. da je $F \in MDA(\Lambda)$. $\square$

Von Mises funkcije ne karakteriziraju kompletno $MDA(\Lambda)$. Mala modifikacija relacije von Mises funkcije (10) vodi do potpune karakterizacije $MDA(\Lambda)$. Postoje dvije karakterizacije maksimalne domene atrakcije Gumbelove distribucije.

Teorem 5 (Karakterizacija I $MDA(\Lambda)$). *Funkcija distribucije F s krajnjom desnom točkom $x_F \leq \infty$ pripada $MDA(\Lambda)$ ako i samo ako postoji neki $z < x_F$ takav da F ima reprezentaciju*

$$\overline{F}(x) = c(x)\cdot exp\{-\int_z^x \frac{g(t)}{a(t)}dt\}, \quad z < x < x_F, \tag{13}$$

gdje su c i g izmjerive funkcije za koje vrijedi $c(x) \to c > 0$, $g(x) \to 1$ kad $x \uparrow x_F$, i $a(x)$ je pozitivna, apsolutno neprekidna (Lebesgue izmjeriva) funkcija s gustoćom $a'(x)$ koja ima $\lim_{x\uparrow x_F} a'(x) = 0$.
Za F sa reprezentacijom (13) možemo uzeti normirajuće konstante

$$d_n = F^{\leftarrow}\left(1-\frac{1}{n}\right), \quad i \quad c_n = a(d_n).$$

Mogući izbor za funkciju a je

$$a(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)}dt, \quad x < x_F.$$

Druga karakterizacija $MDA(\Lambda)$ proizlazi iz dokaza Propozicije 5. gdje smo pokazali da svaka von Mises funkcija zadovoljava (12), tj. postoji pozitivna funkcija $\widetilde{a}$ takva da

$$\lim_{x\uparrow x_F} \frac{\overline{F}(x+t\widetilde{a}(x))}{\overline{F}(x)} = e^{-t}, \quad t\in\mathbb{R}. \tag{14}$$

Teorem 6 (Karakterizacija II $MDA(\Lambda)$). *Funkcija distribucije F pripada $MDA(\Lambda)$ ako i samo ako postoji neka pozitivna funkcija $\widetilde{a}$ takav da vrijedi (14). Mogući izbor za $\widetilde{a} = a$ je analogan kao u Teoremu 5.*

Slično kao i kod maksimalne domene atrakcije Weibullove i Fréchetove distribucije, repna ekvivalencija je koristan alat i za određivanje pripadnosti neke distribucije maksimalnoj domeni atrakcije Gumbelove distribucije, te za računanje normirajućih konstanti. U slučaju Gumbelove distribucije to je čak i važniji alat nego u prethodnim slučajevima zbog velike raznolikosti repova $\overline{F}$.

Propozicija 6 (Svojstvo zatvorenosti $MDA(\Lambda)$ unutar repne ekvivalencije)**.** *Neka su F i G funkcije distribucije s istim krajnjim desnim točkama $x_F = x_G$ i pretpostavimo da je $F \in MDA(\Lambda)$ s normirajućim konstantama $c_n > 0$ i $d_n \in \mathbb{R}$ tako da je*

$$\lim_{n\to\infty} F^n(c_n x + d_n) = \Lambda(x), \quad x \in \mathbb{R}.$$

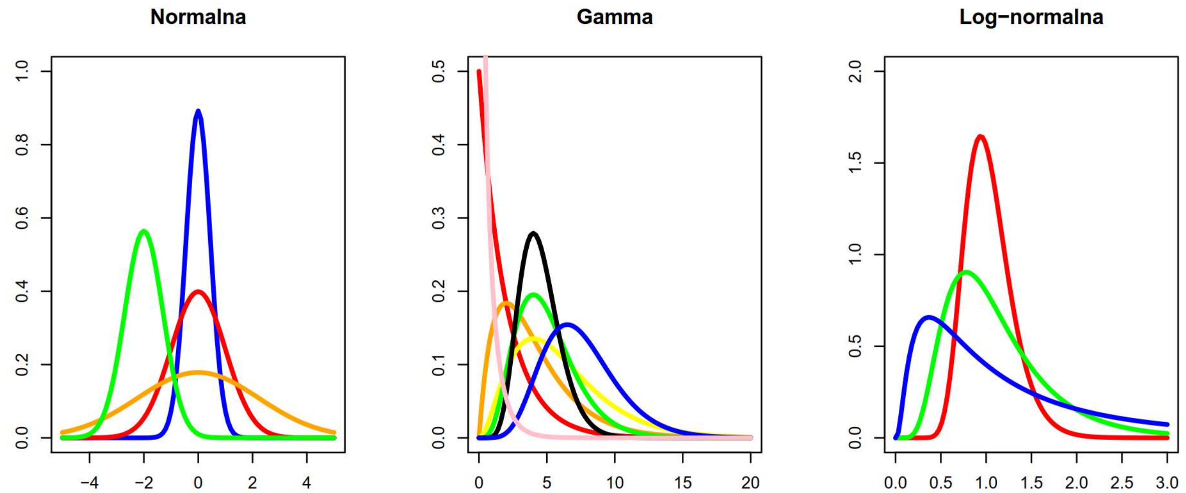
Tada

$$\lim_{n\to\infty} G^n(c_n x + d_n) = \Lambda(x + b), \quad x \in \mathbb{R}$$

ako i samo ako su F i G repno ekvivalentne s $\lim_{x\uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = e^b$.

Vrijedi: $MDA(\Lambda)$ se sastoji od von Mises funkcija i njihovih repno ekvivalentnih funkcija distribucija.

Normalna, gamma i lognormalna distribucija neki su od primjera distribucija koje pripadaju $MDA(\Lambda)$, a njihove funkcije gustoće su sljedeće:
Normalna: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}$
Gamma: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \alpha, \beta > 0$
Lognormalna: $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln x - \mu)^2/2\sigma^2}, \quad x > 0, \mu \in \mathbb{R}, \sigma > 0.$



Slika 4: Normalna, Gamma i Log-normalna funkcija gustoće

4 Generalizirana distribucija ekstremnih vrijednosti

Do sada smo proučavali svaku familiju distribucija ekstremnih vrijednosti $(\Phi_\alpha, \Lambda, \Psi_\alpha)$ zasebno, no sada ćemo pokazati da se ta tri oblika distribucija mogu objediniti u jedan zapis, generaliziranu distribuciju ekstremnih vrijednosti.

Standardne distribucije ekstremnih vrijednosti mogu biti reprezentirane uvođenjem parametra oblika ξ tako da

$$\begin{array}{lll} \xi = \alpha^{-1} > 0 & \text{odgovara Fréchetovoj distribuciji} & \Phi_\alpha, \\ \xi = 0 & \text{odgovara Gumbelovoj distribuciji} & \Lambda, \\ \xi = -\alpha^{-1} < 0 & \text{odgovara Weibullovoj distribuciji} & \Psi_\alpha. \end{array}$$

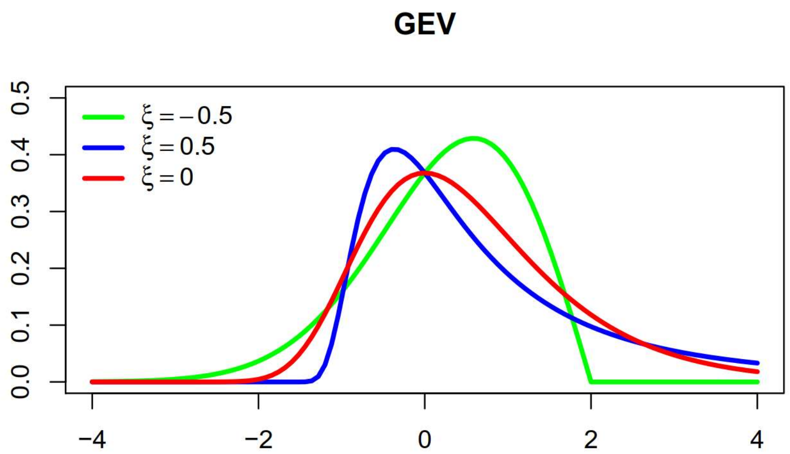
Jenkinson i von Mises uveli su generaliziranu reprezentaciju koja je do sada široko prihvaćena kao standardna reprezentacija:

Definicija 11 (Generalizirana distribucija ekstremnih vrijednosti (GEV))**.** *Definiramo funkciju distribucije H_ξ kao*

$$H_\xi(x) = \begin{cases} exp\{-(1+\xi x)^{-\frac{1}{\xi}}\}, & \xi \neq 0 \\ exp\{-exp\{x\}\}, & \xi = 0 \end{cases}$$

gdje je $1 + \xi x > 0$. H_ξ nazivamo standardna generalizirana distribucija ekstremnih vrijednosti.
Možemo uvesti srodnu familiju koja ima lokacijski parametar μ i parametar skaliranja ψ: $H_{\xi;\mu,\psi}$. Tu familiju dobivamo zamjenimo li gore argument x sa $\frac{x-\mu}{\psi}$ za $\mu \in \mathbb{R}$, $\psi > 0$.

$$H_{\xi;\mu,\psi}(x) = exp\left\{-\left(1+\xi\frac{x-\mu}{\psi}\right)^{-1/\xi}\right\}, \quad 1 + \xi(x-\mu)/\psi > 0.$$



Slika 5: Funkcija gustoće GEV distribucije u ovisnosti o parametru oblika ξ

Razmotrimo podklase odvojeno:

(1) Za $\xi > 0$ je $H_\xi(x) < 1$ za sve x, pa je desna krajnja točka funkcije distribucije beskonačna. Također vrijedi da takve distribucije imaju teške repove.

(2) Za $\xi = 0$ isto vrijedi da je desna krajnja točka beskonačna, dok distribucije imaju lakše repove.

(3) Za $\xi < 0$ desna krajnja točka funkcije distribucije jednaka je $-1/\xi$.

Detaljnije u [3] poglavlje 1.1.3.

Sljedeći teorem je jedan od temeljnih rezultata u teoriji ekstremnih vrijednosti. On objedinjuje osnovne rezultate prethodnog poglavlja.

Prije samog teorema definiramo funkciju

$$U(t) = F^{\leftarrow}(1 - t^{-1}), \quad t > 0.$$

Teorem 7 (Karakterizacija $MDA(H_\xi)$)**.** *Za $\xi \in \mathbb{R}$ sljedeće tvrdnje su ekvivalentne:*

(a) $F \in MDA(H_\xi)$.

(b) Postoji pozitivna, izmjeriva funkcija $a(\cdot)$ takva da za $1 + \xi x > 0$ vrijedi

$$\lim_{u \uparrow x_F} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \xi \neq 0, \\ e^{-x}, & \xi = 0. \end{cases} \tag{15}$$

(c) Za $x, y > 0, y \neq 1$ vrijedi

$$\lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1}, & \xi \neq 0, \\ \frac{\ln x}{\ln y}, & \xi = 0. \end{cases} \tag{16}$$

5 Procjena parametara pod pretpostavkom maksimalne domene atrakcije

Pretpostavimo da su za neki $\xi \in \mathbb{R}$ slučajne varijable $X_1, \ldots, X_n$ n.j.d. iz $F \in MDA(H_\xi)$. Prema propoziciji (1), $F \in MDA(H_\xi)$ je ekvivalentno

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\ln H_\xi(x)$$

za prikladne normirajuće nizove (c_n) i (d_n), te x pripada prikladnoj domeni zavisno o predznaku od ξ.

5.1 Pickandsov procjenitelj za $\xi \in \mathbb{R}$

Najjednostavniji i najstariji procjenitelj za $\xi \in \mathbb{R}$ je Pickandsov procjenitelj (1975.). Ovaj procjenitelj koristi razlike kvantila i baziran je na uređajnim statistikama n-tog reda $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$:

$$\hat{\xi}^{(P)}_{k,n} := \frac{1}{\ln 2} \cdot \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}, \quad k = 1, \ldots, \frac{n}{4}. \tag{17}$$

Pickandsov procjenitelj je konzistentan procjenitelj za $\xi \in \mathbb{R}$:

Teorem 8. *Neka su $X_1, X_2, \ldots$ n.j.d. slučajne varijable s funkcijom distribucije F. Pretpostavimo da je $F \in MDA(H_\xi)$, gdje je $\xi \in \mathbb{R}$. Tada za $n \to \infty$, $k = k(n) \to \infty$, $\frac{k}{n} \to 0$,*

$$\hat{\xi}^{(P)}_{k,n} \xrightarrow{P} \xi.$$

Dokaz. Vidi u [3] dokaz teorema 3.3.1., strana 84. □

Osim svojstva konzistentnosti, Pickandsov procjenitelj zadovoljava i svojstvo asimptotske normalnosti.

Teorem 9. *Neka su $X_1, X_2, \ldots$ n.j.d. slučajne varijable s funkcijom distribucije $F \in MDA(H_\xi), \xi \in \mathbb{R}$.*
Neka je $\hat{\xi}^{(P)} = \hat{\xi}^{(P)}_{k,n}$ Pickandsov procjenitelj (17). Tada

$$\sqrt{k}(\hat{\xi}^{(P)} - \xi) \xrightarrow{d} N(0, \nu(\xi)), \quad n \to \infty,$$

gdje je

$$\nu(\xi) = \frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1)\ln 2)^2}.$$

Dokaz. Vidi u [7] teorem 6.4.1., strana 328. □

5.2 Deckers-Einmahl-de Haanov procjenitelj za $\xi \in \mathbb{R}$

Jedan od najpoznatijih procjenitelja je Hillov procjenitelj. Taj je procjenitelj dizajniran za distribucije koje pripradaju $MDA(H_\xi)$ pri čemu je $\xi > 0$. (više u [7] strana 330.). To se pokazalo kao nedostatak kod primjerice osiguranja i financija gdje se često susrećemo sa negativnim vrijednostima. Iz tog razloga Deckers, Einmahl i de Haan napravili su proširenje Hillovog procjenitelja koji pokriva cijelu klasu H_ξ, $\xi \in \mathbb{R}$:

$$\hat{\xi}^{(M)}_{k,n} = 1 + M_n^{(1)} + \frac{1}{2}\Big(\frac{(M_n^{(1)})^2}{M_n^{(2)}} - 1\Big)^{-1} \tag{18}$$

pri čemu je

$$M_n^{(i)} = \frac{1}{k}\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k+1,n})^i.$$

Obzirom da $M_n^{(1)}$ i $M_n^{(2)}$ možemo interpretirati kao empirijske momente, $\hat{\xi}^{(M)}_{k,n}$ još nazivamo i momentni procjenitelj od ξ.

Teorem 10. *Neka su $X_1, X_2, \ldots$ n.j.d. slučajne varijable s funkcijom distribucije F. Pretpostavimo da je $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$ i $x_F > 0$. Tada za $k = k(n) \to \infty$, $\frac{k}{n} \to 0$, kada $n \to \infty$*

$$\hat{\xi} \xrightarrow{P} \xi.$$

$\hat{\xi}$ je konzistentan procjenitelj za ξ.

Dokaz. Sličan dokazu teorema 8. □

5.3 Procjenitelj krajnje točke x_F

Pretpostavimo da je $F \in MDA(H_\xi)$. Ranije smo komentirali kako je funkcija distribucije ekstremnih vrijednosti konačna ako je $\xi < 0$, odnosno postoji krajnja desna točka u tom slučaju. Iz tog razloga pretpostavimo da je $\xi < 0$. Procjenitelj $\hat{x}_F$ ima oblik:

$$\hat{x}_F = \hat{d}_{n/k} - \frac{\hat{c}_{n/k}}{\hat{\xi}} \tag{19}$$

pri čemu je

$$\hat{c}_{n/k} = (1 - \min\{0, \hat{\xi}\})X_{n-k,n}M_n^{(1)} \quad i \quad \hat{d}_{n/k} = X_{n-k,n}.$$

Više detalja može se vidjeti u [3].

6 Primjena teorije ekstremnih vrijednosti u atletici

Kao što smo već i spomenuli, teorija ekstremnih vrijednosti koristi se u raznim područjima, od financija, osiguranja, hidrologiji pa do sportskih rekorda. U ovom dijelu rada pokušati ćemo primjenom teorije ekstremnih vrijednosti istražiti koji je ultimativni svjetski rekord u trčanju na 100 metara za muškarce i za žene. Nastojat ćemo odgovoriti na pitanje: Koliko brzo možemo trčati? Primjer se temelji na članku [5].

6.1 Baza podataka

Potrebne podatke za istraživanje preuzeli smo s internetske stranice [9]. Obzirom da se kontrole na doping rade tek od 1989.godine, promatrat ćemo ostvarene rezultate počevši od 1.1.1991.g. kako bi izbjegli što više rezultata nastalih pod utjecajem ne dozvoljenih supstanci. Period promatranja rezultata završava 19.6.2008.g. Kako se teorija ekstremnih vrijednosti bavi samo maksimalnim vrijednostima, to znači da je dovoljno uzeti samo najbolje rezultate, odnosono uzeti ćemo samo osobne rekorde svakog pojedinog atletičara. Dobili smo sljedeće:

Disciplina	Broj podataka	Najbolji	Najgori
100m muškarci	3026	9.72	10.55
100m žene	1082	10.65	11.60

Kako će se svaki atletičar pojaviti samo jednom, time ćemo si osigurati nezavisnost podataka. Rezultati u trčanju na 100m su izraženi u stotinkama. Zbog toga će se pojaviti velik broj onih natjecatelja koji imaju jednaki osobni rekord, što može izazvati probleme u procjenama. Kako bi to izbjegli rezultate ćemo rasporediti . Ako m atletičara ima isto vrijeme, na primjer 10.08 sekundi, tada ćemo njihove rezultate rasporediti na intervalu $\langle 10.075, 10.085\rangle$ na sljedeći način:

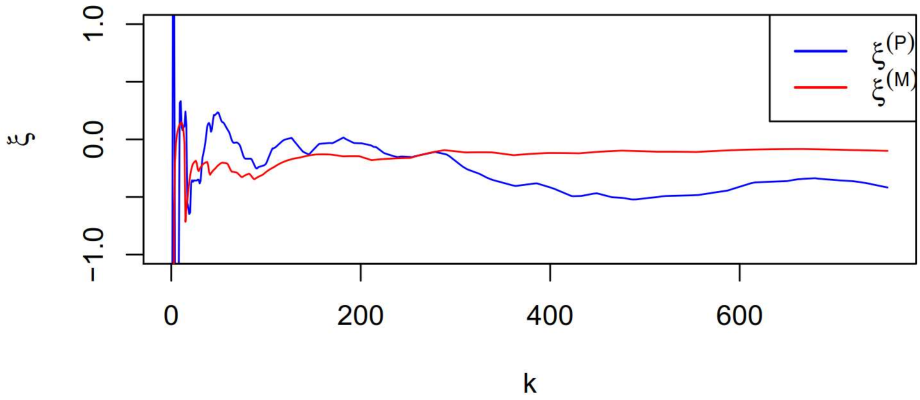
$$t_j = 10.075 + 0.01\frac{2j-1}{2m}, \quad j = 1, \ldots, m.$$

Za kraj ćemo dobivena vremena transformirati u brzine. Time ćemo dobiti da je najbrži rezultat ujedno i najbolji.

6.2 Procjena indeksa ekstremne vrijednosti

U prethodnom poglavlju smo naveli dva procjenitelja indeksa ekstremne vrijednosti, Pickandsov i Deckers-Einmahl-de Haanov (poznat kao momentni procjenitelj). Pomoću tih procjenitelja pokušat ćemo procjeniti vrijednost parametra ξ.

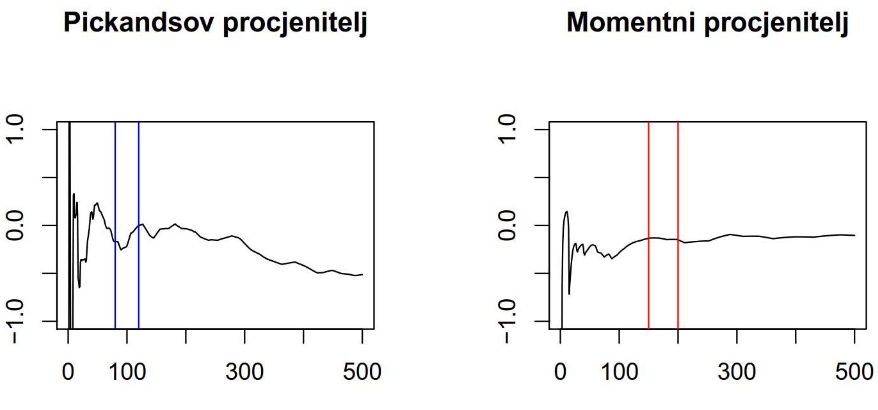
Prvo ćemo nacrtati procjenitelje indeksa razdiobe ekstremnih vrijednosti kao funkcije od k. Za to ćemo koristiti gotove naredbe u R-u. Napomenimo kako program R kod Pickandsovog procjenitelja uzima samo prvu četvrtinu podataka. Razlog tome je što će za velike vrijednosti k graf uvijek početi ili padati ili rasti te nas takvi podaci neće ni zanimati u procjeni. Zbog toga ćemo i za momentni procjenitelj uzeti samo prvu četvrtinu podataka, odnosno prvih 756 podataka kod muškaraca te 271 kod žena.



Slika 6: Procjenitelji $\hat{\xi}$

Uočimo kako je većina procjenjenih vrijednosti negativna što nam dopušta da zaključimo da funkcija distribucije podataka ima desnu krajnju točku. Također možemo uočiti kako je graf za male vrijednosti volatilan, dok za veće vrijednosti graf počinje padati. Pitamo se: Kako odabrati prikladan k za procjenitelje? Upravo je to osnovni problem kod spomenutih procjenitelja. Metode koje se koriste u praksi ima jako puno. Najčešće korištena metoda opisana je u [5]. Tu ćemo metodu i mi iskoristiti na način da ćemo odrediti prvo stabilno područje od k, područje na kojem krivulja ne oscilira jako, te na njemu izračunati prosječnu vrijednost od $\hat{\xi}$.

Kod Pickandsovog procjenitelja uzmimo područje $k \in [70, 120]$ kao stabilno područje, a kod momentnog procjenitelja $k \in [150, 200]$.

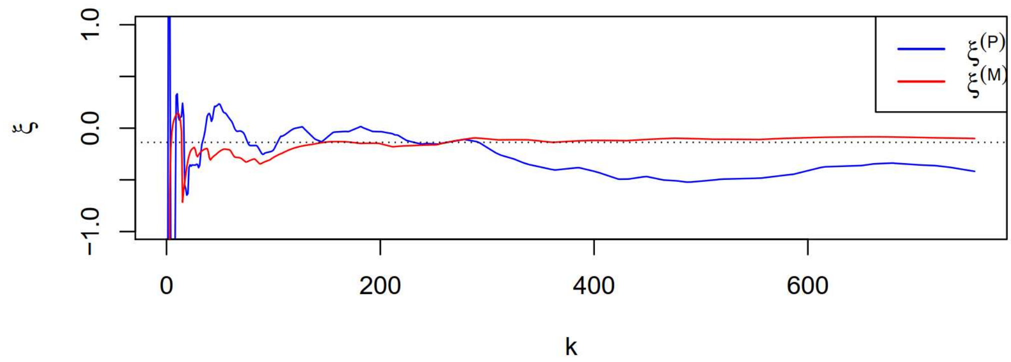


Slika 7: Stabilna područja za procjenitelj $\hat{\xi}$

Izračunavanjem prosječne vrijednosti procjenitelja na spomenutim područjima dobivamo sljedeće vrijednosti:

Procjenitelj	Raspon	$\hat{\xi}$
Pickandsov	70 - 120	-0.13582
Momentni	150 - 200	-0.13898

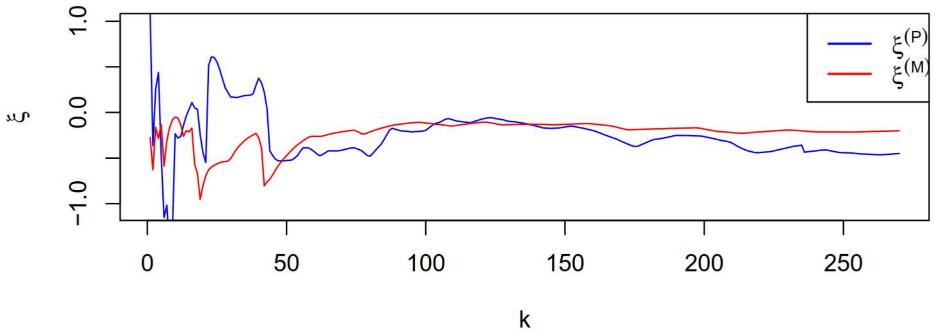
Prosjek dobivenih vrijednosti daje nam konačnu procjenu parametra distribucije ekstremnih vrijednosti $\hat{\xi} = -0.13739$. Prikažimo na slici konačnu procjenu.



Slika 8: Procjenitelj $\hat{\xi}$ za muškarce

Analognim postupkom ćemo odrediti procjenitelj $\hat{\xi}$ za podatke o ženskim osobnim rekordima na 100 metara.

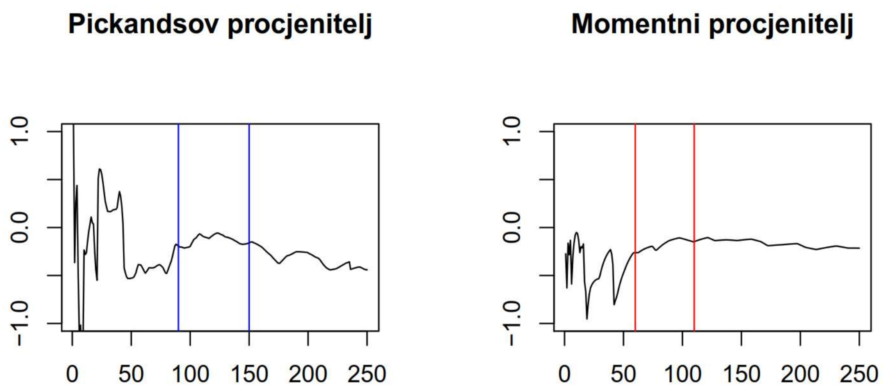
Crtanjem procjenitelja $\hat{\xi}^{(P)}$ i $\hat{\xi}^{(M)}$ u ovisnosti o k dobivamo sljedeće grafove:



Slika 9: Procjenitelji $\hat{\xi}$

Primjetimo da su, kao i kod muškaraca, skoro sve vrijednosti procjenitelja negativne, te također imamo veliku volatilnost u prvom dijelu dok prema kraju imamo da graf počinje lagano padati.

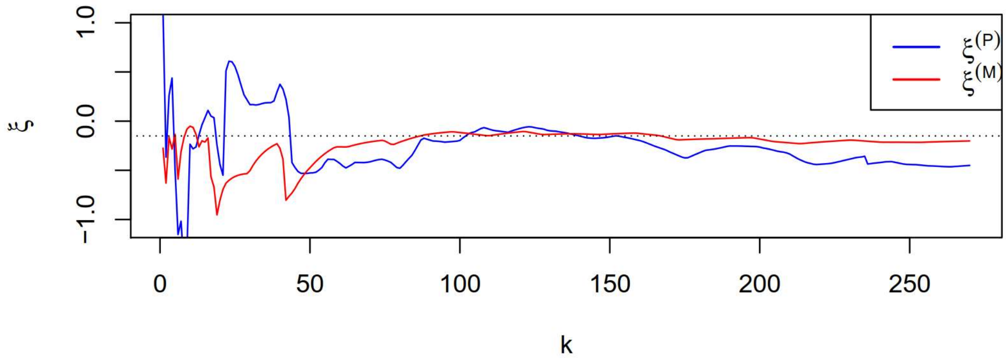
Određivanjem stabilnih područja za Pickandsov i momentni procjenitelj dobivamo sljedeće:



Slika 10: Stabilna područja za procjenitelj $\hat{\xi}$

Procjenitelj	Raspon	$\hat{\xi}$
Pickandsov	90 - 150	-0.13011
Momentni	60 - 110	-0.17491

Konačno, izračunamo li prosjek ovih procjena dobivamo konačnu procjenjenu vrijednost $\hat{\xi} = -0.15051$.



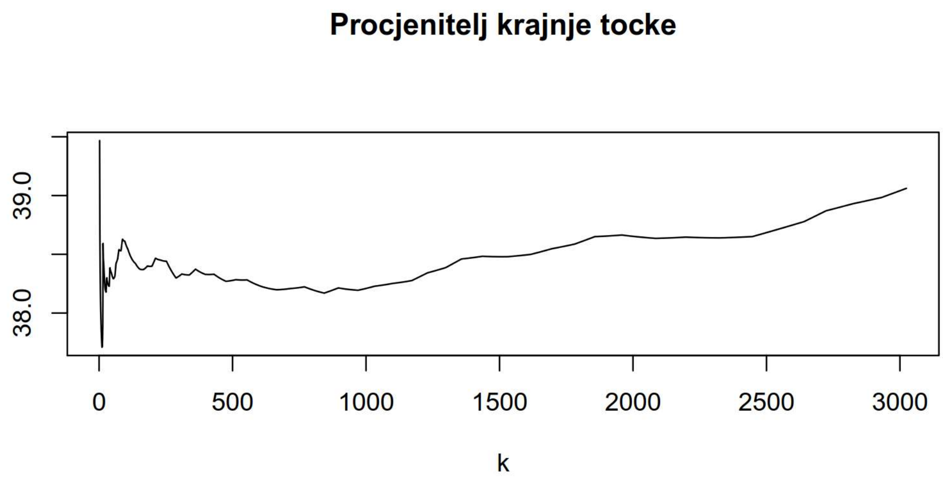
Slika 11: Procjenitelj $\hat{\xi}$ za žene

6.3 Procjena krajnje točke

Ranije smo zaključili da vrijedi $\hat{\xi} < 0$, tj. da distribucija ima krajnju točku pa primjenjujemo formulu (19).

Kako bi procjenili krajnju točku, uvrstit ćemo konačni $\hat{\xi} = -0.13739$ u formulu (19). To znači da izraz (19) ovisi o k samo kroz $M_n^{(1)}$ i $X_{n-k,n}$.

Postupak određivanja procjenitelja $\hat{x}_F$ provesti ćemo na isti način kao kod određivanja parametra oblika. Nacrtajmo graf funkcije $\hat{x}_F$ nasuprot k.

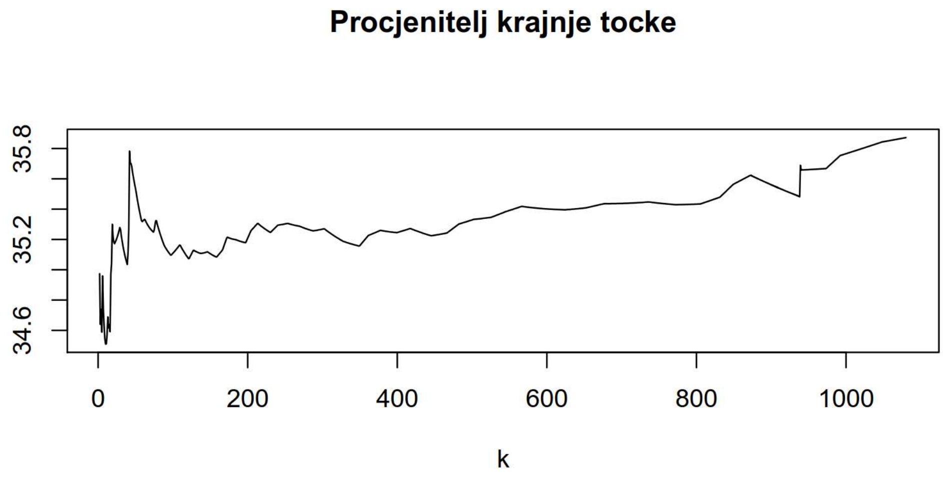


Slika 12: Procjenitelj krajnje točke $\hat{x}_F$

Određivanjem stabilnog područja $k \in [80, 130]$ te izračunavanjem prosjeka $\hat{x}_F$ na tom području dobivamo konačnu procjenu krajnje točke $x_F = 38.54$ (prisjetimo se: podatke iskazane u vremenu pretvorili smo u brzine). Nakon što brzinu transformiramo natrag u sekunde, dolazimo do ultimativnog rekorda od 9.34 sec za muškarce.

Trenutačni svjetski rekord ostvaren 16.08.2009.g. u muškoj konkurenciji drži jamajkanac Usain Bolt, a on iznosi 9.58 sec.

U nastavku ćemo odrediti ultivativni rekord za žene na 100m.



Slika 13: Procjenitelj krajnje točke $\hat{x}_F$

Uzimanjem stabilnog područja $k \in [110, 160]$ te istim postupkom dolazimo do ultimativnog rekorda od 10.25 sec.

Trenutačni svjetski rekord kod žena iznosi 10.49 sec, a ostvarila ga je amerikanka Florence Griffith-Joyner 16.07.1988.g.

Možemo zaključiti kako i u muškoj i u ženskoj konkurenciji ima još mjesta za napredak.

Literatura

[1] J. BEIRLANT, Y. GOEGEBEUR, J. TEUGELS, J. SEGERS, D. DE WAAL, C. FERRO, *Statistics of Extremes, Theory and Applications*, Wiley, Chichester, 2004.

[2] S. G. COLES, *An Introduction to Statistical Modelling of Extreme Values*, Springer, New York, 2001.

[3] L. DE HAAN, A. FERREIRA, *Extreme Value Theory,An Introduction*, Springer, New York, 2006.

[4] A. DEKKERS, J. EINMAHL, L. DE HAAN, *A Moment Estimator for the Index of an Extreme-Value distribution*, The Annals of Statistics, 17 (1989), 1833-1855.

[5] J. EINMAHL, *Ultimate 100m world records throught extreme-value theory*, Statistica Neerlandica, 65(2011), 32-42.

[6] J. EINMAHL, J. MAGNUS, *Records in Athletics Through Extreme-Value Theory*, American Statistical Association, 103(2008), 1382-1391.

[7] P. EMBRECHTS, C. KLUPPELBERG, T. MIKOSCH, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.

[8] S. KOTZ, S. NADARAJAH, *Extreme Value Distributions, Theory and Applications*, Imperial College Press, London, 2000.

[9] *World Athletics* (06.05.2021.), URL: `https://www.worldathletics.org/`

Sažetak

U ovom diplomskom radu opisane su distribucije ekstremnih vrijednosti. Te se distribucije jedine mogu pojaviti kao granične distribucije parcijalnih maksimuma (minimuma), a postoje tri tipa: Fréchetova, Weibullova i Gumbelova distribucija.
Opisani su granični rezultati i identificirane maksimalne domene atrakcije pojedine distribucije. Ova tri oblika distribucija mogu se objediniti u jedan zapis, generaliziranu distribuciju ekstremnih vrijednosti.
U radu su također navedeni neki procjenitelji parametara. Na kraju rada navedena je primjena teorije kroz primjer. Primjer se bazira na članku J. Einmahla: "Ultimate 100m world records throught extreme-value theory".

Ključne riječi

max-stabilna distribucija, Fréchetova distribucija, Weibullova distribucije, Gumbelova distribucija, distribucije ekstremnih vrijednosti, maksimalna domena atrakcije, generalizirana distribucija ekstremnih vrijednosti, Pickandsov procjenitelj, momentni procjenitelj, procjenitelj krajnje točke

Title and summary

Extreme value distributions and application
Extreme value distributions are described in this thesis. Those distributions are the only possible limit laws for maxima (minima) and there exists three types of distribution functions: Fréchet, Weibull and Gumbel distribution.
The limit laws are described and maximum domain of attraction for every distribution is identified. Those three distributions can be represented as one, the generalised extreme value distribution.
The parameter estimation is discussed and a practical example is given. The example is based on the article by J. Einmahl: "Ultimate 100m world records throught extreme value theory".

Keywords
max-stable distribution, Fréchet distribution, Weibull distribution, Gumbel distribution, extreme value distribution, maximum domain of attraction, generalised extreme value distribution, Pickand's esimator, moment estimator, endpoint estimator

Životopis

Ja sam Martina Peranić (djevojačko Filić). Rođena sam 16. veljače 1993. godine u Rijeci. Majka sam blizanaca Lee i Jakova. Pohađala sam Osnovnu školu "Fran Krsto Frankopan" u Omišlju od 1999. do 2007. godine te nakon toga upisala srednju "Ekonomsku školu Mije Mirkovića" u Rijeci, smjer Ekonomist. Godine 2011. sam završila srednju školu te upisala Preddiplomski studij matematike na Odjelu za matematiku Sveučilišta u Rijeci. Preddiplomski studij sam završila 2016. godine uz završni rad "Funkcije izvodnice i slučajne sume" i mentorstvo doc. dr. sc. Danijela Krizmanića. Iste godine sam upisala diplomski studij na Odjelu za metematiku u Osijeku, smjer Financijska matematika i statistika. Tijekom diplomskog studija odradila sam stručnu praksu u Erste banci u Zagrebu na Odjelu za kreditne rizike.