

# Primjena teorije ekstremnih vrijednosti na rezultate u atletici

---

Čigir, Lara

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:501579>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-22**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Lara Čigir

**PRIMJENA TEORIJE EKSTREMNIH**  
**VRIJEDNOSTI NA REZULTATE U**  
**ATLETICI**

Diplomski rad

Voditelj rada:  
prof.dr.sc. Bojan Basrak

Zagreb, studeni 2018.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

*Hvala Ivoni i Ivani na zajedničkom putovanju od osnovne škole do fakulteta.*

*Hvala Ivani i Antoniji na uljepšanim danima diplomskog studija.*

*Hvala mami, tati i Leonu na neizmjernoj podršci i neupitnoj vjeri u mene.*

*Hvala Mariu što je uvijek bio tu.*

# Sadržaj

Sadržaj	iv
Uvod	2
<b>1 Uvod u teoriju ekstremnih vrijednosti</b>	<b>3</b>
1.1 Distribucije ekstremnih vrijednosti . . . . .	5
1.2 Domena privlačnosti . . . . .	7
<b>2 Procjena indeksa ekstremne vrijednosti</b>	<b>9</b>
2.1 Uvjet drugog reda . . . . .	10
2.2 Hillov procjenitelj . . . . .	11
2.3 Procjenitelj metodom maksimalne vjerodostojnosti (MLE) . . . . .	12
2.4 Procjenitelj metodom momenata (MOM) . . . . .	15
2.5 Asimptotska svojstva procjenitelja indeksa . . . . .	17
<b>3 Procjena ekstremnih kvantila i repa razdiobe</b>	<b>19</b>
3.1 Procjena funkcije skale . . . . .	19
3.2 Procjena ekstremnih kvantila . . . . .	20
3.3 Procjena repa . . . . .	21
3.4 Procjena desne krajnje točke razdiobe . . . . .	21
<b>4 Primjena teorije ekstremnih vrijednosti na rekorde u atletici</b>	<b>23</b>
4.1 Podaci . . . . .	23
4.2 Procjena indeksa ekstremne vrijednosti . . . . .	25
4.3 Procjena ultimativnog svjetskog rekorda pojedine discipline . . . . .	37
4.4 Procjena kvalitete trenutnog svjetskog rekorda . . . . .	42
4.5 Osjetljivost procjena na izbor $k$ . . . . .	49
4.6 Zaključak . . . . .	53
<b>Bibliografija</b>	<b>55</b>

# Uvod

Koliko dugačka mora biti zračna pista na aerodromu kako bi se omogućilo sigurno slijetanje aviona? Koliko visoke moraju biti brane u Nizozemskoj kako bi štitele od visokih valova? Koliko dugo možemo živjeti? Koliko brzo možemo trčati? Koliko daleko možemo skočiti?

Sva navedena pitanja povezana su s ekstremima, no u ovome radu posebno nas zanimaju posljednja dva, koja se odnose na svjetske rekorde u atletici. Točnije, želimo odgovoriti na dva glavna pitanja: Koji je ultimativan svjetski rekord u pojedinoj atletskoj disciplini kao što je na primjer ženski maraton? Koliko je "dobar" trenutni rekord tj. koliko ga je teško poboljšati? Ovim pitanjima pristupit ćemo pomoću teorije ekstremnih vrijednosti (TEV) i odgovarajućih statističkih procedura.

TEV je grana statistike koja se bavi ekstremnim odstupanjima od medijana vjerojatnosne distribucije. Ona nastoji procijeniti, iz danog uzorka slučajnih varijabla, vjerojatnost ekstremnih događaja koji su vezani uz rep distribucije. Jedan od značajnijih rezultata teorije ekstremnih vrijednosti je da maksimumi uzorka nezavisnih jednako distribuiranih slučajnih varijabla, nakon odgovarajuće normalizacije, mogu konvergirati po distribuciji samo u jednu od idućih tri distribucija, Gumbelovu, Frèchetovu ili Weibullovu.

Standardna forma ovih triju distribucija naziva se generalizirana distribucija ekstremnih vrijednosti (GEV distribucija), a opisana je pomoću tri parametra, funkcije skale, lokacijske funkcije te indeksa razdiobe ekstremne vrijednosti. GEV distribucija ima široku primjenu u upravljanu rizicima, financijama, osiguranjima, hidrologiji, klimatologiji i mnogim drugim područjima koji se bave ekstremnim događajima.

U ovome radu opisat ćemo glavne rezultate teorije ekstremnih vrijednosti te objasniti kako se procjenjuju sva tri gornja parametra. Posebnu pažnju posvetit ćemo indeksu razdiobe ekstremnih vrijednosti te ćemo u Poglavlju 2 navesti nekoliko njegovih procjenitelja. U Poglavlju 3 opisat ćemo glavne rezultate važne za procjenu ekstremnih

kvantila te konačno procjenu desne krajnje točke razdiobe. U Poglavlju 4 primijenit ćemo opisanu teoriju na prikupljene podatke o osobnim rekordima najboljih atletičara za 10 atletskih disciplina te dati odgovore na gore postavljena pitanja. Podatke ćemo obrađivati u statističkom programu R.

# Poglavlje 1

## Uvod u teoriju ekstremnih vrijednosti

Neka su  $X_1, X_2, \dots, X_n$  nezavisne i jednako distribuirane slučajne varijable s funkcijom distribucije  $F$  te neka je  $x^*$  njezina desna krajnja točka tj.  $x^* := \sup\{x : F(x) < 1\}$  koja može biti beskonačna. Tada

$$M_n = \max(X_1, X_2, \dots, X_n) \xrightarrow{P} x^*, \quad n \rightarrow \infty,$$

jer vrijedi

$$P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = F^n(x),$$

što konvergira u 0 za  $x < x^*$  ili u 1 za  $x \geq x^*$ . Dakle, kako bismo dobili nedegeneriranu graničnu distribuciju za  $M_n$  potrebno je provesti normalizaciju. Pretpostavimo da postoji niz pozitivnih realnih brojeva  $a_n$  te niz realnih brojeva  $b_n$  takvi da

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n}$$

ima nedegeneriranu graničnu distribuciju kako  $n \rightarrow \infty$  tj. da vrijedi

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \tag{1.1}$$

za svaku točku neprekidnosti  $x$  od  $G$ . Klasa funkcija distribucije  $G$  koje se mogu pojaviti kao limes u relaciji (1.1) čine klasu *funkcija distribucije ekstremnih vrijednosti*. Za svaku od tih graničnih distribucija želimo pronaći nužne i dovoljne uvjete za početnu funkciju distribucije  $F$  iz (1.1). Klasu funkcija distribucije  $F$  koje zadovo-



ljavaju (1.1) nazivamo *maksimalna domena atrakcije od  $G$*  ili *domena atrakcije*.

Logaritmiranjem relacije (1.1) dobivamo ekvivalentnu relaciju koja vrijedi za svaku točku neprekidnosti  $x$  za koju je  $0 < G(x) < 1$ ,

$$\lim_{n \rightarrow \infty} n \log F(a_n x + b_n) = \log G(x) . \quad (1.2)$$

Za svaki takav  $x$  vrijedi da  $F(a_n x + b_n) \rightarrow 1$  pa

$$\lim_{n \rightarrow \infty} \frac{-\log F(a_n x + b_n)}{1 - F(a_n x + b_n)} = 1 .$$

Stoga relaciju (1.2) možemo zapisati kao

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log G(x) .$$

**Definicija 1.** *Generalizirana inverzna funkcija neopadajuće funkcije  $f$  tj.  $f^{\leftarrow}$  definirana je relacijom*

$$f^{\leftarrow}(x) := \inf\{y : f(y) \geq x\} .$$

**Teorem 1.0.1.** *Neka je  $a_n$  niz pozitivnih realnih brojeva,  $b_n$  niz realnih brojeva te neka je  $G$  nedegenerirana funkcija distribucije. Tada su sljedeće tvrdnje ekvivalentne:*

1.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x)$$

za svaku točku neprekidnosti  $x$  od  $G$ .

2.

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log G(x) ,$$

za svaku točku neprekidnosti  $x$  od  $G$  za koju vrijedi  $0 < G(x) < 1$ ,  $a(t) := a_{[t]}$  i  $b(t) := b_{[t]}$  (pri čemu je  $[t]$  cjelobrojni dio od  $t$ ).

3.

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) , \quad (1.3)$$

za svaku točku neprekidnosti  $x > 0$  funkcije  $D(x) = G^{\leftarrow}(e^{-1/x})$  pri čemu je  $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ ,  $a(t) := a_{[t]}$ ,  $b(t) := b_{[t]}$ .

*Dokaz:* Vidi u [2] dokaz teorema 1.1.2., str. 6.

## 1.1 Distribucije ekstremnih vrijednosti

Sada možemo identificirati klasu nedegenerirajućih funkcija distribucija koje se pojavljuju kao limes u relaciji (1.1). Tu klasu nazivamo *klasa distribucije ekstremnih vrijednosti*.

**Teorem 1.1.1.** (*Fisher i Tippett (1928), Gnedenko (1943)*) *Klasa distribucije ekstremnih vrijednosti je  $G_\gamma(ax + b)$  gdje je  $a > 0$ ,  $b \in \mathbf{R}$  te*

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad (1.4)$$

za  $\gamma \in \mathbf{R}$  i gdje je za  $\gamma = 0$  desna strana interpretirana kao  $\exp(-e^{-x})$ .

*Dokaz:* Vidi u [2] dokaz teorema 1.1.3., str. 7.

**Definicija 2.** *Parametar  $1/\gamma$  iz relacije (1.4) naziva se repni indeks pri čemu je  $\gamma$  indeks razdiobe ekstremne vrijednosti.*

Rezultat prethodnog teorema pokazuje da granične funkcije distribucije formiraju jednostavnu eksplicitnu jednoparametarsku familiju, do na nizove  $a_n$  i  $b_n$ . Promotrimo posebno slučajeve za podklase  $\gamma > 0$ ,  $\gamma = 0$  te  $\gamma < 0$ :

1. Za  $\gamma > 0$  je  $G_\gamma(x) < 1$  za svaki  $x$  pa je desna krajnja točka funkcije distribucije beskonačna.
2. Za  $\gamma = 0$  desna krajnja točka jednaka je beskonačnosti.
3. Za  $\gamma < 0$  desna krajnja točka funkcije distribucije jednaka je  $-1/\gamma$ .

Ako je  $G = G_\gamma$  u relaciji (1.1) za neki  $\gamma \in \mathbf{R}$  kažemo da je funkcija distribucije  $F$  u domeni atrakcije od  $G_\gamma$  te pišemo  $F \in \mathcal{D}(G_\gamma)$ . Rezultat Teorema 1.1.1. dovodi do reformulacije Teorema 1.0.1.

**Teorem 1.1.2.** *Za indeks ekstremne vrijednosti  $\gamma \in \mathbf{R}$  sljedeće tvrdnje su ekvivalentne:*

1. *Postoje  $a_n > 0$ ,  $b_n \in \mathbf{R}$  takvi da je*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad (1.5)$$

za svaki  $x$  takav da je  $1 + \gamma x > 0$ .

2. *Postoji pozitivna funkcija  $a$  takva da za svaki  $x > 0$  vrijedi*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma}, \quad (1.6)$$

gdje se za  $\gamma = 0$  desna strana jednakosti interpretira kao  $\log x$ .

3. Postoji pozitivna funkcija  $a$  takva da je

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-1/\gamma}, \quad (1.7)$$

za svaki  $x$  takav da je  $1 + \gamma x > 0$ .

4. Postoji pozitivna funkcija  $f$  takva da je

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (1.8)$$

za svaki  $x$  takav da je  $1 + \gamma x > 0$  te gdje je  $x^* = \sup\{x : F(x) < 1\}$ .

*Dokaz:* Vidi u [2] dokaz teorema 1.1.6., str. 10.

Uočimo kako smo tvrdnju (1.6) dobili reformulacijom tvrdnje (1.3) koristeći  $D_\gamma = G_\gamma^{\leftarrow}(e^{-1/x})$ . Inverznu funkciju funkcije  $G_\gamma(e^{-1/x})$  dobili smo računom:

$$\begin{aligned} \exp -(1 + \gamma x)^{-1/\gamma} = y &\Leftrightarrow (1 + \gamma x)^{-1/\gamma} = -\ln y \Leftrightarrow \\ &\Leftrightarrow 1 + \gamma x = \frac{1}{(-\ln y)^\gamma} \Leftrightarrow x = \frac{1/(-\ln y)^\gamma - 1}{\gamma} \end{aligned}$$

iz čega slijedi

$$G_\gamma^{\leftarrow}(e^{-1/x}) = \frac{1/(1/x)^\gamma - 1}{\gamma} = \frac{x^\gamma - 1}{\gamma}.$$

Nadalje, koristili smo  $b(t) = b_{[t]} := U(t)$  te dobili

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

za  $x > 0$ ,  $\gamma \in \mathbf{R}$  i odgovarajuću pozitivnu funkciju  $a$ . Tada vrijedi

$$U(tx) \approx U(t) + a(t) \frac{x^\gamma - 1}{\gamma}.$$

Uočimo sada kako nam relacija (1.6) omogućava procjenu ekstremnog kvantila prilagodbom funkcije  $(x^\gamma - 1)/\gamma$  funkciji kvantila  $U$ .

## 1.2 Domena privlačnosti

Za funkciju distribucije  $F$  želimo pronaći dovoljne uvjete uz koje će postojati niz realnih brojeva  $a_n > 0$  i  $b_n$  takvih da vrijedi

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) .$$

Ti uvjeti, koje je pronašao von Mises (1936), zahtijevaju postojanje prve ili druge derivacije od  $F$ . Sljedeći teorem govori o dovoljnom uvjetu za pripadnost domeni atrakcije, a navedeni uvjet naziva se *von Misesov uvjet*.

**Teorem 1.2.1.** *Neka je  $F$  funkcija distribucije te  $x^*$  njezina desna krajnja točka. Pretpostavimo da  $F''(x)$  postoji i da je  $F'(x)$  pozitivna funkcija za svaki  $x < x^*$ . Ako vrijedi*

$$\lim_{t \uparrow x^*} \left( \frac{1 - F}{F'} \right)' (t) = \gamma ,$$

ili ekvivalentno,

$$\lim_{t \uparrow x^*} \frac{(1 - F(t))F''(t)}{(F'(t))^2} = -\gamma - 1 ,$$

tada je  $F$  u domeni atrakcije od  $G_\gamma$ .

*Dokaz:* Vidi u [2] dokaz teorema 1.1.8., str. 15.

**Napomena 1.** *Uočimo kako u Teoremu 1.2.1. ekvivalencija jednostavno slijedi iz*

$$\begin{aligned} \left( \frac{1 - F}{F'} \right)' \rightarrow \gamma &\Leftrightarrow \frac{(-F')F' - (1 - F)F''}{(F')^2} \rightarrow \gamma \Leftrightarrow \frac{-(F')^2 - (1 - F)F''}{(F')^2} \rightarrow \gamma \\ &\Leftrightarrow -1 - \frac{(1 - F)F''}{(F')^2} \rightarrow \gamma \Leftrightarrow \frac{(1 - F)F''}{(F')^2} \rightarrow -\gamma - 1 \end{aligned}$$

Idući teorem govori o nužnim i dovoljnim uvjetima za pripadnost funkcije distribucije  $F$  domeni atrakcije od  $G_\gamma$ .

**Teorem 1.2.2.** *Funkcija distribucije  $F$  je u domeni atrakcije od distribucije graničnih vrijednosti  $G_\gamma$  ako i samo ako*

1. za  $\gamma > 0$ :  $x^*$  je beskonačno i

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma} \quad (1.9)$$

za svaki  $x > 0$ .

2. za  $\gamma < 0$ :  $x^*$  je konačno i

$$\lim_{t \downarrow 0} \frac{1 - F(x^* - tx)}{1 - F(x^* - t)} = x^{-1/\gamma}$$

za svaki  $x > 0$ .

3. za  $\gamma = 0$ :

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}$$

za svaki  $x \in \mathbf{R}$  gdje je  $f$  neka pozitivna funkcija. Uočimo kako  $x^*$  može biti konačno ili beskonačno.

*Dokaz:* Vidi u [2] dokaz teorema 1.2.1., str. 19.

## Poglavlje 2

# Procjena indeksa ekstremne vrijednosti

Prisjetimo se relacije (1.8) koja kaže da postoji pozitivna funkcija  $f$  takva da za sve  $x$  koji zadovoljavaju  $1 + \gamma x > 0$  vrijedi

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma},$$

gdje je  $x^* = \sup\{x : F(x) < 1\}$ . Neka je  $X$  slučajna varijabla s funkcijom distribucije  $F$  te neka je  $F \in \mathcal{D}(G_\gamma)$ . Tada nam prethodna relacija govori da za  $x > 0$ ,  $x < (0 \vee (-\gamma))^{-1}$  vrijedi

$$\lim_{t \uparrow x^*} P\left(\frac{X - t}{f(t)} > x \mid X > t\right) = (1 + \gamma x)^{-1/\gamma}. \quad (2.1)$$

Uvjetna distribucija od  $(X - t)/f(t)$  uz uvjet  $X > t$  ima graničnu distribuciju kada  $t \uparrow x^*$

$$H_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma}, \quad 0 < x < (0 \vee (-\gamma))^{-1}, \quad (2.2)$$

gdje se za  $\gamma = 0$  desna strana interpretira kao  $1 - e^{-x}$ . Ta klasa funkcija distribucije naziva se *klasa generaliziranih Pareto distribucija (GPD)*. Relacija (1.8) nam zapravo govori da se od nekog praga  $t$  nadalje tj. za  $X > t$  funkcija distribucije može aproksimativno zapisati kao

$$1 - F(x) \approx (1 - F(t)) \left(1 - H_\gamma\left(\frac{x - t}{f(t)}\right)\right), \quad x > t. \quad (2.3)$$

Neka su  $X_1, X_2, \dots$  nezavisne jednako distribuirane slučajne varijable s funkcijom distribucije  $F$  te neka  $F_n$  odgovara *empirijskoj funkciji distribucije*,  $F_n(x) = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq x\}}$ . Zamijenimo li u relaciji (2.3)  $t$  sa  $X_{n-k,n}$  dobivamo

$$1 - F(x) \approx (1 - F(X_{n-k,n})) \left(1 - H_\gamma \left(\frac{x - X_{n-k,n}}{f(\frac{n}{k})}\right)\right).$$

Kako je  $1 - F(X_{n-k,n}) \approx 1 - F_n(X_{n-k,n}) = k/n$  konačno dobijemo

$$1 - F(x) \approx \frac{k}{n} \left(1 - H_\gamma \left(\frac{x - X_{n-k,n}}{f(\frac{n}{k})}\right)\right). \quad (2.4)$$

Da bismo mogli primijeniti dobivenu aproksimaciju (2.4) potrebno je procijeniti indeks razdiobe ekstremne vrijednosti  $\gamma$ .

## 2.1 Uvjet drugog reda

Prisjetimo se relacije (1.6)

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma},$$

gdje je  $U = (1/(1 - F))^\leftarrow$ . Pretpostavimo da postoji funkcija  $A$  konstantnog predznaka takva da ako  $\lim_{t \rightarrow \infty} A(t) = 0$  za svaki  $x > 0$ , tada

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - D_\gamma(x)}{A(t)} \quad (2.5)$$

postoji. Funkcija  $A$  može biti pozitivna i negativna. Neka je  $H$  granična funkcija te neka je  $a_1 = aA$  pri čemu su  $a$  i  $A$  funkcije iz (2.5). Tada relaciju (2.5) možemo zapisati kao

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t) - a(t)D_\gamma(x)}{a_1(t)} = H(x). \quad (2.6)$$

**Definicija 3.** Funkcija  $U$  zadovoljava uvjet drugog reda ako za neku pozitivnu funkciju  $a$  i neku pozitivnu ili negativnu funkciju  $A$  takvu da  $\lim_{t \rightarrow \infty} A(t) = 0$  vrijedi

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} =: H(x), \quad x > 0, \quad (2.7)$$

pri čemu  $H$  nije oblika  $H(x) = cD_\gamma(x)$  gdje je  $c$  konstanta, odnosno  $H$  nije linearna transformacija funkcije  $D_\gamma(x) = (x^\gamma - 1)/\gamma$ . Posebno,  $H$  ne smije biti jednako 0.

**Korolar 1.** Pretpostavimo da vrijedi relacija (2.7) za sve  $x > 0$ . Tada postoji pozitivna funkcija  $a$  i pozitivna ili negativna funkcija  $A$  takva da

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - D_\gamma(x)}{A(t)} = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds =: H_{\gamma,\rho}(x). \quad (2.8)$$

**Napomena 2.** Funkciju  $H_{\gamma,\rho}$  iz (2.8) ponekad ćemo pisati kao

$$H_{\gamma,\rho} := \frac{1}{\rho} \left( \frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right), \quad (2.9)$$

pri čemu je parametar  $\rho \leq 0$  i jednaka je:

$$\begin{cases} \frac{1}{\gamma} \left( x^\gamma \log x - \frac{x^\gamma - 1}{\gamma} \right), \text{ za } \rho = 0 \neq \gamma, \\ \frac{1}{\rho} \left( \frac{x^\rho - 1}{\rho} - \log x \right), \text{ za } \rho \neq 0 = \gamma, \\ \frac{1}{2} (\log x)^2, \text{ za } \rho = 0 = \gamma. \end{cases}$$

## 2.2 Hillov procjenitelj

Prisjetimo se relacije (1.9) iz Teorema 1.2.2. koji kaže da je funkcija  $F \in \mathcal{D}(G_\gamma)$  za  $\gamma > 0$  ako i samo ako

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \gamma > 0.$$

Ekvivalentna forma relacije (1.9) je

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \gamma. \quad (2.10)$$

Parcijalnom integracijom dobivamo

$$\int_t^\infty (1 - F(s)) \frac{ds}{s} = \int_t^\infty (\log u - \log t) dF(u)$$

stoga relaciju (2.10) možemo zapisati kao

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (\log u - \log t) dF(u)}{1 - F(t)} = \gamma. \quad (2.11)$$

Zamijenimo li u relaciji (2.11) parametar  $t$  s uređajnom statistikom  $X_{n-k,n}$  te  $F$  s empirijskom funkcijom distribucije  $F_n$  dobivamo *Hillov procjenitelj*  $\hat{\gamma}_H$

$$\hat{\gamma}_H := \frac{\int_{X_{n-k,n}}^\infty \log u - \log X_{n-k,n} dF_n(u)}{1 - F_n(X_{n-k,n})}$$

ili

$$\hat{\gamma}_H := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}. \quad (2.12)$$

Idući teorem govori o uvjetima uz koje je  $\hat{\gamma}_H$  asimptotski normalan.



**Teorem 2.2.1.** *Pretpostavimo da funkcija distribucije  $F$  zadovoljava uvjet drugog reda tj. neka za  $x > 0$  vrijedi*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad (2.13)$$

ili ekvivalentno,

$$\lim_{t \rightarrow \infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-1/\gamma} \frac{x^{\gamma/\rho} - 1}{\gamma\rho},$$

gdje je  $\gamma > 0$ ,  $\rho \leq 0$  i  $A$  funkcija takva da  $\lim_{t \rightarrow \infty} A(t) = 0$ . Tada

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \quad (2.14)$$

gdje je  $N$  standardna normalna distribucija i

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda$$

uz konačan  $\lambda$ .

*Dokaz:* Vidi u [2] dokaz teorema 3.2.5., str. 74.

## 2.3 Procjenitelj metodom maksimalne vjerodostojnosti (MLE)

Klasa funkcija distribucije koje zadovoljavaju  $F \in \mathcal{D}(G_\gamma)$  za neki  $\gamma \in \mathbf{R}$  ne mogu biti parametrizirane konačnim brojem parametara, stoga procjenitelj metodom maksimalne vjerodostojnosti za  $\gamma$  ne postoji.

No relacija (2.1) ukazuje da je za veće realizacije slučajne varijable  $X$  tj. za  $X > t$ , uvjetna distribucija od  $\frac{X-t}{f(t)}$  GP distribucija. Kako je klasa GP distribucija parametrizirana samo jednim parametrom  $\gamma$ , možemo primijeniti metodu maksimalne vjerodostojnosti i na taj način dobiti koristan procjenitelj za  $\gamma$ . Kako bi ispravno koristili uvjet  $X > t$  u (2.1) potrebna nam je iduća lema. No prije samog iskaza leme prisjetit ćemo se pojma uvjetne gustoće.

**Definicija 4.** *Neka je  $(X, Y)$  slučajni vektor. Distribuciju slučajne varijable  $X$  za danu vrijednost  $y$  slučajne varijable  $Y$  zovemo uvjetnom distribucijom od  $X$  za danu*

2.3. PROCJENITELJ METODOM MAKSIMALNE VJERODOSTOJNOSTI  
(MLE)

13

$Y = y$ . Uvjetne distribucije zadaju se uvjetnim gustoćama. Uvjetna gustoća (ili uvjetna funkcija vjerojatnosti) od  $X$  za dano  $Y = y$  zadaje se

$$f_{X|Y}(x|y) := P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbf{R}.$$

To je ujedno funkcija gustoće neprekidne vjerojatnosne razdiobe koja odgovara uvjetnoj distribuciji od  $X$  uz dano  $Y = y$ . Naime, neka je sa  $P(a \leq X \leq b|Y = y)$  označena uvjetna vjerojatnost da će  $X$  poprimiti vrijednosti u intervalu  $[a, b]$  uz dano  $Y = y$ . Tada je

$$P(a \leq X \leq b|Y = y) := \int_a^b f_{X|Y}(x|y) dx.$$

Primijetimo da se lijeva strana gornje jednakosti ne računa po formuli za uvjetnu vjerojatnost  $P(A|B) := P(A \cap B)/P(B)$  (jer je to nemoguće zbog  $P(Y = y) = 0$ ) iako ima istu interpretaciju.

**Lema 1.** Neka su  $X, X_1, X_2, \dots, X_n$  nezavisne jednako distribuirane slučajne varijable sa zajedničkom distribucijom  $F$  te neka su  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  uređajne statistike  $n$ -tog reda. Zajednička distribucija od  $\{X_{i,n}\}_{i=n-k+1}^n$  uz dano  $X_{n-k,n} := t$  za neki  $k = 1, \dots, n-1$  jednaka je zajedničkoj distribuciji vektora  $\{X_{i,k}^*\}_{i=1}^k$  nezavisnih i jednako distribuiranih slučajnih varijabli  $\{X_i^*\}_{i=1}^k$  s funkcijom distribucije

$$F_t(x) = P(X \leq x|X > t) = \frac{F(x) - F(t)}{1 - F(t)}, \quad x > t.$$

*Dokaz:* Neka su  $E_{1,n} \leq \dots \leq E_{n,n}$  uređajne statistike nezavisnih i jednako distribuiranih slučajnih varijabli s eksponencijalnom distribucijom tj.  $F(x) = P(E > x) = e^{-x}$ ,  $x > 0$ . Tada je uvjetna distribucija od  $(E_{n-k+1,n}, \dots, E_{n,n})$  uz uvjet  $E_{n-k,n} = t$ , jednaka distribuciji od  $(E_{1,k}^*, \dots, E_{k,k}^*)$  tj.

$$F(x) = P(E^* > x) = e^{-(x-t)}, \quad \text{za } x > t. \quad (2.15)$$

Neka je  $V(x) := (-\log(1 - F(x)))^\leftarrow$ . Uvjetna distribucija od  $(V(E_{n-k+1,n}), \dots, V(E_{n,n}))$  uz uvjet  $V(E_{n-k,n}) = V(t)$  jednaka je distribuciji od  $(V(E_{1,k}^*), \dots, V(E_{k,k}^*))$ . Sada za  $x > V(t)$  vrijedi  $P(V(E^*) > x) = P(-\log(1 - F(E^*)^\leftarrow) > x) = P(E^* > -\log(1 - F(x)))$ , što je prema (2.15) jednako

$$P(E^* > -\log(1 - F(x))) = e^{-(-\log(1 - F(x)) - t)} = e^{\log(1 - F(x))} e^t = (1 - F(x)) e^t.$$

Zapišemo li  $t = V^\leftarrow(V(t)) = -\log(1 - F(V(t)))$  dobivamo

$$P(V(E^*) > x) = (1 - F(x)) e^t = (1 - F(x)) e^{-\log(1 - F(V(t)))} = \frac{1 - F(x)}{1 - F(V(t))},$$

što je jednako

$$\frac{1 - F(x)}{1 - F(V(t))} = \frac{P(X > x)}{P(X > V(t))} = \frac{P(X > x \cap X > V(t))}{P(X > V(t))} = P(X > x | X > V(t)).$$

Ovime smo dokazali tvrdnju leme za svaku neprekidnu strogo rastuću funkciju distribucije  $F$ . Na sličan se način tvrdnja dokaže i za opći slučaj.

Neka su  $X_1, \dots, X_n$  nezavisne jednako distribuirane slučajne varijable sa zajedničkom distribucijom  $F$ . Kako bismo procijenili  $\gamma$  potrebno je koncentrirati se na  $k$ -torku statistika gornjeg reda tj. na  $(X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n})$  ili ekvivalentno na

$$(Z_0, Z_1, \dots, Z_k) := (X_{n-k,n}, X_{n-k+1,n} - X_{n-k,n}, \dots, X_{n,n} - X_{n-k,n}).$$

Funkcija vjerodostojnosti dobiva se iz uvjetne distribucije od  $(Z_1, \dots, Z_k)$  uz uvjet  $Z_0 = t$ , što je prema prethodnoj lemi jednako distribuciji uređajne statistike  $k$ -tog reda uzorka  $(Z_1^*, \dots, Z_k^*)$  sa zajedničkom distribucijom

$$F_t(t+x) = (F(t+x) - F(t))/(1 - F(t)), x > 0. \quad (2.16)$$

Kako je poredak nebitan u funkciji vjerodostojnosti i kako su po pretpostavci  $X_i$  nezavisne, slijedi da su i  $Z_i^*$  nezavisne. Time dobivamo  $k$  nezavisnih jednako distribuiranih slučajnih varijabla s funkcijom distribucije jednakom (2.16). Primjenom metode maksimalne vjerodostojnosti na Pareto distribuciju, procjenitelj metodom maksimalne vjerodostojnosti je dobiven maksimizacijom aproksimativne funkcije vjerodostojnosti  $L(\gamma, \sigma, z_1, \dots, z_k) = \prod_{i=1}^k h_{\gamma, \sigma}(z_i)$  s obzirom na  $\gamma$  i  $\sigma$  pri čemu je  $z_i = x_{n-i+1,n} - x_{n-k,n}$  i  $h_{\gamma, \sigma} = \partial H_\gamma(x/\sigma)/\partial x$ .

Primijetimo da ova aproksimativna uvjetna funkcija teži k  $\infty$  za  $\gamma < -1$  i za  $\gamma/\sigma \downarrow -(X_{n,n} - X_{n-k,n})^{-1}$ , stoga maksimum po svim mogućim vrijednostima od  $(\gamma, \sigma)$  ne postoji. Koncentrirajmo se na područje  $(\gamma, \sigma) \in (-1/2, \infty) \times (0, \infty)$ . Jednadžbe vjerodostojnosti su

$$\begin{cases} \frac{\partial \log h_{\gamma, \sigma}(z)}{\partial \gamma} = \frac{1}{\gamma^2} \log \left( 1 + \frac{\gamma}{\sigma} z \right) - \left( \frac{1}{\gamma} + 1 \right) \frac{\frac{z}{\sigma}}{1 + \frac{\gamma}{\sigma} z}, \\ \frac{\partial \log h_{\gamma, \sigma}(z)}{\partial \sigma} = -\frac{1}{\sigma} - \left( \frac{1}{\gamma} + 1 \right) \frac{-\frac{\gamma}{\sigma^2} z}{1 + \frac{\gamma}{\sigma} z}. \end{cases}$$

Završne jednadžbe u terminima  $X_{n-i+1,n} - X_{n-k,n}$  mogu se pojednostaviti do

$$\begin{cases} \frac{1}{k} \sum_{i=1}^k \log \left( 1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n}) \right) = \gamma, \\ \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})} = \frac{1}{\gamma+1}. \end{cases} \quad (2.17)$$

**Teorem 2.3.1.** *Neka su  $X_1, X_2, \dots$  nezavisne jednako distribuirane slučajne varijable s funkcijom distribucije  $F$ . Pretpostavimo da funkcija  $F$  zadovoljava uvjet drugog reda uz  $\gamma > -1/2$ . Tada za  $k = k(n) \rightarrow \infty, k/n \rightarrow 0$  i*

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda$$

za konačan  $\lambda$ , sustav (2.17) ima niz rješenja  $(\hat{\gamma}_{MLE}, \hat{\sigma}_{MLE})$  koja zadovoljavaju

$$\sqrt{k} \left( \hat{\gamma}_{MLE} - \gamma, \frac{\hat{\sigma}_{MLE}}{a\left(\frac{n}{k}\right)} - 1 \right) \xrightarrow{d} N(\lambda b_{\gamma, \rho}, \Sigma),$$

pri čemu  $N$  označava standardnu normalnu distribuciju,

$$b_{\gamma, \rho} := \begin{cases} \left( \frac{\gamma+1}{(1-\rho)(1+\gamma-\rho)}, \frac{-\rho}{(1-\rho)(1+\gamma-\rho)} \right), & \rho < 0 \\ (1, 0) & , \rho = 0, \end{cases}$$

a matrica  $\Sigma$  je dana s

$$\begin{pmatrix} (1+\gamma)^2 & -(1+\gamma) \\ -(1+\gamma) & 1+(1+\gamma)^2 \end{pmatrix}.$$

*Dokaz:* Vidi u [2] dokaz teorema 3.4.2., str. 92.

## 2.4 Procjenitelj metodom momenata (MOM)

U poglavlju 2.2 gdje smo se upoznali s Hillovim procjeniteljem zaključili smo da je definiran samo za  $\gamma > 0$ . Kako Hillov procjenitelj ima široku upotrebu, željeli bismo razviti njemu sličan procjenitelj, ali takav da se može koristiti ne samo za  $\gamma > 0$  već za svaki  $\gamma \in \mathbf{R}$ .

**Lema 2.** *Neka su  $X_1, X_2, \dots$  nezavisne jednako distribuirane slučajne varijable s funkcijom distribucije  $F$  i neka je  $F \in \mathcal{D}(G_\gamma)$ ,  $x^* = U(\infty) > 0$  tj. za  $x > 0$*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}.$$

Definirajmo za  $j = 1, 2$ ,

$$M_n^{(j)} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j. \quad (2.18)$$

Tada za  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ ,  $n \rightarrow \infty$ ,

$$\frac{M_n^{(j)}}{\left(a \binom{n}{k} / U \binom{n}{k}\right)^j} \xrightarrow{P} \prod_{i=1}^j \frac{i}{1 - i\gamma_-}$$

gdje je  $\gamma_- = \min(0, \gamma)$ .

*Dokaz:* Vidi u [2] dokaz teorema 3.5.1., str. 101.

Ova lema nam pomaže da pronađemo konzistentan procjenitelj za  $\gamma$  kada je  $\gamma < 0$  jer pod tim uvjetima

$$\frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} \xrightarrow{P} \frac{1 - 2\gamma_-}{2(1 - \gamma_-)} . \quad (2.19)$$

Kao što smo i spomenuli, vrijedi

$$\hat{\gamma}_H \xrightarrow{P} \gamma_+ .$$

Kombinacijom Hillovog procjenitelja i statistike iz (2.18) dobivamo iduću relaciju kojom definiramo procjenitelja za  $\gamma \in \mathbf{R}$ :

$$\hat{\gamma}_M := M_n^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} \right)^{-1} . \quad (2.20)$$

**Teorem 2.4.1.** *Neka su  $X_1, X_2, \dots$  nezavisne jednako distribuirane slučajne varijable s funkcijom distribucije  $F$ . Pretpostavimo da je  $F \in \mathcal{D}(G_\gamma)$  te da je  $x^* > 0$ . Tada*

$$\hat{\gamma}_M \xrightarrow{P} \gamma$$

za  $\gamma \in \mathbf{R}$ . Slijedi da je  $\hat{\gamma}_M$  konzistentan procjenitelj za  $\gamma$ .

**Napomena 3.** Procjenitelj  $\hat{\gamma}_M$  iz relacije (2.20) naziva se procjenitelj metodom momenata.

**Napomena 4.** U idućem teoremu pokazat ćemo uz koje je uvjete procjenitelj metodom momenata asimptotski normalan. Pritom ćemo koristiti funkciju  $Q$  konstantnog predznaka

$$Q(t) := \begin{cases} A(t), & \gamma < \rho \leq 0 \\ \gamma_+ - \frac{a(t)}{U(t)}, & \rho < \gamma \leq 0 \text{ ili } \gamma = -\rho, \\ \frac{\rho}{\gamma + \rho} A(t), & \gamma > -\rho > 0, \\ A(t), & \gamma > \rho = 0. \end{cases}$$

**Teorem 2.4.2.** (Dekkers, Einmahl, de Haan (1989)) Neka su  $X_1, X_2, \dots$  jednako distribuirane slučajne varijable s funkcijom distribucije  $F$ . Pretpostavimo da vrijedi uvjet drugog reda definiran relacijom (2.8). Ako vrijedi

$$\lim_{n \rightarrow \infty} \sqrt{k} Q\left(\frac{n}{k}\right) = \lambda$$

gdje je  $Q$  funkcija definirana u Napomeni 4 te  $\lambda$  konačna, tada

$$\sqrt{k}(\hat{\gamma}_M - \gamma) \rightarrow^d N(\lambda b_{\gamma, \rho}, \text{var}_\gamma) \quad (2.21)$$

gdje je  $N$  standardna normalna distribucija,

$$b_{\gamma, \rho} := \begin{cases} \frac{(1-\gamma)(1-2\gamma)}{(1-\gamma-\rho)(1-2\gamma-\rho)}, & \gamma < \rho \leq 0, \\ \frac{\gamma(1+\gamma)}{(1-\gamma)(1-3\gamma)}, & \rho < \gamma \leq 0, \\ -\frac{\gamma}{(1+\gamma)^2}, & 0 < \gamma < -\rho, \\ \frac{\gamma-\rho+\rho}{\rho(1-\rho)}, & 0 < \gamma < -\rho \text{ ili } \gamma \geq -\rho > 0, \\ 1, & \gamma > \rho = 0, \end{cases}$$

$$\text{var}_\gamma := \begin{cases} \gamma^2 + 1, & \gamma \geq 0 \\ \frac{(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)}, & \gamma < 0. \end{cases}$$

*Dokaz:* Vidi u [2] dokaz teorema 3.5.4., str. 104.

## 2.5 Asimptotska svojstva procjenitelja indeksa

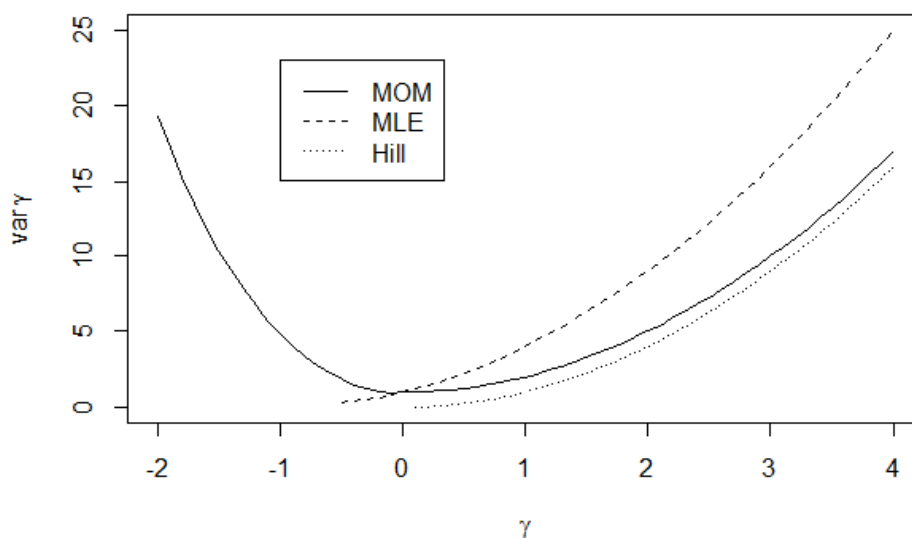
Do sada smo spomenuli nekoliko procjenitelja indeksa ekstremnih vrijednosti, Hillov procjenitelj ( $\hat{\gamma}_H$ ), procjenitelj metodom momenata (MOM) ( $\hat{\gamma}_M$ ) i procjenitelj metodom maksimalne vjerodostojnosti (MLE) ( $\hat{\gamma}_{MLE}$ ). Prisjetimo se kako je Hillov procjenitelj definiran samo za  $\gamma > 0$ , MOM za  $\gamma \in \mathbf{R}$ , a MLE za  $\gamma > -\frac{1}{2}$ .

U prethodnim poglavljima smo vidjeli kako su uz određene uvjete sva tri procjenitelja asimptotski normalna, što znači da za  $n$  nezavisnih i jednako distribuiranih slučajnih varijabla uz konstantu  $k$  vrijedi

$$\sqrt{k}(\hat{\gamma} - \gamma) \approx \sqrt{\text{var}_\gamma} N + \lambda b_{\gamma, \rho}$$

gdje je  $N$  standardna normalna distribucija, a konstante  $\lambda$ ,  $\text{var}_\gamma$ ,  $b_{\gamma, \rho}$  poznate (vidi Teoreme 2.2.1, 2.3.1 i 2.4.2).

Za asimptotsku normalnost procjenitelja MLE zahtjevamo uvjet drugog reda definiran relacijom (2.7) uz pomoćnu funkciju drugog reda  $A$  i uz  $k$  koji zadovoljava  $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbf{R}$ . Za asimptotsku normalnost Hillovog procjenitelja zahtjevamo uvjet drugog reda definiran relacijom (2.13) uz pomoćnu funkciju drugog reda  $A$ , a za MOM uvjet drugog reda iz relacije (2.8) gdje je  $\rho \neq \gamma$ , uz pomoćnu funkciju  $Q$  definiranu u Napomeni 4 i  $k$  takav da  $\sqrt{k}Q(n/k) \rightarrow \lambda \in \mathbf{R}$ .



Slika 2.1: Asimptotska varijabilnost

Na Slici 2.1 usporedili smo asimptotsku varijabilnost procjenitelja metodom momenta, procjenitelja metodom maksimalne vjerodnostojnosti i Hillovog procjenitelja. Uočimo kako Hillov procjenitelj ima najmanju asimptotsku varijabilnost na skupu  $\gamma \in (0, \infty)$ . Za pozitivne vrijednosti  $\gamma$ , MLE ima najveću asimptotsku varijabilnost, a Hillov procjenitelj najmanju. Negativne vrijednosti od  $\gamma$  procjenjuje MOM stoga on ima najmanju varijabilnost za  $\gamma < 0$ . Također možemo uočiti kako se MLE i MOM ne razlikuju značajno kada je  $\gamma$  blizu nule.

## Poglavlje 3

# Procjena ekstremnih kvantila i repa razdiobe

### 3.1 Procjena funkcije skale

Neka je  $F$  u domeni atrakcije distribucije ekstremnih vrijednosti tj. neka za  $a_n > 0$  i  $b_n$  te  $\gamma \in \mathbf{R}$  vrijedi

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), 1 + \gamma x > 0. \quad (3.1)$$

Tada za neku pozitivnu funkciju  $a$  i za funkciju  $b(t) = U(t) = F^{\leftarrow}(1 - 1/t)$  vrijedi

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \quad (3.2)$$

za  $x > 0$ . Nadalje pretpostavimo da za  $x > 0$ ,  $\rho \leq 0$  i funkciju  $A$  konstantnog predznaka takvu da  $A(t) \rightarrow 0$ ,  $t \rightarrow \infty$  vrijedi uvjet drugog reda,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \frac{1}{\rho} \left( \frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right). \quad (3.3)$$

Sada možemo definirati procjenitelj za funkciju skale  $a$  vezanog uz moment procjenitelja od  $\gamma$ ,  $\hat{\gamma}_M$ . Definirajmo prvo

$$\hat{\gamma}_- := 1 - \frac{1}{2} \left( 1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} \right)^{-1}, \quad (3.4)$$

gdje je  $M_n^{(j)}$  za  $j = 1, 2$  definiran relacijom (2.18). Sada je procjenitelj za funkciju skale jednak



$$\hat{a}\left(\frac{n}{k}\right) = \hat{\sigma}_M := X_{n-k,n} M_n^{(1)}(1 - \hat{\gamma}_-). \quad (3.5)$$

## 3.2 Procjena ekstremnih kvantila

Kao i kod procjene funkcije skale u potpoglavlju (3.1) uzmimo relacije (3.1)–(3.3) kao polazne pretpostavke. Pretpostavimo da je za odgovarajuće procjenitelje  $\hat{\gamma}$ ,  $\hat{a}(n/k)$  i  $\hat{b}(n/k)$ ,

$$\sqrt{k} \left( \hat{\gamma} - \gamma, \frac{\hat{a}\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} - 1, \frac{\hat{b}\left(\frac{n}{k}\right) - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} \right) \rightarrow (\Gamma, \Lambda, B), \quad n \rightarrow \infty \quad (3.6)$$

gdje su  $(\Gamma, \Lambda, B)$  normalne slučajne varijable. Sada možemo razmatrati procjenu ekstremnih kvantila. Neka je  $x_p := U(1/p)$  kvantil koji želimo procijeniti. Slučajevi koji nas naročito zanimaju su oni gdje je očekivani broj opažanja iznad  $x_p$ , tj.  $np$  mali broj. To znači da želimo ekstrapolirati kvantile i izvan opsega dostupnih opažanja. Stoga pretpostavljamo da  $p$  ovisi o  $n$ , tj.  $p = p_n$ , i da

$$\lim_{n \rightarrow \infty} p_n = 0.$$

Želimo  $x_{p_n}$  procijeniti pomoću relacije  $1 - F(x_{p_n}) = p_n$  ili ekvivalentno iz  $x_{p_n} = U(1/p_n)$  s  $p_n \rightarrow 0$  dok  $n \rightarrow \infty$ .

**Teorem 3.2.1.** *Pretpostavimo da za neku funkciju  $A$  takvu da  $A(t) \rightarrow 0$ ,  $t \rightarrow \infty$  vrijedi uvjet drugog reda (3.3). Pretpostavimo nadalje:*

1. parametar drugog reda  $\rho$  je negativan ili jednak 0, a  $\gamma$  je negativan;
2.  $k = k(n) \rightarrow \infty$ ,  $n/k \rightarrow \infty$  i  $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbf{R}$ ,  $n \rightarrow \infty$ ;
3. uvjet (3.6) vrijedi za odgovarajuće procjenitelje od  $\gamma$ ,  $a(n/k)$  i  $U(n/k)$ ;
4.  $np_n = o(k)$  i  $\log(np_n) = o(\sqrt{k})$ ,  $n \rightarrow \infty$ .

Definirajmo

$$\hat{x}_{p_n} := \hat{b}\left(\frac{n}{k}\right) + \hat{a}\left(\frac{n}{k}\right) \frac{\left(\frac{k}{np_n}\right)^{\hat{\gamma}} - 1}{\hat{\gamma}} \quad \text{i} \quad x_{p_n} := U\left(\frac{1}{p_n}\right). \quad (3.7)$$

Tada, kako  $n \rightarrow \infty$  vrijedi,

$$\sqrt{k} \frac{\hat{x}_{p_n} - x_{p_n}}{a\left(\frac{n}{k}\right) q_\gamma(d_n)} \rightarrow \Gamma + (\gamma_-)^2 B - \gamma_- \Gamma - \lambda \frac{\gamma_-}{\gamma_- + \rho}$$

gdje su  $\Gamma$  i  $B$  normalno distribuirane slučajne varijable,  $d_n := k/(np_n)$ ,  $\gamma_- := \min(0, \gamma)$  i gdje za  $t > 1$  vrijedi

$$q_\gamma(t) := \int_1^t s^{\gamma-1} \log s \, ds.$$

*Dokaz:* Vidi u [2] dokaz teorema 4.3.1., str. 135.

**Napomena 5.** Kod procjene kvantila koristimo se procjeniteljem za funkciju skale  $\hat{a}$  te procjeniteljem indeksa ekstremne vrijednosti  $\hat{\gamma}$  koje smo naveli u prethodnim poglavljima, kao i procjeniteljem lokacijske funkcije  $b$  koju do sada nismo spomenuli. Za procjenitelj funkcije  $b_{n/k} := U(n/k)$  uzima se njegov empirijski analog tj.  $\hat{b} := \hat{b}_{n/k} := X_{n-k,n}$ .

**Napomena 6.** Uvjet  $np_n = o(k)$  iz Teorema 3.2.1. znači da  $\lim_{n \rightarrow \infty} np_n/k = 0$  tj. da  $k$  raste puno brže od  $np_n$ . Posebno, kada  $np_n \rightarrow 0$  uvjet  $\log(np_n) = o(\sqrt{k})$  nam zapravo govori da ne smijemo pretjerati s ekstrapoliranjem.

### 3.3 Procjena repa

Neka je  $F$  u domeni atrakcije distribucije ekstremnih vrijednosti. Za velike vrijednosti  $x$  želimo procijeniti

$$p = 1 - F(x).$$

Kao kod procjene funkcije skale i kvantila neka su relacije (3.1)-(3.3) polazne pretpostavke. Ovdje posebno,  $x$  ovisi o  $n$ , pa je  $x = x_n$  te  $p_n = 1 - F(x_n) \rightarrow 0$ . Definirajmo dodatno:

$$d_n := \frac{k}{n(1-F(x_n))}.$$

Za procjenu repa koristeći relaciju (2.4) dobivamo

$$\hat{p}_n := \frac{k}{n} \left( \max \left( 0, 1 + \hat{\gamma} \frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})} \right) \right)^{-1/\hat{\gamma}}. \quad (3.8)$$

uz poznati  $x_n$ .

### 3.4 Procjena desne krajnje točke razdiobe

Pretpostavimo da je  $F \in \mathcal{D}(G_\gamma)$  za neki negativan  $\gamma$ . U tom je slučaju desna krajnja točka  $x^*$  sigurno konačna. Sada, ako u relaciji (3.7) postavimo  $p_n = 0$  dobijemo

## 22 POGLAVLJE 3. PROCJENA EKSTREMNIH KVANTILA I REPA RAZDIOBE

procjenitelj desne krajnje točke

$$\hat{x}^* := \hat{b} \left( \frac{n}{k} \right) - \frac{\hat{a} \left( \frac{n}{k} \right)}{\hat{\gamma}}. \quad (3.9)$$

## Poglavlje 4

# Primjena teorije ekstremnih vrijednosti na rekorde u atletici

Primjenom teorije ekstremnih vrijednosti na svjetske rekorde u atletici pokušat ćemo odgovoriti na dva pitanja: koji je ultimativan tj. konačan svjetski rekord u pojedinoj atletskoj disciplini te koliko je "dobar" trenutni rekord pojedine discipline, odnosno koliko je teško poboljšati ga?

U gotovo svim dosadašnjim pokušajima predviđanja ultimativnih svjetskih rekorda promatralo se poboljšanje rekorda kroz vrijeme. U ovom radu, gore navedenim problemima pristupit ćemo na drugačiji način. Koristit ćemo *sva* najbolja sportska postignuća u pojedinoj disciplini tj. osobne rekorde najboljih svjetskih atletičara. Tako dobivena procjena ultimativnog rekorda reći će nam što je moguće postići u bliznoj budućnosti s obzirom na današnje fizičke sposobnosti, znanje i sportsku opremu te regulativu o zabranjenim supstancama u sportu.

Odgovore na gornja pitanja pokušat ćemo dati za 10 atletskih disciplina: 5 trkačkih, 2 skakačke i 3 bacačke discipline, a to su, trčanje na 100 metara, trčanje na 400 metara, trčanje na 800 metara, trčanje na 1500 metara, maraton, skok u dalj, skok u vis te bacanje kladiva, bacanje koplja i bacanje diska.

### 4.1 Podaci

Za svaku od 20 disciplina (10 za muškarce i 10 za žene) prikupili smo podatke o osobnim rekordima najboljih atletičara. Pritom smo brinuli da lista prikupljenih podataka bude što potpunija. Točnije, ako se pojedini atletičar pojavljuje na listi, tada

se *svi* atletičari s boljim osobnim rekordom također moraju pojaviti na listi.

Kako smo napomenuli, naglasak je na najboljim osobnim rekordima, a ne na razvoju rekorda kroz vrijeme. Zbog toga se svaki atletičar na našoj listi pojavljuje točno jednom čak i ako je oborio svjetski rekord nekoliko puta.

Podatke smo prikupili sa službene web stranice Međunarodnog atletskog saveza (International Association of Athletics Federation (IAAF)) te smo promatrali sve zabilježene najbolje osobne rezultate pojedinih disciplina zaključno s krajem 2017. godine. Na stranicama IAAF-e po pojedinoj disciplini određena je gornja (trkačke discipline) odnosno donja granica (skakačke i bacačke discipline) za uvrštavanje pojedinog atletičara na listu najboljih svjetskih rekorda. Svi rekordi koje smo uzeli u obzir postignuti su na natjecanjima s elektronskim mjerenjem vremena.

Baza podataka sastoji se od približno 16000 podataka, pri čemu oko 9000 podataka čine postignuća atletičara te oko 7000 podataka postignuća atletičarki. U Tablici 4.1 dat ćemo kratak pregled baze podataka s brojem zabilježenih rekorda po disciplini te s najboljim odnosno najslabijim rezultatom.

Disciplina	Muškarci			Žene		
	Broj podataka	Najbolji rezultat	Najslabiji rezultat	Broj podataka	Najbolji rezultat	Najslabiji rezultat
Trčanje na 100 m	1757	9.58	10.30	1649	10.49	11.50
Trčanje na 400 m	697	43.03	45.50	729	47.60	52.00
Trčanje na 800 m	647	1:40.91	1:46.00	920	1:53.28	2:02.00
Trčanje na 1500 m	912	3:26.00	3:38.00	856	3:50.07	4:10.00
Maraton	1806	2:02:57	2:12:00	794	2:15:25	2:30:00
Skok u vis	769	2.45	2.25	198	2.09	1.95
Skok u dalj	856	8.95	8.00	285	7.52	6.75
Bacanje diska	577	74.08	62.00	133	76.80	65.00
Bacanje kladiva	455	86.74	74.01	924	82.98	60.00
Bacanje koplja	359	98.48	80.00	496	72.28	55.04

Tablica 4.1: Sažetak podataka

Sve udaljenosti u skakačkim i bacačkim disciplinama iskazane su u metrima pa je veći rezultat ujedno i bolji. Sva vremena u trkačkim disciplinama su iskazana u sekundama, i što je manje sekunda to je rezultat bolji. Kako bismo uravnotežili podatke svih triju disciplina, vremena trčanja u sekundama smo transformirali u prosječnu brzinu. Na taj način dobili smo veličinu izraženu u jedinici km/h. Jasno, što je veća brzina bolji je rezultat.

Kod pojedinih disciplina, pogotovo trkačkih na kraćim relacijama, pojavljuje se veliki broj istih rezultata. Na primjer, 137 muških atletičara istrčalo je 100 metara za 10.30 sekundi, njih 110 za 10.29 sekundi itd. To ne mora značiti da su pravi rezultati zbilja isti, već da mjerenje vremena nije savršeno. Veliki broj ponavljanja može uzrokovati probleme u procjenama stoga smo ih "izgladili". Na primjer, pretpostavimo da  $m$  atletičara ima isti osobni rekord u trčanju na 100 metara od 10.30 sekundi. Tada smo tih  $m$  rezultata razvukli duž intervala (10.295, 10.305) formulom

$$d_j = 10.295 + 0.01 \frac{2j - 1}{2m}, \quad j = 1, \dots, m .$$

## 4.2 Procjena indeksa ekstremne vrijednosti

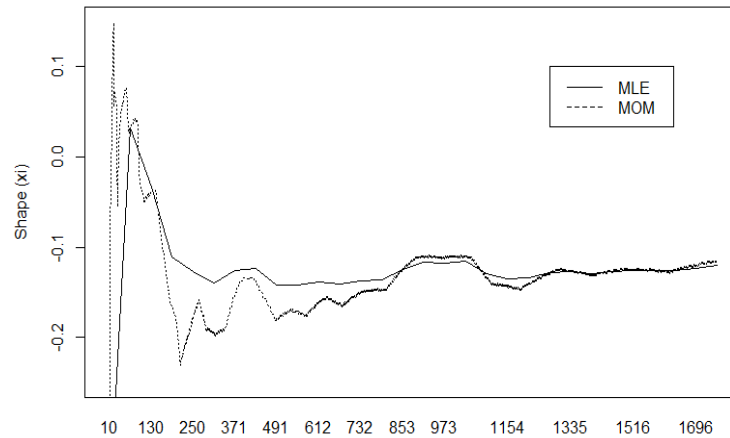
Primjenom procjenitelja definiranih u prethodnim poglavljima na podatke opisane u poglavlju 4.1 pokušat ćemo odgovoriti na pitanja: koji je ultimativan svjetski rekord te koliko je dobar trenutni rekord?

Uzmimo u obzir jednu atletsku disciplinu, npr. trčanje na 100 metara kod muškaraca te s  $X_1, X_2, \dots, X_n$  označimo  $n$  osobnih rekorda *svih* atletičara u toj disciplini. Kako je točnu definiciju "atletičara" teško odrediti, tako je i mjerenje točnog broja  $n$  težak zadatak. Naravno, broj  $n$  mnogo je veći od broja podataka u pojedinoj disciplini (točnije, 1757 kod trčanja na 100 m) jer smo našu bazu podataka napravili od osobnih rekorda *najboljih* atletičara. No u našim procjenama točan broj  $n$  neće nam biti važan.

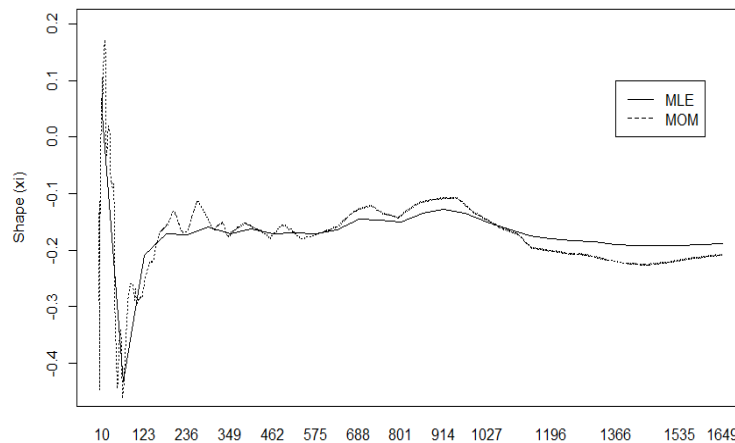
Sada, neka je  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  uređajna statistika pri čemu  $X_{n,n}$  označava svjetski rekord u toj disciplini. Prisjetimo se kako smo vrijeme trčanja transformirali u brzinu, stoga vrijedi što je veća brzina to je rezultat bolji, analogno što je veća visina skoka ili veća udaljenost bacanja to je rezultat bolji.

Prvi korak u našim procjenama je procjena indeksa ekstremlne vrijednosti za 10 odabranih atletske discipline, posebno za muškarce i za žene. Procjenitelj indeksa razdiobe

ekstremne vrijednosti nacrtat ćemo kao funkciju od  $k$  pri čemu je  $1 \leq k < l$ , a  $l$  označava broj osobnih rekorda najboljih atletičara u toj disciplini tj. broj podataka pojedine discipline iz Tablice 4.1.



Slika 4.1: MLE i MOM procjenitelj indeksa razdiobe ekstremne vrijednosti za disciplinu trčanje na 100 m (muškarci)



Slika 4.2: MLE i MOM procjenitelj indeksa razdiobe ekstremne vrijednosti za disciplinu trčanje na 100 m (žene)

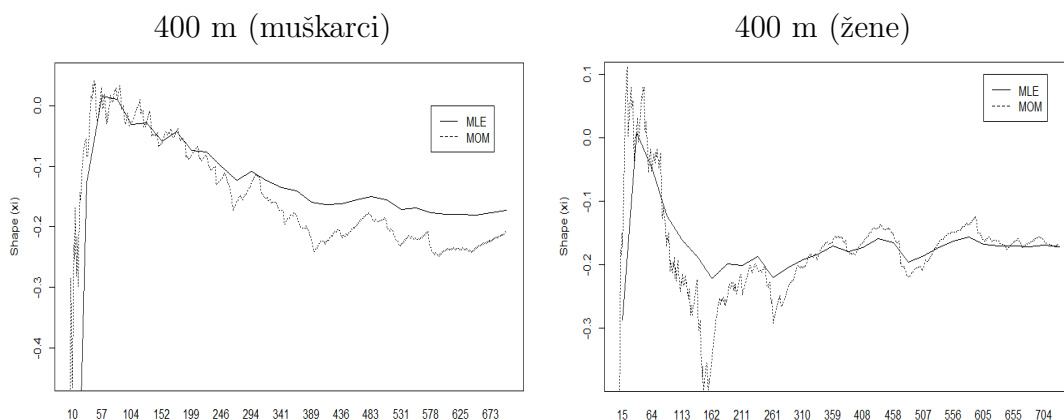
Na Slikama 4.1 i 4.2 prikazali smo procjenitelje metodom maksimalne vjerodostojnosti ( $\hat{\gamma}_{MLE}$ ) te procjenitelje metodom momenata ( $\hat{\gamma}_M$ ) kao funkcije od  $k$ , pri čemu je kod

procjene indeksa razdiobe ekstremne vrijednosti za trčanje muškaraca,  $1 \leq k \leq 1756$ , a kod žena,  $1 \leq k \leq 1648$ .

Graf funkcije procjenitelja metodom momenata dobili smo samostalnom implementacijom formula (2.18) i (2.20) u programskom jeziku *R*. Nadalje, kako ne postoji eksplicitna formula za maksimizaciju funkcije vjerodostojnosti, za dobivanje MLE procjenitelja potrebne su odgovarajuće numeričke metode. U *R* – *u* postoji nekoliko paketa koja sadrže razne statističke procedure iz teoriju ekstremnih vrijednosti. Graf funkcije MLE procjenitelja nacrtali smo koristeći gotovu naredbu u *R* – *u*.

Uočimo na slikama kako je za male vrijednosti  $k$  varijabilnost oba procjenitelja veća te kako se povećanjem  $k$  varijabilnost značajno smanjuje. Također vidimo kako se oba procjenitelja ponašaju dosta slično, s manjim odstupanja. Za jako male vrijednosti od  $k$  indeks ekstremne vrijednosti je pozitivan, ali povećanjem vrijednosti od  $k$ ,  $\gamma$  je manji od nule.

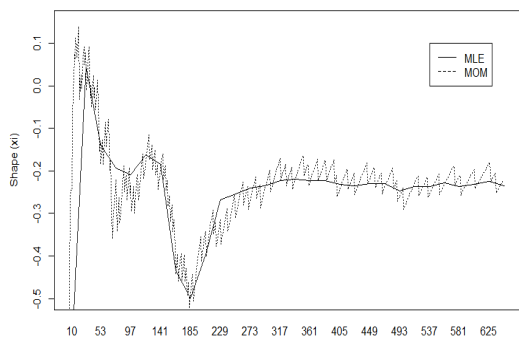
Crtaњem preostalih grafova možemo zaključiti kako je za sve discipline gotovo uvijek  $\gamma < 0$ . Također vidimo kako zaključci doneseni na temelju prva dva grafa, vrijede i za ostale.



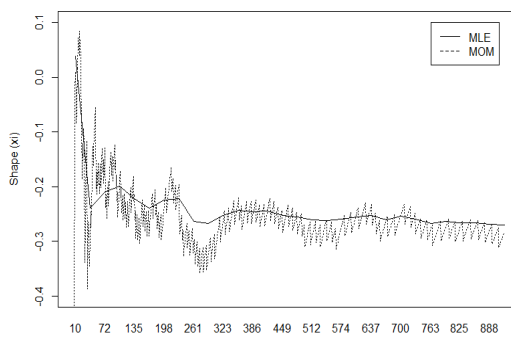


## POGLAVLJE 4. PRIMJENA TEORIJE EKSTREMNIH VRIJEDNOSTI NA REKORDE U ATLETICI

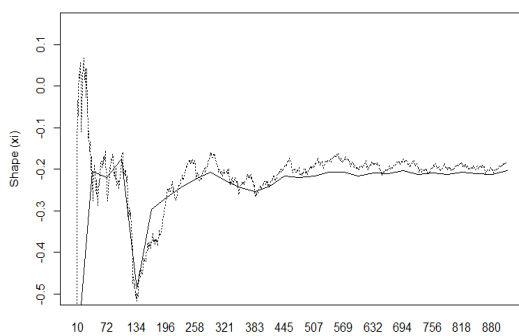
800 m (muškarci)



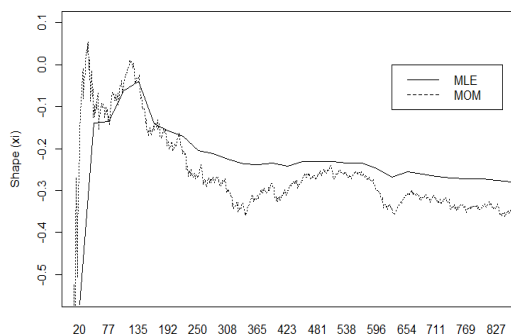
800 m (žene)



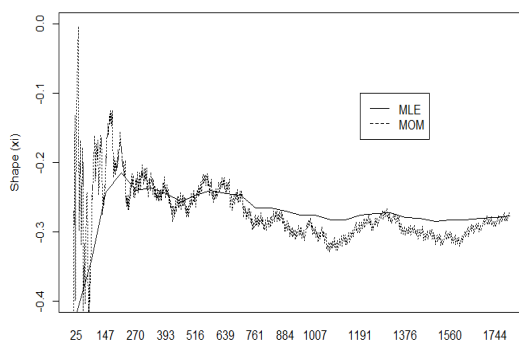
1500 m (muškarci)



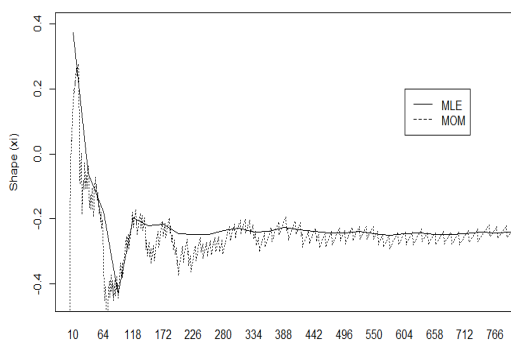
1500 m (žene)



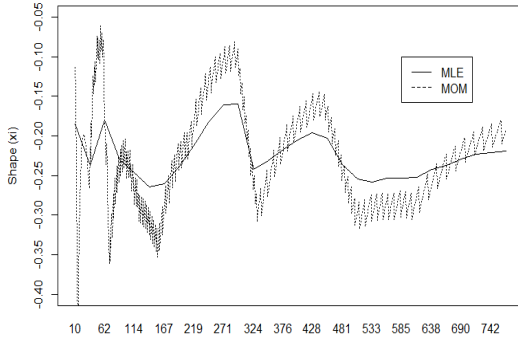
Maraton (muškarci)



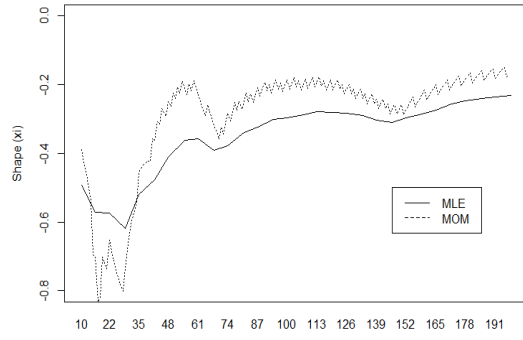
Maraton (žene)



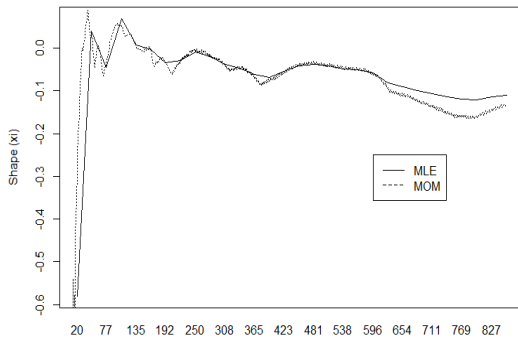
Skok u vis (muškarci)



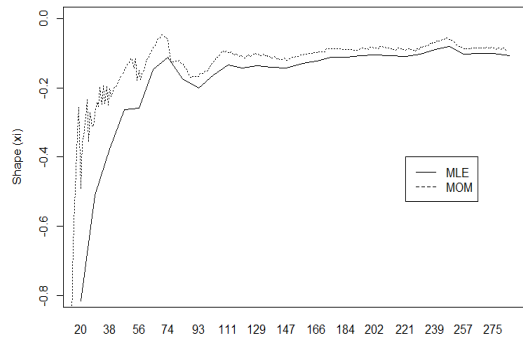
Skok u vis (žene)



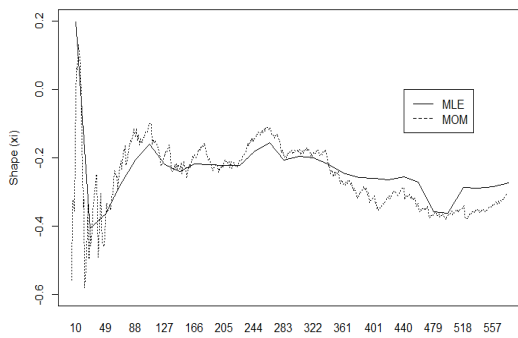
Skok u dalj (muškarci)



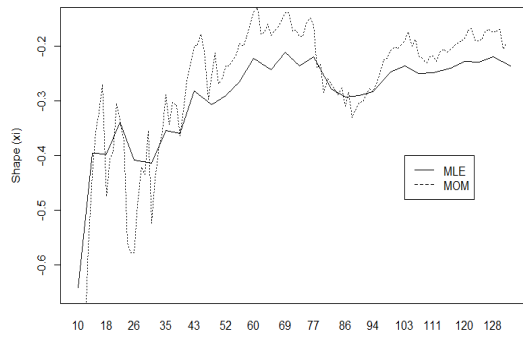
Skok u dalj (žene)

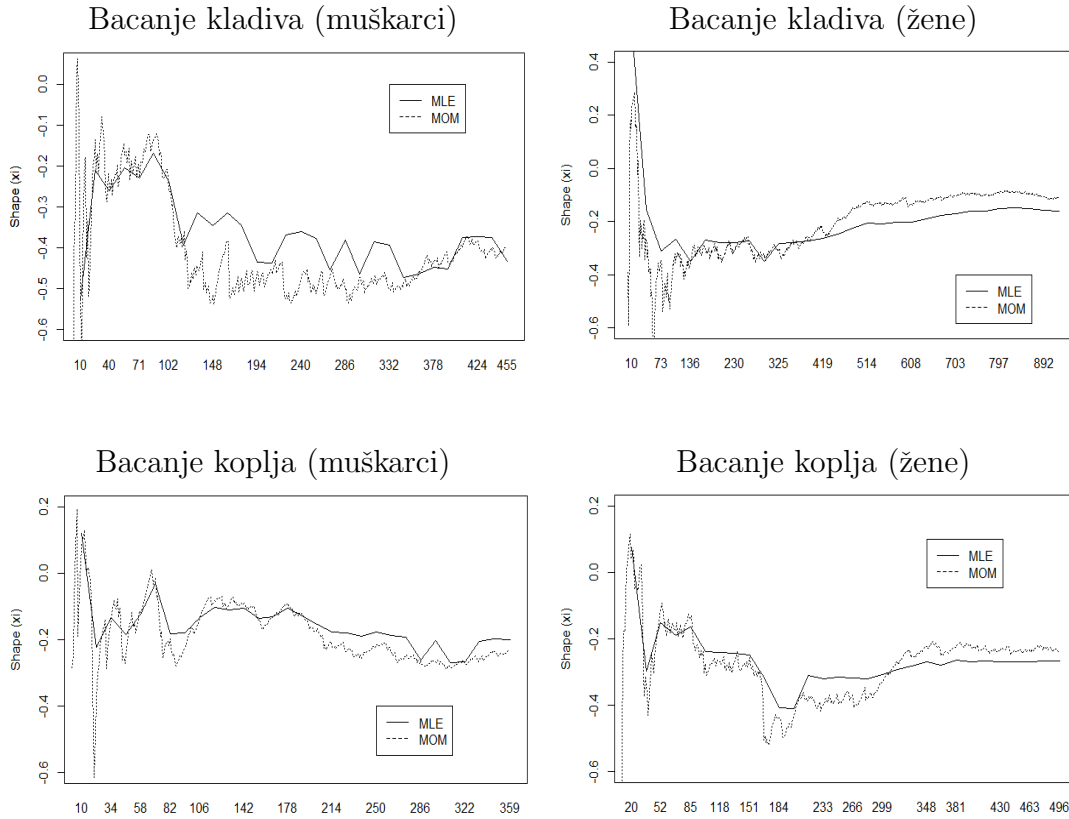


Bacanje diska (muškarci)



Bacanje diska (žene)





Tablica 4.2: MLE i MOM procjenitelji indeksa razdiobe ekstremne vrijednosti za preostale atletske discipline (kod muškaraca i žena)

Na temelju dobivenih grafova na Slikama 4.1 i 4.2, ali i na ostalim grafovima u Tablici 4.2 nije odmah očito koju ćemo procjenu indeksa ekstremne vrijednosti uzeti. Stoga ćemo uzeti u obzir još jedan dodatan procjenitelj koji ima dobra svojstva kada je  $\gamma < 0$ . Uočimo prvo kako je formula za  $\hat{\gamma}_H$  iz relacije (2.12) jednaka formuli za  $M_n^{(1)}$  iz relacije (2.18). Prisjetimo se sada relacije (2.20) koja nam daje formulu za računanje MOM procjenitelja

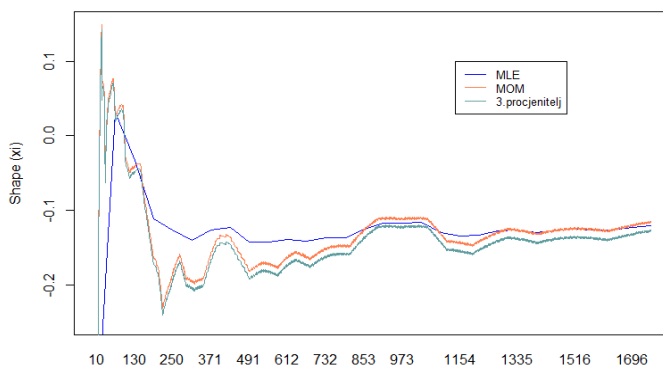
$$\hat{\gamma}_M := M_n^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}$$

te uočimo kako je "prvi dio" MOM procjenitelja upravo Hillov procjenitelj. Kako je Hillov procjenitelj definiran samo za  $\gamma > 0$ , koristit ćemo samo "drugi dio" MOM

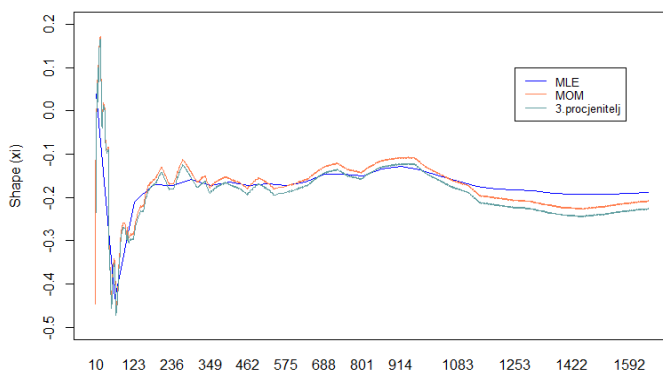
procjenitelja tj.

$$\hat{\gamma}_3 := 1 - \frac{1}{2} \left( 1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}.$$

Na Slikama 4.3 i 4.4 prikazat ćemo sva tri procjenitelja na istom grafu,  $\hat{\gamma}_{MLE}$ ,  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  za disciplinu trčanje na 100 metara kod muškaraca i kod žena.



Slika 4.3: Procjenitelji  $\hat{\gamma}_{MLE}$ ,  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  za disciplinu trčanje na 100 m (muškarci)

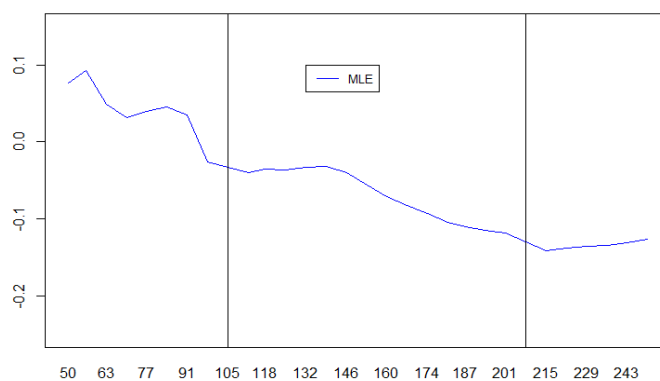


Slika 4.4: Procjenitelji  $\hat{\gamma}_{MLE}$ ,  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  za disciplinu trčanje na 100 m (žene)

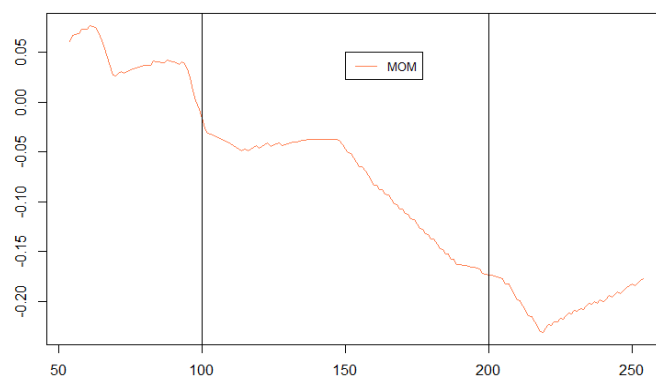
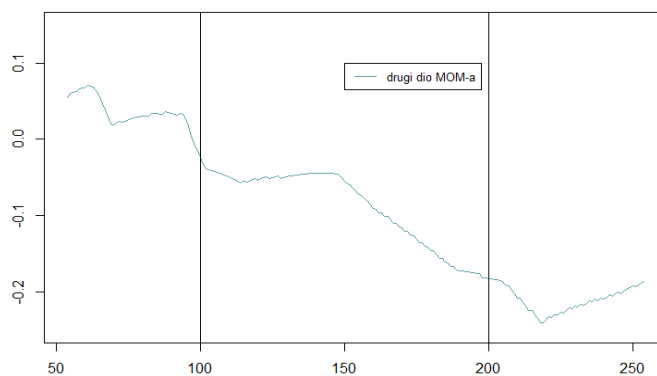
Za svaku disciplinu promatrali smo grafove tih triju procjenitelja te smo pokušali pronaći prvu stabilnu regiju od  $k$ . Točnije, za vrijednosti od  $k$  između 50 i 250 pokušali smo pronaći prvi skup od barem 50 uzastopnih  $k$  takvih da procjenjene vrijednosti indeksa za takve  $k$  ne osciliraju previše.

Općenito, pronalazak dobrih vrijednosti za  $k$  na kojima ćemo bazirati naše procjenitelje, problem je koji se često javlja u praksi. Metoda koju smo gore opisali i koju smo koristili prilikom procjene često je korištena u literaturi i u praksi i upravo je ta metoda korištena u [1]. U poglavlju 4.5 dodatno ćemo se osvrnuti na pitanje koliko je sam proces procjene indeksa razdiobe ekstremne vrijednosti osjetljiv na loše odabrane vrijednosti od  $k$ .

Procijenimo sada indekse ekstremnih vrijednosti svih disciplina. Postupak procjene detaljno ćemo prikazati na disciplini trčanje na 100 metara kod muškaraca i kod žena. Na Slikama 4.5, 4.6 i 4.7 prikazat ćemo zasebno procjenitelje  $\hat{\gamma}_{MLE}$ ,  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  za disciplinu trčanje na 100 m (muškarci) koncentrirajući se na vrijednosti  $50 \leq k \leq 250$  te ćemo vertikalnim crtama označiti donju, odnosno gornju granicu regije od  $k$  koju smo odredili kao prvu stabilnu regiju od  $k$ .



Slika 4.5: Procjenitelj  $\hat{\gamma}_{MLE}$  (100 m muškarci)

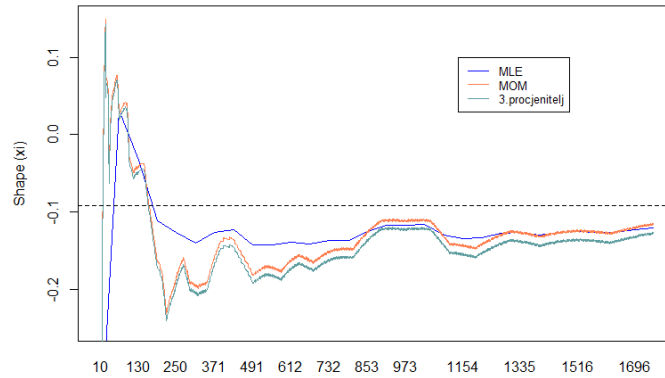
Slika 4.6: Procjenitelj  $\hat{\gamma}_M$  (100 m muškarci)Slika 4.7: Procjenitelj  $\hat{\gamma}_3$  (100 m muškarci)

Kod procjenitelja  $\hat{\gamma}_{MLE}$  takvu stabilnu regiju odabrali smo između  $k = 100$  i  $k = 210$ , a kod  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  procjenitelja između  $k = 100$  i  $k = 200$ . Nakon određivanja regija, odredili smo prosječne vrijednosti svakog pojedinog procjenitelja na toj regiji.

	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_M$	$\hat{\gamma}_3$
Prosječne vrijednosti	-0.086	-0.092	-0.097

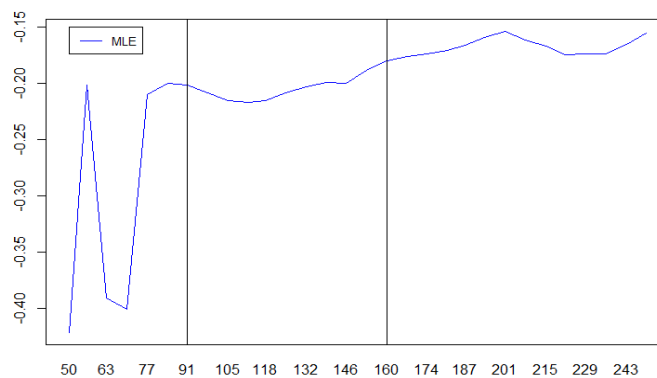
Tablica 4.3: Prosječne vrijednosti svih procjenitelja (100 m muškarci)

Prosjeck tih triju procjena daju nam konačnu procjenu indeksa razdiobe ekstremne vrijednosti za disciplinu trčanje na 100 metara kod muškaraca ( $\hat{\gamma} = -0.092$ ). Na Slici 4.8 horizontalnom iscrtkanom linijom nacrtali smo konačnu procjenu  $\hat{\gamma}$ .

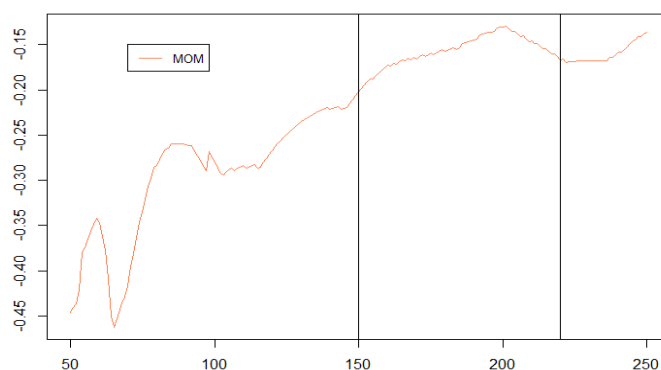
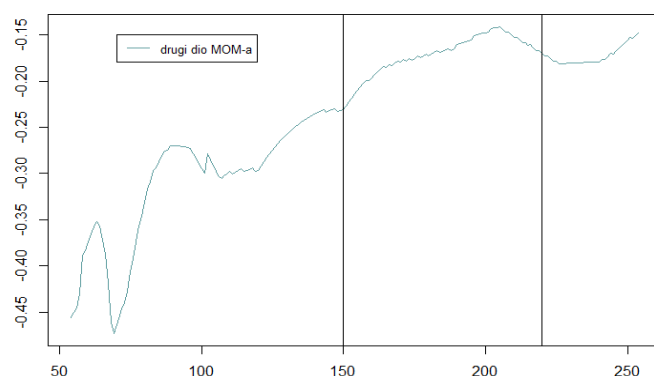


Slika 4.8: Procjenitelj  $\hat{\gamma}$  (100 m muškarci)

Na Slikama 4.9, 4.10 i 4.11 prikazat ćemo zasebno sva tri procjenitelja za disciplinu trčanje na 100 m kod žena koncentrirajući se na vrijednosti  $50 \leq k \leq 250$  te ćemo, kao i kod trčanja muškaraca na 100 metara, označiti stabilne regije i izračunati prosjeke tih regija kod svih procjenitelja.



Slika 4.9: Procjenitelj  $\hat{\gamma}_{MLE}$  (100 m žene)

Slika 4.10: Procjenitelj  $\hat{\gamma}_M$  (100 m žene)Slika 4.11: Procjenitelj  $\hat{\gamma}_3$  (100 m žene)

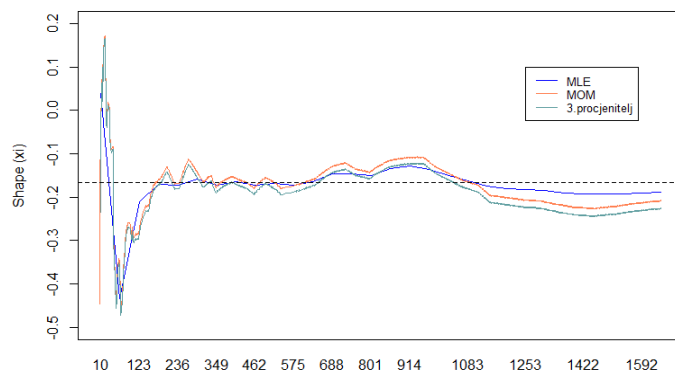
	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_M$	$\hat{\gamma}_3$
Prosječne vrijednosti	-0.182	-0.155	-0.169

Tablica 4.4: Prosječne vrijednosti svih procjenitelja (100 m žene)

Kod procjenitelja  $\hat{\gamma}_{MLE}$ , stabilnu regiju za  $k$  smo odabrali između  $k = 90$  i  $k = 180$ , a kod  $\hat{\gamma}_M$  i  $\hat{\gamma}_3$  između  $k = 150$  i  $k = 220$ .

U Tablici 4.4 prikazali smo dobivene rezultate iz kojih, računajući prosjek, konačno dobivamo procjenitelj indeksa ekstremne vrijednosti za disciplinu trčanje na 100 metara kod žena ( $\hat{\gamma} = -0.168$ ). Na Slici 4.12 horizontalnom iscertkanom linijom nacrtali smo konačnu procjenu od  $\gamma$ .





Slika 4.12: Procjenitelj  $\hat{\gamma}$  (100 m žene)

Na isti način procijenili smo indekse razdiobe ekstremne vrijednosti i za preostale discipline. Dobiveni rezultati prikazani su u Tablici 4.5. Vidimo kako su doista sve procijenjene vrijednosti od  $\gamma$  manje od 0, osim jedne pozitivne vrijednosti  $\hat{\gamma} = 0.012$  kod discipline skoka u dalj kod muškaraca.

Disciplina	Muškarci	Žene
Trčanje na 100 m	-0.092	-0.168
Trčanje na 400 m	-0.047	-0.160
Trčanje na 800 m	-0.226	-0.247
Trčanje na 1500 m	-0.224	-0.271
Maraton	-0.243	-0.190
Skok u vis	-0.231	-0.254
Skok u dalj	0.012	-0.117
Bacanje diska	-0.245	-0.149
Bacanje kladiva	-0.401	-0.332
Bacanje koplja	-0.162	-0.316

Tablica 4.5: Procijenjene vrijednosti  $\hat{\gamma}$

### 4.3 Procjena ultimativnog svjetskog rekorda pojedine discipline

Nakon procjene indeksa razdiobe ekstremne vrijednosti za sve discipline, iduće pitanje kojim ćemo se baviti je koji je ultimativan svjetski rekord pojedine discipline. Procjenom krajnje desne točke funkcije distribucije  $F$  tj.  $x^* := \sup\{x | F(x) < 1\}$  zapravo procjenjujemo ultimativan svjetski rekord.

Kod procjene krajnje desne točke pretpostavljamo da je  $\gamma < 0$  jer je za  $\gamma \geq 0$ ,  $x^* = \infty$  što smo i komentirali u poglavlju 1.2. Zbog te pretpostavke nećemo procjenjivati  $x^*$  kod disciplina kod kojih je  $\hat{\gamma}$  pozitivan ili jako blizu nuli. Takav slučaj pojavljuje se kod dvije discipline, trčanje na 400 m kod muškaraca i skok u dalj kod muškaraca. Relativno visoke vrijednosti od  $\hat{\gamma}$  ukazuju kako je moguće ostvariti značajan napredak trenutnog svjetskog rekorda.

Prisjetimo se relacije (3.9) pomoću koje smo dobili procjenitelj krajnje desne točke

$$\hat{x}^* := \hat{b}\left(\frac{n}{k}\right) - \frac{\hat{a}\left(\frac{n}{k}\right)}{\hat{\gamma}},$$

pri čemu je procjenitelj za funkciju skale  $a\left(\frac{n}{k}\right)$  dan relacijom (3.5)

$$\hat{a}\left(\frac{n}{k}\right) := X_{n-k,n} M_n^{(1)} (1 - \hat{\gamma}_-),$$

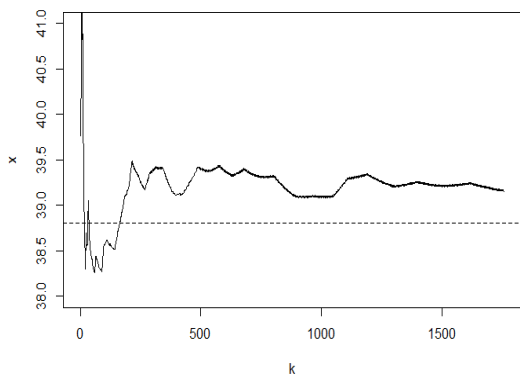
gdje je  $\hat{\gamma}_-$  jednak relaciji (3.4), a procjenitelj funkcije  $b\left(\frac{n}{k}\right)$  je  $\hat{b}\left(\frac{n}{k}\right) := X_{n-k,n}$ .

Sada ćemo, koristeći gornje relacije i dobivene podatke o procijenjenim vrijednostima od  $\gamma$  iz Tablice 4.5, nacrtati grafove funkcija procjenitelja krajnje desne točke kao funkcije od  $k$ . Kako koristimo fiksne procjenitelje od  $\gamma$ , ovisnost o  $k$  se očitava samo kroz  $X_{n-k,n}$  i  $M_n^{(1)}$ . Nadalje, koristeći istu metodu kao i kod procjene indeksa razdiobe ekstremne vrijednosti procijenit ćemo krajnju desnu točku svake funkcije distribucije tj. ultimativan svjetski rekord svake discipline.

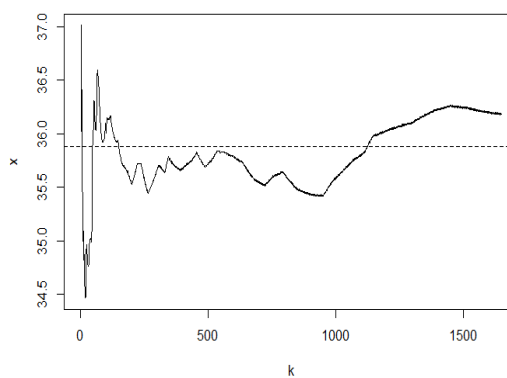
U Tablici 4.6 grafički ćemo prikazati procjenitelje krajnje desne točke kao funkcije od  $k$  i to za sve navedene atletske discipline osim trčanja na 400 m kod muškaraca i skoka u dalj kod muškaraca. Zatim ćemo dobivene procjenitelje pojedinih disciplina iskazanih u brzinama ponovno transformirati u vrijeme. Konačno, sve dobivene rezultate prikazat ćemo u Tablici 4.7 zajedno s trenutnim svjetskim rekordom za svaku disciplinu.

## POGLAVLJE 4. PRIMJENA TEORIJE EKSTREMNIH VRIJEDNOSTI NA REKORDE U ATLETICI

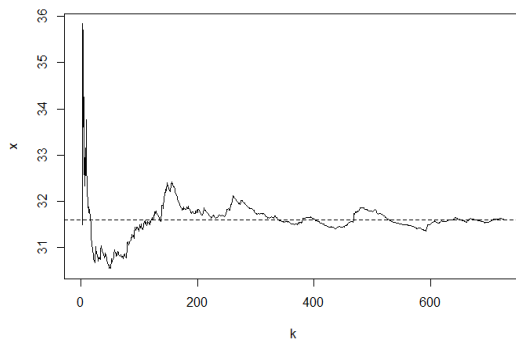
100 m (muškarci)



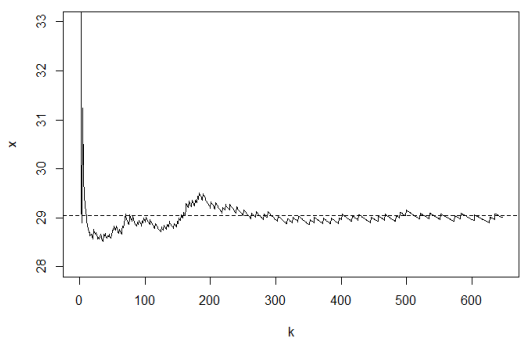
100 m (žene)



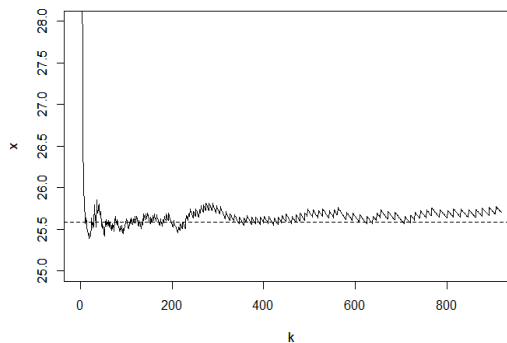
400 m (žene)



800 m (muškarci)

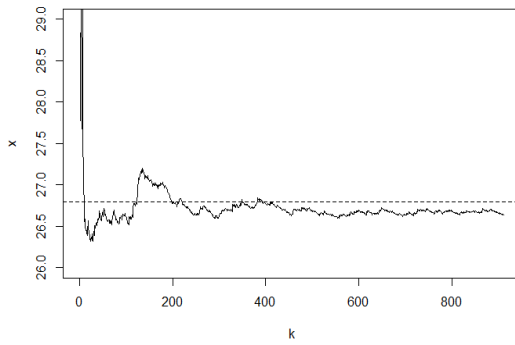


800 m (žene)

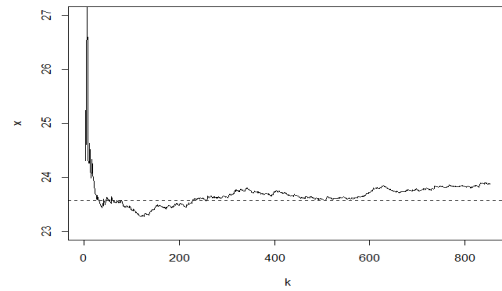


4.3. PROCJENA ULTIMATIVNOG SVJETSKOG REKORDA POJEDINE DISCIPLINE

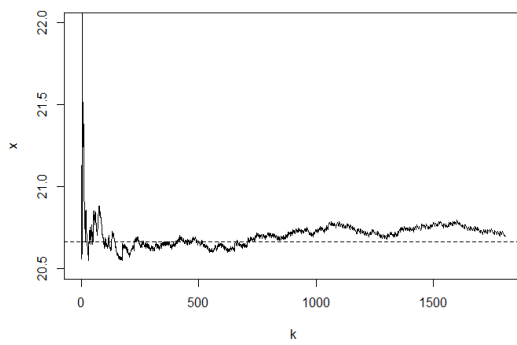
1500 m (muškarci)



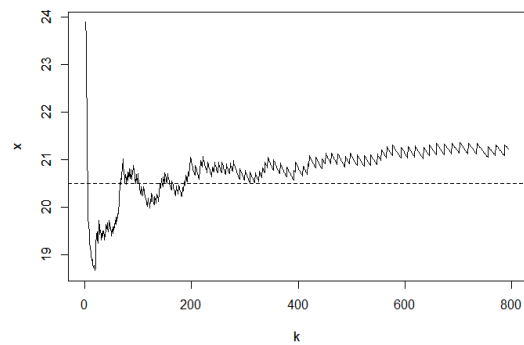
1500 m (žene)



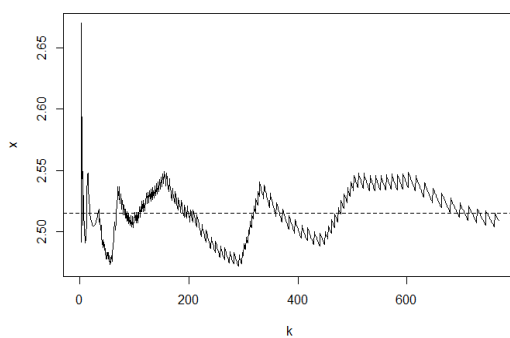
Maraton (muškarci)



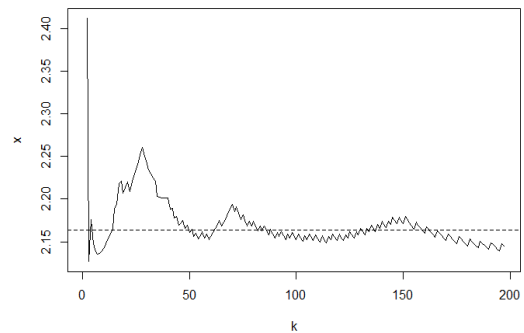
Maraton (žene)



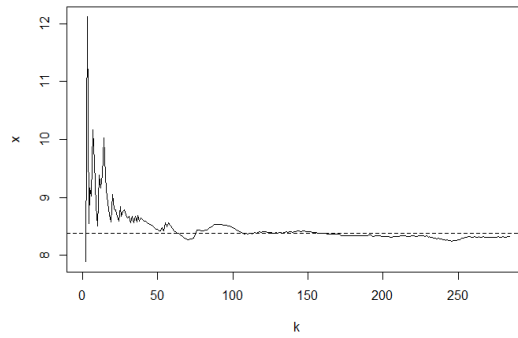
Skok u vis (muškarci)



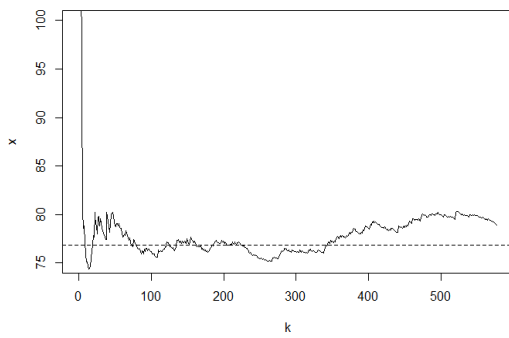
Skok u vis (žene)



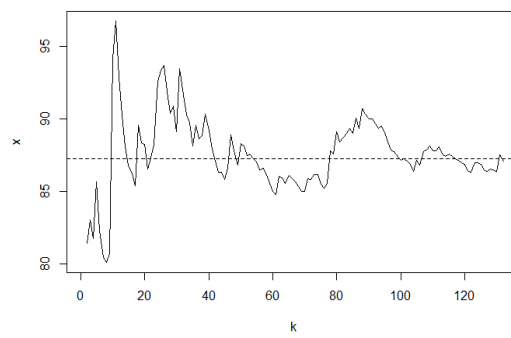
Skok u dalj (žene)



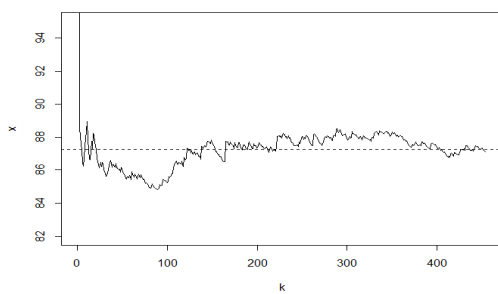
Bacanje diska (muškarci)



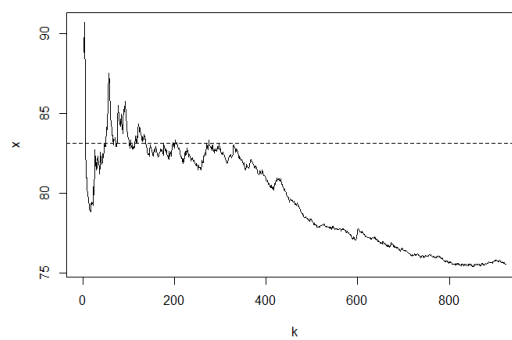
Bacanje diska (žene)

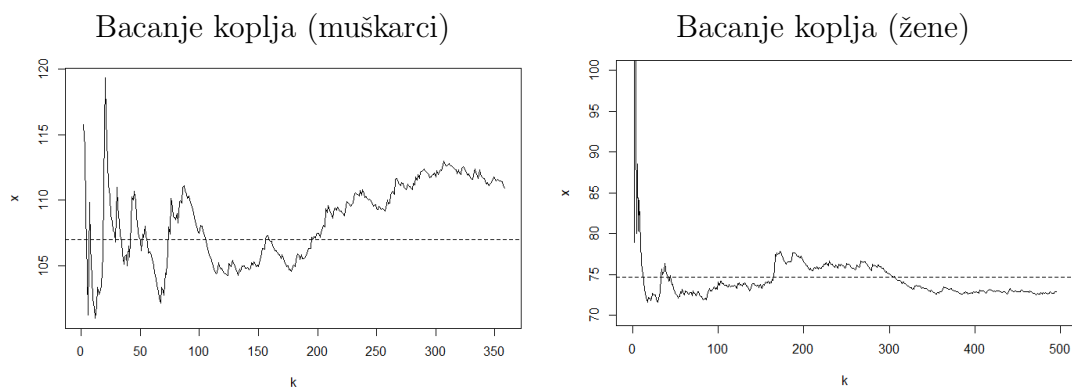


Bacanje kladiva (muškarci)



Bacanje kladiva (žene)





Tablica 4.6: Procjenitelji krajnje desne točke atletskih disciplina (kod muškaraca i žena)

Disciplina	Muškarci		Žene	
	Krajnja točka	Svjetski rekord	Krajnja točka	Svjetski rekord
Trčanje na 100 m	9.28	9.58	10.03	10.49
Trčanje na 400 m	/	43.18	45.59	47.60
Trčanje na 800 m	1:39.16	1:40.91	1:52.55	1:53.28
Trčanje na 1500 m	3:21.50	3:26.00	3:49.10	3:50.07
Maraton	2:01:56	2:02:57	2:02:56	2:15:25
Skok u vis	2.52	2.45	2.16	2.09
Skok u dalj	/	8.95	8.37	7.52
Bacanje diska	76.83	74.08	87.27	76.80
Bacanje kladiva	87.27	86.74	83.15	82.98
Bacanje koplja	107.00	98.48	74.65	72.28

Tablica 4.7: Procijenjene vrijednosti  $\hat{x}^*$  zajedno s trenutnim svjetskim rekordom

Rezultati dobiveni našom analizom prikazani u Tablici 4.7 pokazuju kako je kod trčanja na kraće relacije (100 m, 400 m, 800 m, 1500 m) manje prostora za poboljšanje rekorda. Također, po ovoj analizi u muškom maratonu moguće je postići mali napredak od 1 minute (0.83%), dok je kod ženskog maratona moguće znatno popraviti trenutni svjetski rekord i to za gotovo 13 minuta (9.22%). S druge strane, u disciplini bacanja koplja kod muškaraca moguće je veliko poboljšanje trenutnog rekorda za čak 8 metara (8.65%), dok je kod žena u toj disciplini moguće poboljšanje od svega 2 metra (3.28%). Konačno, prema našoj analizi najveće poboljšanje je ostvarivo u bacanju diska kod žena (13.63%).

## 4.4 Procjena kvalitete trenutnog svjetskog rekorda

Iduće važno pitanje kojim ćemo se baviti je kvaliteta trenutnog svjetskog rekorda pojedine discipline, odnosno koliko je teško poboljšati ga. *Kvalitetu* svjetskog rekorda mjerimo pomoću  $n(1 - F(X_{n,n}))$ , što je očekivani broj prekoračenja trenutnog svjetskog rekorda  $X_{n,n}$ . Što je taj broj manji, to je teže poboljšati trenutni rekord, odnosno to je rekord bolji.

Razlika između procijenjenog ultimativnog svjetskog rekorda i trenutnog svjetskog rekorda tj.  $x^* - X_{n,n}$  čini nam se kao prirodan način za računanje kvalitete trenutnog svjetskog rekorda, međutim na taj način ne uzimamo u obzir ponašanje repa funkcije distribucije  $F$ .

Sada, prisjetimo se relacije (3.8)

$$\hat{p}_n := \frac{k}{n} \left( \max \left( 0, 1 + \hat{\gamma} \frac{X_{n,n} - \hat{b}}{\hat{a}} \right) \right)^{-1/\hat{\gamma}}.$$

Kako je  $\hat{p}_n = 1 - F(X_{n,n})$  te ako sve pomnožimo s  $n$  dobit ćemo upravo procjenu od  $n(1 - F(X_{n,n}))$ . Traženi procjenitelj za kvalitetu svjetskog rekorda je

$$Q := n(1 - F(X_{n,n})) = k \left[ \max \left( 0, 1 + \hat{\gamma} \frac{X_{n,n} - \hat{b}}{\hat{a}} \right) \right]^{-1/\hat{\gamma}}. \quad (4.1)$$

U idućem teoremu pokazat ćemo svojstva granične distribucije od  $Q$ .

**Teorem 4.4.1.** *Neka je  $\gamma > -1/2$ . Neka je  $F$  neprekidna funkcija distribucije te pretpostavimo da  $U$  zadovoljava uvjet drugog reda (2.7) gdje je  $\rho < 0$ . Pretpostavimo*

da je  $k \rightarrow \infty$ ,  $k/n \rightarrow 0$  i  $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbf{R}$ . Nadalje pretpostavimo da su

$$\sqrt{k} \left( \frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} - 1 \right), \sqrt{k} \left( \frac{b(\frac{n}{k}) - b(\frac{n}{k})}{a(\frac{n}{k})} \right), \sqrt{k}(\hat{\gamma} - \gamma)$$

svi  $O_p(1)$ . Tada vrijedi

$$Q \xrightarrow{d} \text{Exp}(1),$$

kako  $n \rightarrow \infty$ .

*Dokaz:* Vidi u [1] dokaz Teorema 1.

**Napomena 7.** Matematička oznaka  $O_p(1)$  opisuje granično ponašanje funkcije te nam govori kako vrijednosti triju navedenih funkcija u prethodnom teoremu teže ka konstanti dok  $k \rightarrow \infty$ .

U postupku procjene kvalitete svjetskog rekorda

$$Q := k \left[ \max \left( 0, 1 + \hat{\gamma} \frac{X_{n,n} - \hat{b}}{\hat{a}} \right) \right]^{-1/\hat{\gamma}}$$

koristit ćemo istu metodu kao i kod procjene indeksa razdiobe ekstremne vrijednosti i krajnje desne točke, tražeći prvu stabilnu regiju od  $k$  gdje je  $50 \leq k \leq 250$ .

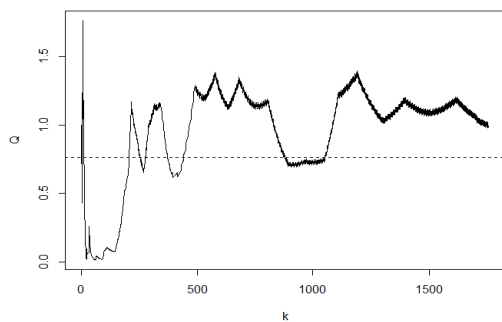
Uočimo pritom u relaciji (4.1) kako procjenitelj  $Q$  ovisi i o  $n$ , stoga nije potpuno trivijalan s obzirom da različite vrijednosti od  $n$  donekle otežavaju usporedbu kvalitete svjetskog rekorda po disciplinama. Međutim, prilikom naše procjene nećemo koristiti konkretan  $n$ , gdje  $n$  predstavlja broj svih atletičara u pojedinoj disciplini iz razloga jer je sama definicija atletičara nejasna pa je i definiranje i mjerenje broja  $n$  teško.

Također, prilikom procjene koristit ćemo fiksne procjenitelje od  $\gamma$  iz Tablice 4.5. U Tablici 4.8 grafički ćemo prikazati procijenjene kvalitete svjetskih rekorda svake atletske discipline kao funkcije od  $k$ .

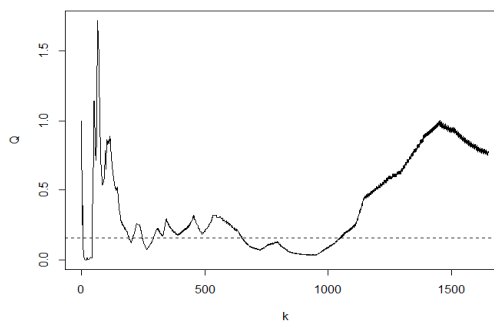


POGLAVLJE 4. PRIMJENA TEORIJE EKSTREMNIH VRIJEDNOSTI NA  
REKORDE U ATLETICI

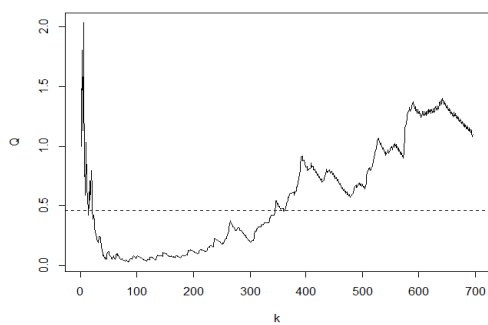
100 m (muškarci)



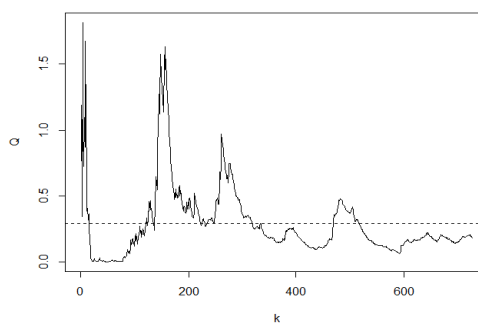
100 m (žene)



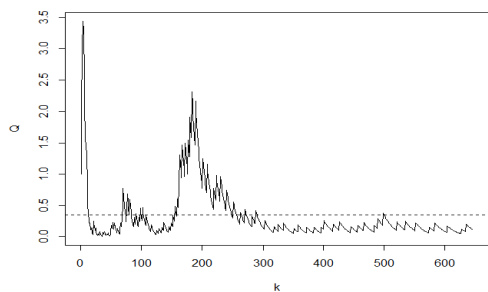
400 m (muškarci)



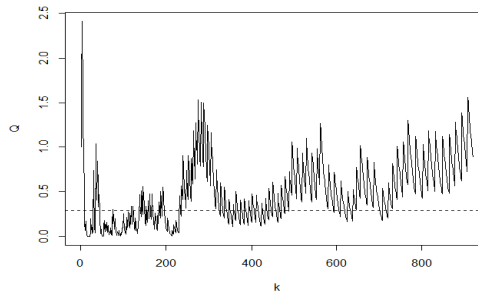
400 m (žene)



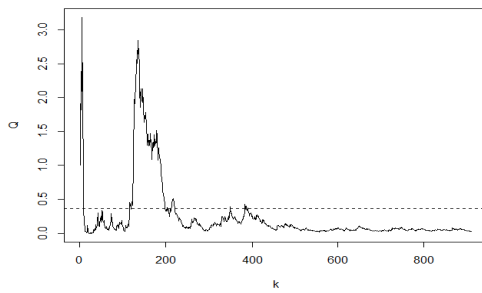
800 m (muškarci)



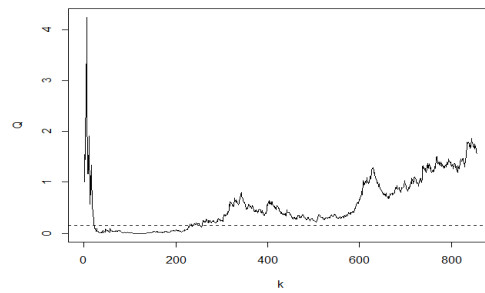
800 m (žene)



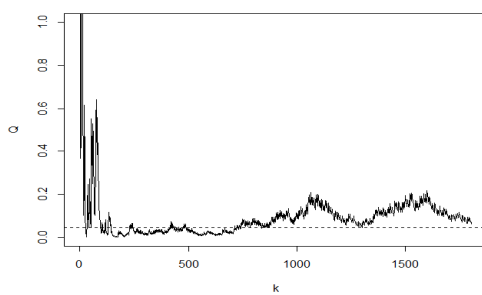
1500 m (muškarci)



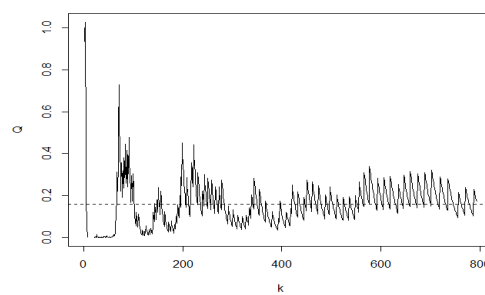
1500 m (žene)



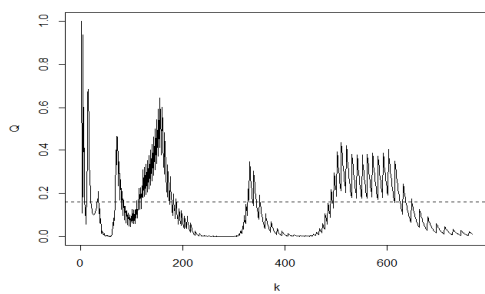
Maraton (muškarci)



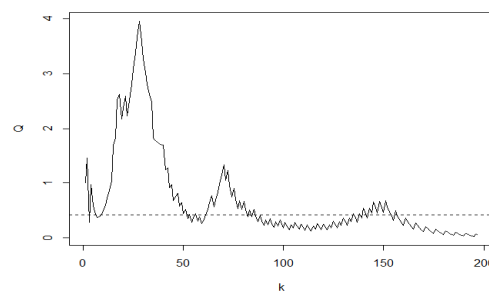
Maraton (žene)



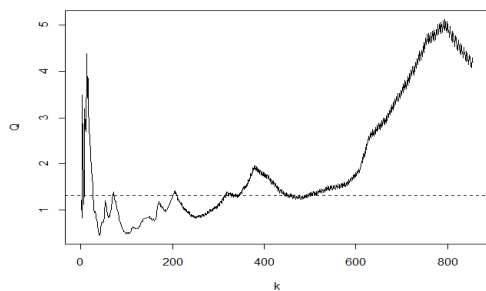
Skok u vis (muškarci)



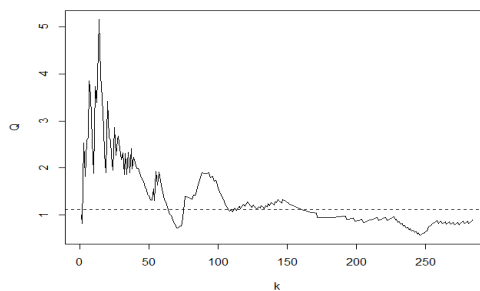
Skok u vis (žene)



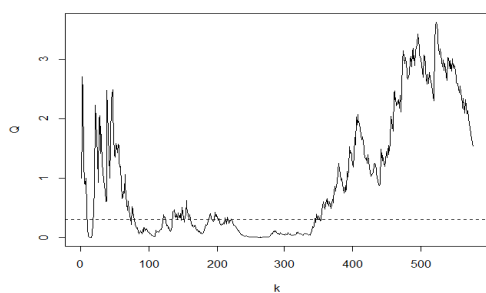
Skok u dalj (muškarci)



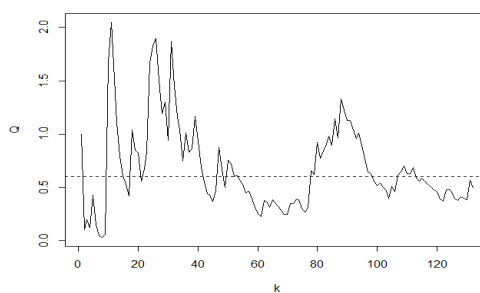
Skok u dalj (žene)



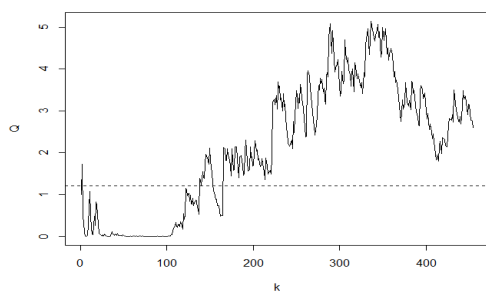
Bacanje diska (muškarci)



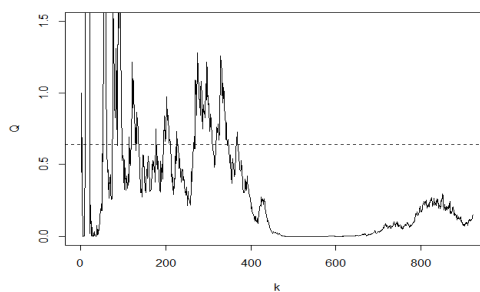
Bacanje diska (žene)

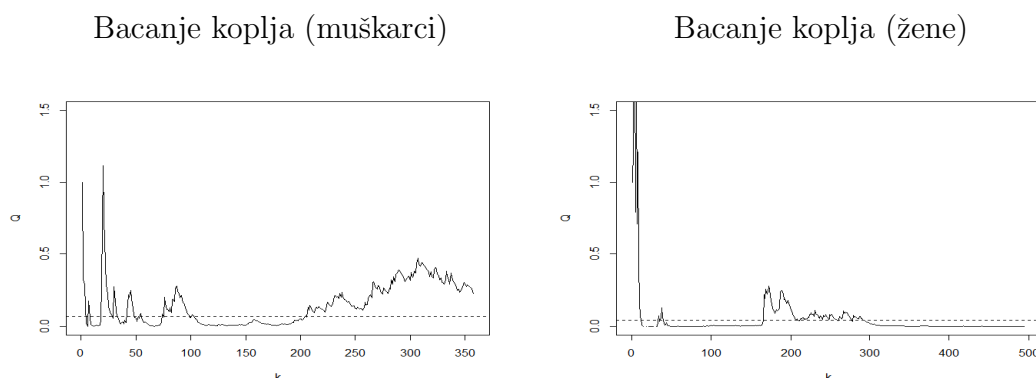


Bacanje kladiva (muškarci)



Bacanje kladiva (žene)





Tablica 4.8: Procjenitelj kvalitete trenutnog svjetskog rekorda  $Q$  za pojedine atletske discipline (kod muškaraca i žena)

Na temelju rezultata iz Teorema 4.4.1. umjesto samog  $Q$  prezentirat ćemo  $e^{-Q}$  koji konvergira k uniformnoj distribuciji na intervalu  $(0, 1)$  te ćemo na taj način dobiti ne samo relativan već i apsolutan kriterij za određivanje kvalitete svjetskih rekorda. Pokažimo sada tvrdnju: ako je slučajna varijabla  $X$  eksponencijalno distribuirana s parametrom 1 tj.  $X \sim Exp(1)$ , tada je slučajna varijabla  $e^{-X}$  uniformno distribuirana na intervalu  $[0, 1]$  tj.  $e^{-X} \sim U(0, 1)$ .

Neka je  $X \sim Exp(1)$ . Definirajmo slučajnu varijablu  $Y := e^{-X}$ . Dokažimo da je tako definirana slučajna varijabla  $Y \sim U(0, 1)$ . Vrijedi

$$F_Y(y) = P(Y \leq y) = P(e^{-X} \leq y) = P(X \geq -\ln y) = 1 - P(X \leq -\ln y) = 1 - F_X(-\ln y).$$

Kako je po pretpostavci  $X \sim Exp(1)$ , funkcija distribucije od  $X$  izgleda  $F_X(x) = 0$  za  $x < 0$ ,  $F_X(x) = 1 - e^{-x}$ , za  $x \geq 0$ . Sada, za  $-\ln y < 0$  vrijedi  $F_X(-\ln y) = 0$ , pri čemu  $-\ln y < 0 \implies \ln y > 0 \implies y > 1$ . Konačno, za  $y > 1$   $F_Y(y) = 1 - F_X(-\ln y) = 1 - 0 = 1$ . Za  $-\ln y \geq 0$  vrijedi  $F_X(-\ln y) = 1 - e^{\ln y} = 1 - y$ , pri čemu  $-\ln y \geq 0 \implies \ln y \leq 0 \implies y \leq 1$ . Konačno, za  $y \leq 1$  imamo  $F_Y(y) = 1 - F_X(-\ln y) = 1 - 1 + y = y$ .

Po definiciji funkcije prirodnog logaritma znamo  $\ln : \mathbf{R}^+ \rightarrow \mathbf{R}$  pa je  $y > 0$ . Stoga vrijedi  $y < 0 \implies F_Y(y) = 0$ . Promotrimo li sve navedene slučajeve, zaključujemo kako smo dobili upravo funkciju distribucije uniformne slučajne varijable  $Y$  na intervalu  $(0, 1)$ . Time smo završili dokaz.

U Tablici 4.9 prikazat ćemo procijenjene vrijednosti kvalitete  $Q$  i  $e^{-Q}$  dobivene našom analizom, za svih 20 atletske discipline zajedno sa trenutnim svjetskim rekordom.

Disciplina	Atletičar	Rekord	Godina	$Q$	$e^{-Q}$
Bacanje koplja (Ž)	Barbora Špotáková	72.28	2008.	0.041	0.959
Maraton (M)	Denis Kipruto Kimettó	2:02:37	2014.	0.044	0.956
Bacanje koplja (M)	Jan Zelezný	98.48	1996.	0.066	0.935
Trčanje na 1500 m (Ž)	Genzebe Dibaba	3:50.07	2015.	0.154	0.856
Trčanje na 100 m (Ž)	Florence Griffith-Joyner	10.49	1988.	0.156	0.855
Maraton (Ž)	Paula Radcliffe	2:15:25	2003.	0.158	0.853
Skok u vis (M)	Javier Sotomayor	2.45	1993.	0.159	0.852
Trčanje na 800 m (Ž)	Jarmila Kratochvílová	1:53.28	1983.	0.285	0.751
Trčanje na 400 m (Ž)	Marita Koch	47.60	1985.	0.289	0.748
Bacanje diska (M)	Jürgwn Schult	74.08	1986.	0.296	0.743
Trčanje na 800 m (M)	David Rudisha	1:40.91	2012.	0.351	0.703
Trčanje na 1500 m (M)	Hicham El Guerrouj	3:26.00	1998.	0.357	0.699
Skok u vis (Ž)	Stefka Kostadinova	2.09	1987.	0.411	0.662
Trčanje na 400 m (M)	Wayde Van Niekerk	43.03	2016.	0.461	0.630
Bacanje diska (Ž)	Gabriele Reinsch	76.80	1988.	0.601	0.548
Bacanje kladiva (Ž)	Anita Wlodarczyk	82.98	2016.	0.642	0.526
Trčanje na 100 m (M)	Usain Bolt	9.58	2009.	0.737	0.478
Skok u dalj (Ž)	Galina Chistyakova	7.52	1988.	1.121	0.325
Bacanje kladiva (M)	Yutiy Sedykh	86.74	1986.	1.198	0.301
Skok u dalj (M)	Mike Powell	8.95	1991.	1.312	0.269

Tablica 4.9: Procijenjene kvalitete trenutnih svjetskih rekorda  $Q$  i  $e^{-Q}$  (silazno sortirane)

Napomenimo kako smo u Tablici 4.9 naveli svjetske rekorde iz baze podataka koju smo napravili na početku ovog rada, a u kojoj smo zabilježili rezultate zaključno s krajem 2017. godine. Nositelj svjetskog rekorda u muškom maratonu bio je Denis Kipruto Kimettó koji je 2014. godine postavio rekord od 2:02:37. U međuvremenu, točnije 16.9.2018. rekord je oboren na 2:01:39 od strane altetičara Eliuda Kipchoge.

## 4.5 Osjetljivost procjena na izbor $k$

Sve naše procjene  $\hat{\gamma}$ ,  $\hat{x}^*$  i  $Q$  ovise o izboru prve stabilne regije od  $k$ . Za manje vrijednosti od  $k$ , uglavnom između 50 i 250, pronalazili smo skupove od barem 50 uzastopnih  $k$  takvih da procijenjene vrijednosti na toj regiji ne variraju previše. Međutim ponekad između 50 i 250 nismo mogli pronaći stabilnu regiju jer su vrijednosti u toj regiji jako oscilirale, stoga smo stabilnu regiju od  $k$  pokušavali pronaći i za veće vrijednosti od  $k$ .

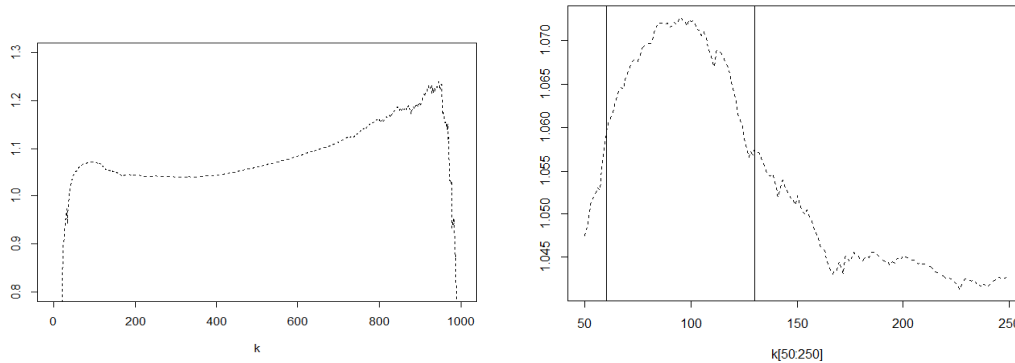
Kako je ova metoda često korištena u literaturi i u praksi, bitno je dotaknuti se pitanja osjetljivosti cjelokupne procjene na naš izbor  $k$ . U ovom ćemo poglavlju pokazati kako naše procjene nisu osjetljive na izbor  $k$ .

Našu tvrdnju pokazat ćemo na dva slučaja, prvi kada je  $\gamma > 0$ , a drugi za  $\gamma < 0$ . Za ispitivanje tvrdnje, simulirat ćemo deset slučajnih uzorka duljine 1000.

U prvom slučaju, simulirat ćemo uzorke iz standardne Cauchyjeve distribucije od kojih ćemo uzeti apsolutnu vrijednost. Standardna Cauchyjeva distribucija ima funkciju gustoće

$$f(x; 0, 1) := \frac{1}{\pi(1 + x^2)}, \quad x \geq 0.$$

Ta funkcija distribucije je u domeni atrakcije od distribucije graničnih vrijednosti  $G_\gamma$  s indeksom ekstremne vrijednosti  $\gamma = 1$ . Koristeći metodu momenata prikazat ćemo procjenitelj indeksa ekstremne vrijednosti za prvi slučajni uzorak kao funkciju od  $k$  te odrediti prvu stabilnu regiju za  $50 \leq k \leq 250$ .



Tablica 4.10: Procjenitelj  $\hat{\gamma}$  standardne Cauchyjeve distribucije kao funkcija od  $k$  (lijeva slika); Prva stabilna regija:  $60 \leq k \leq 130$  (desna slika)

Određivanjem prosječne vrijednosti procjenitelja na odabranoj regiji dobili smo  $\hat{\gamma} = 1.07$ . Na temelju dobivenog procjenitelja vidimo kako naša metoda primijenjena na ovaj uzorak rezultira dobrim procjeniteljem, no moguće je da se radi o slučajnoj podudarnosti. Stoga ćemo postupak ponoviti na preostalim devet slučajnih uzoraka. Dobivene procijenjene vrijednosti su prikazane su u Tablici 4.11.

Slučajni uzorci	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
$\hat{\gamma}$	1.07	0.89	0.94	0.87	1.06	0.96	0.86	0.99	0.98	0.99

Tablica 4.11: Procijenjene vrijednosti  $\hat{\gamma}$  za deset slučajnih uzoraka iz standardne Cauchyjeve distribucije

Zaključujemo kako smo dobili deset *dobrih* procjena s prosječnom vrijednosti od 0.96 i standardnom devijacijom od 0.07.

U drugom slučaju koristit ćemo podatke koji dolaze iz distribucije sa konačnom desnom krajnjom točkom, odnosno gdje je  $\gamma \leq 0$ . Podatke ćemo simulirati iz obrnute Burr distribucije

$$F(x) := 1 - \left( \frac{\beta}{\beta + (x^* - x)^{-\tau}} \right)^\lambda \quad x < x^*, \quad (4.2)$$

s parametrima  $\beta > 0$ ,  $\tau > 0$ ,  $\lambda > 0$  i krajnjom desnom točkom  $x^* > 0$ . Ovo je također distribucija teškog repa te predstavlja poopćenje Pareto distribucije.

Pokažimo, prvo, kako iz Burrove distribucije

$$\bar{F}(x) = \left( \frac{\beta}{\beta + x^\tau} \right)^\lambda$$

generirati obrnutu Burrovu distribuciju. Neka je  $Y'$  slučajna varijabla na  $(-\infty, x^*)$  te neka je  $Y := 1/(x^* - Y')$  slučajna varijabla koja dolazi iz Burrove distribucije. Pokažimo da je tada  $Y'$  iz obrnute Burrove distribucije. Vrijedi

$$\begin{aligned} \bar{F}_{Y'}(x) &= P(Y' > x) = P(x^* - Y' < x^* - x) = P\left(\frac{1}{x^* - Y'} > \frac{1}{x^* - x}\right) = \\ &= P(Y > (x^* - x)^{-1}) = \left(\frac{\beta}{\beta + (x^* - x)^{-\tau}}\right)^\lambda, \end{aligned}$$

pri čemu je  $x < x^*$ . Za  $x \nearrow x^*$  vrijedi

$$\left(\frac{\beta}{\beta + (x^* - x)^{-\tau}}\right)^\lambda \sim \frac{1}{(x^* - x)^{-\tau\lambda}},$$

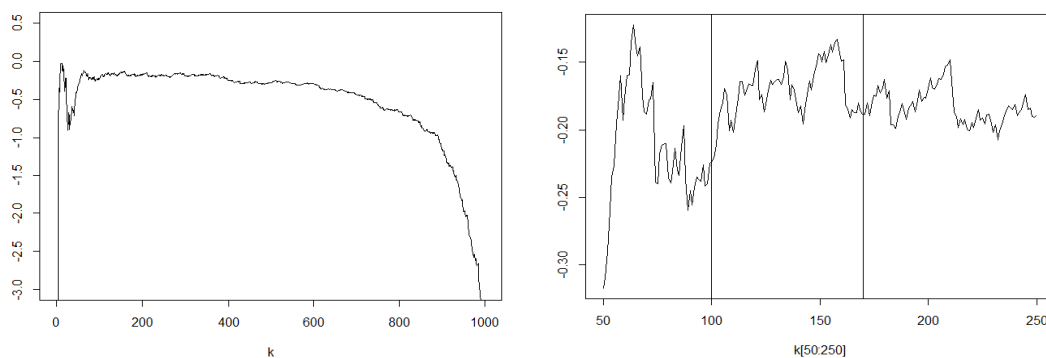
stoga je repni indeks jednak  $-\tau\lambda$  tj.  $1/\gamma = -\tau\lambda$ , iz čega slijedi  $\gamma = -1/(\tau\lambda)$ .

Koristit ćemo proizvoljno odabranu vrijednost indeksa razdiobe ekstremne vrijednosti  $\gamma = -0.20$ , stoga su vrijednosti preostalih parametara  $\beta = 1$ ,  $\tau = 10$ ,  $\lambda = 1/2$  i  $x^* = 10$  i vrijedi  $\gamma = -1/(10 \cdot 1/2) = -1/5 = -0.20$ . U  $R - u$  ne postoji gotova naredba za simulaciju podataka iz obrnute Burr distribucije, stoga ćemo podatke simulirati na sljedeći način. Zamijenit ćemo  $F(x)$  s  $U$ , gdje  $U$  predstavlja uniformno distribuirane podatke na intervalu  $(0, 1)$  te ćemo izraziti  $x$ . Računamo

$$\begin{aligned} U &= 1 - \left(\frac{1}{1 + (10 - x)^{-10}}\right)^{1/2} \iff 1 - U = (1 + (10 - x)^{-10})^{-1/2} \\ \iff (1 - U)^{-2} - 1 &= (10 - x)^{-10} \iff ((1 - U)^{-2} - 1)^{-1/10} = 10 - x \\ \iff x &= 10 - \left(\frac{1}{(1 - U)^2} - 1\right)^{-1/10}. \end{aligned}$$

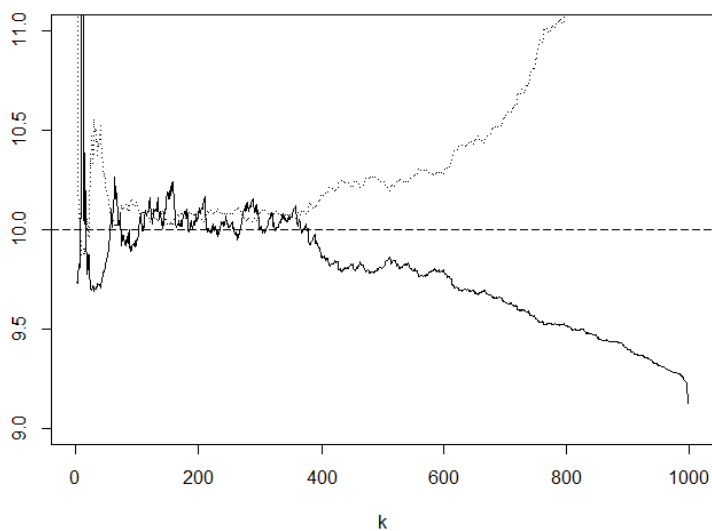
Sada možemo simulirati prvi slučajni uzorak duljine 1000 iz obrnute Burr distribucije te koristeći metodu momenata prikazati procjenitelj indeksa razdiobe ekstremne vrijednosti kao funkciju od  $k$ .





Tablica 4.12: Procjenitelj  $\hat{\gamma}$  obrnute Burrove distribucije kao funkcija od  $k$  (lijeva slika); Prva stabilna regija:  $100 \leq k \leq 170$  (desna slika)

Određivanjem prosječne vrijednosti procjenitelja na odabranoj regiji dobili smo  $\gamma = -0.17$ . Sada ćemo procijeniti krajnju desnu točku  $x^*$  na dva načina, koristeći procijenjene vrijednosti od  $\gamma$  u ovisnosti o  $k$  i koristeći fiksnu procjenu  $\hat{\gamma} = -0.17$ . Rezultate jednog uzorka dobivene na oba dva načina prikazat ćemo na istom grafu na Slici 4.13 i to tako da ćemo punom linijom prikazati vrijednosti dobivene koristeći  $\gamma$  u ovisnosti o  $k$ , točkastom linijom koristeći fiksnu vrijednost od  $\gamma$  te iscrtkanom linijom pravu vrijednost  $x^* = 10$ .



Slika 4.13: Procjenitelji desne krajnje točke kao funkcije od  $k$

Koristeći metodu odabira prve stabilne regije od  $k$  za procjenu krajnje desne točke u prvom slučaju ( $\gamma$  u ovisnosti o  $k$ ) dobijemo  $\hat{\gamma} = 10.03$ , a u drugom slučaju (fiksni  $\gamma$ )  $\hat{\gamma} = 10.08$ . Ponovimo li postupak i na preostalim slučajnim uzorcima dobivamo procijenjene vrijednosti prikazane u Tablici 4.13.

Slučajni uzorci	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
$\hat{\gamma}$	-0.17	-0.16	-0.23	-0.24	-0.14	-0.21	-0.32	-0.21	-0.29	-0.18
$\hat{x}^*$ ( $\gamma$ u ovisnosti o $k$ )	10.03	10.12	9.87	9.97	10.12	9.94	9.77	9.94	9.87	9.94
$\hat{x}^*$ (fiksni $\gamma$ )	10.08	10.10	9.95	9.96	10.18	9.85	9.80	9.95	9.89	9.99

Tablica 4.13: Procijenjene vrijednosti  $\gamma$  i  $\hat{x}^*$  za deset slučajnih uzoraka obrnute Burrove distribucije

Rezultati pokazuju kako je prosječna vrijednost indeksa ekstremne vrijednosti -0.22, a standardna devijacija 0.06, prosječna vrijednost procjenitelja krajnje desne točke uz  $\gamma$  ovisan o  $k$ , 9.96, a standardna devijacija 0.11 te prosječna vrijednost od  $\hat{x}^*$  uz fiksni  $k$  jednaka 9.98, a standardna devijacija 0.12.

Na temelju gornjih simulacija, možemo zaključiti kako je metoda procjene koristeći prvu stabilnu regiju od  $k$  gdje  $50 \leq k \leq 250$  slabo (nedovoljno) osjetljiva na odabir regije  $k$ .

## 4.6 Zaključak

Naše procjene krajnjih točaka svjetskih disciplina tj. procjene ultimativnih svjetskih rekorda (Tablica 4.7) pokazuju kako je i dalje moguće ostvariti značajan napredak u odnosu na trenutne svjetske rekorde.

Rezultati naše analize pokazuju kako je najveći napredak moguće ostvariti u bacanju diska kod žena, i to za 13.63%, a slijedi moguće poboljšanje od 11.30% u skoku u dalj kod žena, poboljšanje od 9.22% u ženskom maratonu te poboljšanje trenutnog rekorda za 8.65% u bacanju koplja kod muškaraca. Promotrimo li procijenjene vrijednosti indeksa razdiobe ekstremne vrijednosti (Tablica 4.5) navedenih disciplina uočavamo kako su relativno visoke u odnosu na ostale discipline i to redom,

$-0.149, -0.117, -0.190, -0.162.$

S druge strane, najmanji napredak ostvariv je u bacanju kladiva kod žena (0.20%), trčanju na 1500 metara kod žena (0.42%), bacanju kladiva kod muškaraca (0.61%) te trčanju na 800 metara kod žena (0.64%). Procijenjene vrijednosti indeksa razdiobe ekstremne vrijednosti ovih disciplina su pak relativno male te iznose  $-0.332, -0.271, -0.401, -0.247$ , redom.

Na temelju gornjih rezultata možemo zaključiti da povećanjem vrijednost indeksa razdiobe ekstremne vrijednosti raste mogućnost poboljšanja trenutnog svjetskog rekorda.

Nadalje, kvalitetu trenutnog svjetskog rekorda mjerili smo pomoću  $n(1 - F(X_{n,n}))$ , što je očekivani broj prekoračenja trenutnog svjetskog rekorda. Svjetske rekorde visoke kvalitete teško je poboljšati, dok je za rekorde niske kvalitete vrlo vjerojatno da će u skorijoj budućnosti biti oboreni.

Na temelju procijenjenih kvaliteta svjetskih rekorda (Tablica 4.9) dobivenih našom analizom, dobili smo da su discipline najveće kvalitete bacanje koplja kod žena (0.959), muški maraton (0.956) te bacanje koplja kod muškaraca (0.935) stoga je trenutni rekord u tim disciplinama najteže poboljšati. S druge strane, skoro poboljšanje vrlo je vjerojatno kod skoka u dalj kod muškaraca (0.269), bacanja kladiva kod muškaraca (0.301) te skoka u dalj kod žena (0.325).

Usporedimo li rezultate prikazane u Tablicama 4.7 i 4.9 zaključujemo kako kvaliteta svjetskog rekorda i mogućnost poboljšanja rekorda nisu ovisni jedno o drugome. To znači da rekord može imati veliku mogućnost poboljšanja iako je visoke kvalitete, ali visoka kvaliteta upućuje kako je to poboljšanje teško ostvariti. Na primjer, to vrijedi kod bacanja koplja kod muškaraca. U ovoj disciplini moguće je ostvariti veliki napredak (8.65%), ali je isto tako trenutni rekord teško oboriti jer je visoke kvalitete (0.935). Vrijedi i obrat, kao što je na primjer kod bacanja kladiva kod muškaraca gdje je moguće ostvariti mali napredak (0.61%), ali je isto tako trenutni rekord niske kvalitete (0.301).

# Bibliografija

- [1] J. H. J. Einmahl i J. R. Magnus (2008): Records in Athletics Through Extreme - Value Theory, *Journal of the American Statistical Association*, 103:484, 1382-1391.
- [2] L. de Haan i A. Ferreira: *Extreme Value Theory: An Introduction*. Springer, New York (2006).
- [3] S. I. Resnick: *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York (2007).
- [4] N. Sarapa: *Teorija vjerojatnosti*. Školska knjiga, Zagreb (2002).



# Sažetak

Primjenom teorije ekstremnih vrijednosti na svjetske rekorde u atletici pokušali smo dati odgovore na dva glavna pitanja: Koji je konačan tj. ultimativan svjetski rekord u pojedinoj disciplini? Koliko su trenutni rekordi "dobri" tj. koliko ih je teško poboljšati? Pritom smo u obzir uzeli 10 atletskih disciplina gdje smo posebno razlikovali muškarce i žene, a to su trčanje na 100 m, 400 m, 800 m, 1500 m, maraton, skok u vis i dalj te bacanje diska, kladiva i koplja. Tako dobivena procjena ultimativnog rekorda može nam reći što je moguće postići u bližoj budućnosti s obzirom na današnje fizičke sposobnosti, znanje i sportsku opremu te regulativu o zabranjenim supstancama u sportu.

Važan rezultat teorije ekstremnih vrijednosti koji smo koristili kaže da maksimumi uzorka nezavisnih jednako distribuiranih slučajnih varijabla, nakon odgovarajuće normalizacije, mogu konvergirati po distribuciji samo u jednu od idućih tri distribucija, Gumbelovu, Frèchetovu ili Weibullovu. Za našu analizu posebno je važna Weibullova koja odgovara razdiobi s konačnim nosačem. Prilikom procjene parametara važnih za ovu analizu, susreli smo se s problemom izbora vrijednosti za  $k$  na kojima temeljimo naše procjenitelje te smo ispitali osjetljivost procjena na izbor  $k$ .

Dobivene procjene pokazuju kako je i dalje moguće ostvariti značajan napredak u odnosu na trenutne svjetske rekorde. Najveći napredak je moguće ostvariti u bacanju diska i skoku u dalj kod žena, a najmanji u bacanju kladiva i trčanju na 1500 metara kod žena. Također zaključujemo kako neki od trenutnih svjetskih rekorda imaju višu kvalitetu od ostalih, što znači da ih je teže poboljšati.



# Summary

By applying extreme value theory to world records in athletics, we attempted to answer two questions: What is the ultimate world record in a specific athletic event? How "good" is the current world record, i.e. how difficult is it to improve? We consider 10 events for both men and women, running at 100 m, 400 m, 800 m, 1500m, marathon, high jump, long jump, hammer throw, discus throw and javelin throw. Our estimated ultimate record tells us what is possible in the near future, given the present knowledge, sports equipment and drug regulation.

Important result in extreme value theory used in our analysis says that the maximum of a sample of independent identically distributed random variables, after proper normalization, can only converge in distribution to one of three possible distributions, Gumbel, Frèchet or Weibull. Especially important for our analysis is Weibull, which is a distribution with finite support. While estimating parameters, we face the problem of finding a good value for  $k$  so we tested procedure sensitivity to the choice of  $k$ .

Our estimations shows that substantial improvements over current record are still possible, especially in discus throwing and long jump for women. On the other hand, very little improvement is possible in hammer throwing and running at 1500 meters for women. Also, some world records have higher quality then the rest, which means they are more difficult to improve.





# Životopis

Lara Čigir rođena je 23. veljače 1993. godine u Zagrebu. Pohađala je II. opću gimnaziju te nakon polaganja državne mature 2011. godine upisuje Preddiplomski studij Matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu. Nakon završetka prediplomskog studija, godine 2016. upisuje Diplomski studij Matematičke statistike na istom, Matematičkom odsjeku.