Asimptotika procjenitelja maksimalne vjerodostojnosti

Krtinić, Marin

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Marin Krtinić

ASIMPTOTIKA PROCJENITELJA MAKSIMALNE VJERODOSTOJNOSTI

Diplomski rad

Voditelj rada:
prof. dr. sc. Miljenko Huzak

Zagreb, rujan, 2021

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. ____________________________

Sadržaj

Uvod

Procjena parametara metodom maksimalne vjerodostojnosti jedna je od najpopularnijih i najčešće korištenih metoda procjene parametara u inferencijalnoj statistici. Osim što su procjenitelji dobiveni metodom maksimalne vjerodostojnosti često jednostavni za odrediti, dodatni razlog popularnosti je što tako dobiveni procjenitelji u regularnim parametarskim modelima imaju lijepa svojstva: konzistentni su, asimptotski su efikasni i normalno distribuirani. U ovom radu ćemo formulirati uvjete za navedena svojstva i dokazati da pod tim uvjetima i vrijede. Rad je podijeljen u 5 poglavlja. Na početku definiramo osnovne pojmove matematičke statistike i navodimo rezultate iz teorije vjerojatnosti koje ćemo koristiti u ostatku rada. U drugom poglavlju dajemo motivaciju za uniformni jaki zakon velikih brojeva, te iskazujemo i dokazujemo jednu verziju uniformnog jakog zakona velikih brojeva. U trećem poglavlju definiramo procjenitelj maksimalne vjerodostojnosti, na primjerima pokazujemo da procjenitelj maksimalne vjerodostojnosti općenito ne mora postojati, da čak i ako postoji ne mora biti jako konzistentan, te na kraju formuliramo uvjete koji garantiraju egzistenciju i jaku konzistentnost niza procjenitelja maksimalne vjerodostojnosti. U četvrtom poglavlju uvodimo dodatne pretpostavke regularnosti modela, podsjećamo se nekih rezultata iz diferencijalnog računa vektorskih funkcija više varijabli, te nakon toga iskazujemo i dokazujemo teorem koji daje uvjete uz koje postoji jako konzistentan i asimptotski normalan niz korijena jednadžbe vjerodostojnosti. Na kraju četvrtog poglavlja na primjeru pokazujemo potencijalne probleme metode maksimalne vjerodostojnosti te komentiramo kako se navedeni teorem može iskoristiti za dokazivanje konzistentnosti i asimptotske normalnosti niza procjenitelja maksimalne vjerodostojnosti. U zadnjem poglavlju uvodimo pojam asimptotske efikasnosti, motiviramo metodu kojom iz konzistentnog i asimptotski normalnog niza procjenitelja možemo dobiti asimptotski efikasan niz te dokazujemo teorem koji opravdava navedenu metodu. Na kraju na primjeru ilustriramo primjenu navedenog teorema i dajemo primjer superefikasnog niza procjenitelja.

Poglavlje 1

Osnovni pojmovi i rezultati

U ovom poglavlju definiramo osnovne pojmove matematičke statistike i navodimo neke od rezultata iz teorije vjerojatnosti koje ćemo koristiti u ostatku rada.

1.1 Osnovni pojmovi matematičke statistike

Definicija 1.1.1. *Neka je* $(\Omega, \mathcal{F})$ *izmjeriv prostor i* $\mathcal{P}$ *množina vjerojatnosnih mjera na* $(\Omega, \mathcal{F})$*. Tada uređenu trojku* $(\Omega, \mathcal{F}, \mathcal{P})$ *zovemo statistička struktura.*

U ovom radu pretpostavljamo da je množina $\mathcal{P}$ parametrizirana:

$$\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$$

gdje je $\Theta \subset \mathbb{R}^m$, $m \geq 1$. Nadalje, pretpostavljamo da je parametrizacija injektivna:

$$\theta_1 \neq \theta_2 \implies \mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}\ , \quad \text{za sve } \theta_1, \theta_2 \in \Theta.$$

Pokažimo na primjeru dva primjera statističkih struktura.

Primjer 1.1.2.

(i) Stavimo: $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $\Theta = \mathbb{R} \times \langle 0, \infty \rangle$, $\theta = (\mu, \sigma^2)$ *i za* $B \in \mathcal{B}(\mathbb{R})$ *definiramo*

$$\mathbf{P}_\theta(B) = \int_B \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} d\lambda(x).$$

Tada je $\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$ *i* $(\Omega, \mathcal{F}, \mathcal{P})$ *je jedan primjer statističke strukture.*

(ii) $\Omega = \mathbb{Z}_+$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\Theta = \langle 0, \infty \rangle$ *i za* $k \in \Omega$ *definiramo*

$$\mathbf{P}_\theta(\{k\}) = \frac{\theta^k}{k!} e^{-\theta}.$$

Time je za svaki $\theta \in \Theta$ definirana vjerojatnosna mjera $\mathbf{P}_\theta$ na $(\Omega, \mathcal{F})$, pa imamo familiju vjerojatnosnih mjera $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ i $(\Omega, \mathcal{F}, \mathcal{P})$ je statistička struktura.

Definicija 1.1.3. *Slučajni uzorak duljine n na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je konačan niz $X_1, X_2, \ldots, X_n$ slučajnih varijabli na $(\Omega, \mathcal{F})$ takvih da su nezavisne i jednako distribuirane u odnosu na svaku vjerojatnost $\mathbf{P} \in \mathcal{P}$.*

Koristeći sljedeći teorem slijedi da za zadanu familiju gustoća $\{f(\cdot|\theta) : \theta \in \Theta\}$ postoji statistička struktura $(\Omega, \mathcal{F}, \mathcal{P})$, $\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$ i niz slučajnih varijabli $(X_n : n \in \mathbb{N})$ na $(\Omega, \mathcal{F})$, koje su nezavisne i jednako distribuirane u odnosu na svaku vjerojatnost $\mathbf{P}_\theta \in \mathcal{P}$ i imaju gustoću $f(\cdot|\theta)$. Dakle, ima smisla izraz "neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s gustoćom $f(\cdot|\theta)$, $\theta \in \Theta$". Nadalje, možemo $\mathcal{P}$ poistovjetiti s $\mathcal{P}' = \{f(\cdot|\theta) : \theta \in \Theta\}$ pa ćemo reći da je $X_1, \ldots, X_n$ slučajni uzorak iz modela $\mathcal{P}'$ ako je $X_1, \ldots, X_n$ slučajni uzorak na $(\Omega, \mathcal{F}, \mathcal{P})$ i X_i ima gustoću $f(\cdot|\theta)$.

Teorem 1.1.4. *Neka je $(F_n : n \in \mathbb{N})$ proizvoljan niz funkcija distribucije. Tada postoji vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbf{P})$ i niz $(X_n : n \in \mathbb{N})$ nezavisnih slučajnih varijabli na Ω takav da je $F_{X_n} = F_n$.*

Dokaz. Neka su $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$, $n \in \mathbb{N}$ vjerojatnosni prostori, takvi da je za sve $n \in \mathbb{N}$, $\Omega_n = \mathbb{R}$, $\mathcal{F}_n = \mathcal{B}(\mathbb{R})$ i $\mathbf{P}_n = \mathbf{P}_{F_n}$, gdje je $\mathbf{P}_{F_n}$ vjerojatnosna mjera na $\mathcal{B}(\mathbb{R})$ generirana sa F_n. Dakle, vrijedi $P_n(\langle -\infty, x]) = F_n(x)$, $x \in \mathbb{R}$. Neka je $(\Omega, \mathcal{F}, \mathbf{P}) = (\prod_{j=1}^{\infty} \Omega_j, \prod_{j=1}^{\infty} \mathcal{F}_j, \prod_{j=1}^{\infty} \mathbf{P}_j)$ vjerojatnosni prostor koji je jednak produktu vjerojatnosnih prostora $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$, $n \in \mathbb{N}$. Za $n \in \mathbb{N}$ definiramo funkcije $X_n : \Omega \to \mathbb{R}$ tako da stavimo

$$X_n(\omega) = \omega_n \ , \quad \omega = (\omega_1, \omega_2, \ldots) \in \Omega = \mathbb{R}^{\mathbb{N}}.$$

Funkcije X_n, $n \in \mathbb{N}$ su slučajne varijable, tj. one su $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ izmjerive. Naime, za $B \in \mathcal{B}(\mathbb{R})$ imamo $X_n^{-1}(B) = \{\omega \in \Omega : \omega_n \in B\} \in \mathcal{F}$, jer je to izmjeriv pravokutnik (pogledati [14, str. 335]). Za $x \in \mathbb{R}$ vrijedi

$$\begin{aligned} F_{X_n}(x) = \mathbf{P}(X_n \le x) &= \mathbf{P}(\{\omega \in \Omega : \omega_n \in \langle -\infty, x]\}) \\ &= \mathbf{P}_n(\langle -\infty, x]) \\ &= F_n(x), \end{aligned} \tag{1.1}$$

dakle $F_{X_n} = F_n$. Nadalje, za proizvoljni $k \in \mathbb{N}$ i proizvoljne $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R})$ imamo

$$\begin{aligned} \mathbf{P}(X_1 \in B_1, \ldots, X_k \in B_k) &= \mathbf{P}(\{\omega \in \Omega : \omega_1 \in B_1, \ldots, \omega_k \in B_k\}) \\ &= \prod_{j=1}^{k} \mathbf{P}_j(B_j) = \prod_{j=1}^{k} \mathbf{P}(X_j \in B_j). \end{aligned} \tag{1.2}$$

Dakle, $(X_n : n \in \mathbb{N})$ je niz nezavisnih slučajnih varijabli. □

Napomena 1.1.5. *U ovom radu pretpostavljamo da radimo s apsolutno neprekidnim ili diskretnim slučajnim varijablama. Dakle, funkcije $f(\cdot|\theta)$ su gustoće apsolutno neprekidne ili diskretne slučajne varijable.*

Definicija 1.1.6. *Statistika na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je svaka slučajna veličina koja je izmjeriva funkcija nekog slučajnog uzorka na toj statističkoj strukturi.*

Definicija 1.1.7. *Neka je $(X_1, X_2, \ldots, X_n)$ slučajni uzorak iz modela $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$. Bilo koju statistiku $T = t(X_1, X_2, \ldots, X_n)$ koja je iste dimenzije kao i θ nazivamo procjeniteljem od θ.*

Iako se dosta toga može reći o svojstvima procjenitelja za fiksnu duljinu slučajnog uzorka, nas u ovom radu više zanimaju svojstva i ponašanje procjenitelja kada duljina uzorka teži u beskonačnost. Odnosno, od interesa su nam nizovi procjenitelja i njihova asimptotska svojstva. Pretpostavimo da na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ imamo niz $(X_n : n \in \mathbb{N})$ nezavisnih i jednako distribuiranih slučajnih varijabli u odnosu na svaku vjerojatnost $\mathbf{P}_\theta$, $\theta \in \Theta$. Tada je za svaki $n \in \mathbb{N}$, $X_1, X_2, \ldots, X_n$ slučajni uzorak na toj strukturi i za takav niz neka je $T_n = t_n(X_1, X_2, \ldots, X_n)$ niz procjenitelja za θ.

Definicija 1.1.8. *Niz procjenitelja $(T_n : n \in \mathbb{N})$ za θ je:*

(i) jako konzistentan ako vrijedi

$$(\forall \theta \in \Theta) \;\; T_n \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} \theta \,, \quad n \to \infty,$$

(ii) slabo konzistentan ako vrijedi

$$(\forall \theta \in \Theta) \;\; T_n \xrightarrow{\mathbf{P}_\theta} \theta \,, \quad n \to \infty.$$

Napomena 1.1.9. *Često se za slabo konzistentan niz procjenitelja kaže da je konzistentan. Međutim, u ovom radu nas zanima jaka konzistentnost, pa sukladno tome kada kažemo da je niz procjenitelja konzistentan, osim ako nije drugačije naglašeno, mislimo na jaku konzistentnost.*

1.2 Rezultati iz teorije vjerojatnosti

Za početak ćemo dokazati dva rezultata koja su korisna kod dokazivanja konvergencije gotovo sigurno niza slučajnih varijabli.

Propozicija 1.2.1. *Neka je* $(A_n : n \in \mathbb{N})$ *niz događaja na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbf{P})$. *Ako* $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$*, tada je* $\mathbf{P}(A_n \text{ b.m.p.}) = \mathbf{P}(\overline{\lim_n} A_n) = 0$. *Gdje je*

$$\overline{\lim_n} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Dokaz. Za svaki $n \in \mathbb{N}$ vrijedi

$$\overline{\lim_n} A_n \subset \bigcup_{k=n}^{\infty} A_k,$$

odakle slijedi da za svaki $n \in \mathbb{N}$ vrijedi

$$\mathbf{P}(\overline{\lim_n} A_n) \leq \mathbf{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k). \tag{1.3}$$

Desna strana u (1.3) teži u 0 za $n \to \infty$, jer je to ostatak konvergentnog reda, pa slijedi tražena tvrdnja. □

Propozicija 1.2.2. *Neka je* $(X_n : n \in \mathbb{N})$ *niz slučajnih varijabli na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbf{P})$. *Niz* $(X_n)_{n \in \mathbb{N}}$ *konvergira gotovo sigurno prema slučajnoj varijabli* X *ako i samo ako za svaki* $\varepsilon > 0$ *vrijedi*

$$\lim_{n\to\infty} \mathbf{P}\left(\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}\right) = 0.$$

Dokaz. Neka je $\varepsilon > 0$. Za $n \in \mathbb{N}$ stavimo

$$A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\} \;\text{ i }\; A(\varepsilon) = \overline{\lim_n} A_n(\varepsilon).$$

Tada je $A(\varepsilon)$ presjek padajućeg niza događaja $\cup_{k=n}^{\infty} A_k(\varepsilon)$, $n \in \mathbb{N}$, pa je zbog neprekidnosti vjerojatnosti u odnosu na padajući niz događaja

$$\mathbf{P}(A(\varepsilon)) = \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{k=n}^{\infty} A_k(\varepsilon)\right).$$

Neka je $D = \{\omega \in \Omega : X_n \not\to X \text{ za } n \to \infty\}$. Tada je

$$D = \bigcup_{\varepsilon > 0} A(\varepsilon) = \bigcup_{m=1}^{\infty} A\left(\frac{1}{m}\right),$$

jer je $A(\varepsilon_1) \subset A(\varepsilon_2)$ za $\varepsilon_1 > \varepsilon_2$. Dakle, vrijedi

$$X_n \xrightarrow{g.s.} X \iff \mathbf{P}(D) = 0 \iff (\forall \varepsilon > 0)\, \mathbf{P}(A(\varepsilon)) = 0 \iff$$
$$(\forall \varepsilon > 0) \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}\right) = 0.$$

□

Sljedeća dva teorema pokazuju kako se s limesom može ući pod integral.

Teorem 1.2.3 (Lebesgueov teorem o monotonoj konvergenciji)**.** *Neka je $(X_n : n \in \mathbb{N})$ rastući niz nenegativnih slučajnih varijabli i neka $(g.s.)$ $\lim_{n\to\infty} X_n = X$. Tada je*

$$\lim_{n\to\infty} \mathbf{E}X_n = \mathbf{E}X. \tag{1.4}$$

Za dokaz prethodnog teorema pogledati [14].

Korolar 1.2.4. *Neka je $(X_n : n \in \mathbb{N})$ niz nenegativnih slučajnih varijabli. Tada je*

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbf{E}X_n. \tag{1.5}$$

Dokaz. Niz $\left(\sum_{k=1}^{n} X_k : n \in \mathbb{N}\right)$ je rastući i konvergira prema $\sum_{n=1}^{\infty} X_n$, pa primjenom Lebesgueovog teorema o monotonoj konvergenciji slijedi

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n\right) = \lim_{n\to\infty} \mathbf{E} \sum_{k=1}^{n} X_k = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{E}X_k = \sum_{n=1}^{\infty} \mathbf{E}X_n.$$

□

Dokaz sljedećeg teorema se također može naći u [14].

Teorem 1.2.5 (Lebesgueov teorem o dominiranoj konvergenciji)**.** *Neka je $(X_n : n \in \mathbb{N})$ niz slučajnih varijabli takav da je $(g.s.)$ $\lim_{n\to\infty} X_n = X$ i neka je $|X_n| \leq Y$ za sve n, pri čemu je $\mathbf{E}|Y| < \infty$. Tada je*

$$\lim_{n\to\infty} \mathbf{E}X_n = \mathbf{E}X. \tag{1.6}$$

Za zadani niz slučajnih varijabli $(X_n : n \in \mathbb{N})$ neka je $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$ pripadni niz parcijalnih suma i $\overline{X}_n = \frac{S_n}{n}$, $n \in \mathbb{N}$ pripadni niz parcijalnih aritmetičkih sredina. Sljedeći teorem, čiji se dokaz može naći u [14], daje nužne i dovoljne uvjete za $(g.s.)$ konvergenciju niza $(\overline{X}_n : n \in \mathbb{N})$.

Teorem 1.2.6 (Kolmogorov)**.** *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli. Tada niz parcijalnih aritmetičkih sredina $(\overline{X}_n : n \in \mathbb{N})$ gotovo sigurno konvergira konačnom limesu ako i samo ako je $\mathbf{E}|X_1| < \infty$ i u tom slučaju je*

$$(g.s.)\ \lim_{n\to\infty} \overline{X}_n = \mathbf{E}X_1. \tag{1.7}$$

Od koristi može biti i sljedeće poopćenje prethodnog teorema na koje ćemo se pozivati kada kažemo da koristimo jaki zakon velikih brojeva. Prije iskaza teorema podsjetimo se da matematičko očekivanje slučajne varijable X postoji ako je $\mathbf{E}X^+ < +\infty$ ili $\mathbf{E}X^- < +\infty$ i u tom slučaju je $\mathbf{E}X = \mathbf{E}X^+ - \mathbf{E}X^- \in \mathbb{R} \cup \{-\infty, +\infty\}$.

Teorem 1.2.7. *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli. Ako postoji matematičko očekivanje $\mathbf{E}X_1 \in \mathbb{R} \cup \{-\infty, +\infty\}$, tada je*

$$(g.s.)\ \lim_{n\to\infty} \overline{X}_n = \mathbf{E}X_1. \tag{1.8}$$

Dokaz. Ako je $\mathbf{E}|X_1| < \infty$ tvrdnja teorema slijedi iz teorema 1.2.6. Dakle, preostaje slučaj $\mathbf{E}X_1 = \pm\infty$. Kako je $(-X)^+ = X^-$ i $(-X)^- = X^+$, bez smanjenja općenitosti možemo pretpostaviti da je $\mathbf{E}X_1^+ = +\infty$ i $\mathbf{E}X_1^- < +\infty$. Za $k > 0$ definiramo $X_n^{(k)} = \min\{X_n, k\}$. Niz $\left(X_n^{(k)}\right)_{n\in\mathbb{N}}$ je niz nezavisnih jednako distribuiranih slučajnih varijabli i $\mathbf{E}|X_1^{(k)}| < \infty$, pa iz teorema 1.2.6 slijedi $\overline{X^{(k)}}_n \xrightarrow{g.s.} \mathbf{E}X_1^{(k)}$. Kako je $X_n^{(k)} \le X_n$ imamo

$$\liminf_{n\to\infty} \overline{X}_n \ge \liminf_{n\to\infty} \overline{X^{(k)}}_n = \mathbf{E}X_1^{(k)}\ \ (g.s.). \tag{1.9}$$

Primjenom Lebesgueovog teorema o monotonoj konvergenciji slijedi

$$\lim_{k\to\infty} \mathbf{E}\left(X_1^{(k)}\right)^+ = \mathbf{E}X_1^+ = +\infty, \tag{1.10}$$

pa kako je $\left(X_1^{(k)}\right)^- = X_1^-$ slijedi

$$\lim_{k\to\infty} \mathbf{E}X_1^{(k)} = \lim_{k\to\infty}\left(\mathbf{E}\left(X_1^{(k)}\right)^+ - \mathbf{E}X_1^-\right) = +\infty. \tag{1.11}$$

Tvrdnja teorema sada slijedi iz (1.9) puštanjem $k \to +\infty$. □

Često nam je od interesa asimptotska razdioba niza procjenitelja, odnosno promatramo konvergenciju po distribuciji niza slučajnih vektora ili varijabli. Sljedeća dva teorema su od velike važnosti za rješavanje takvih problema. Dokazi oba teorema se mogu naći u [4].

Teorem 1.2.8 (Centralni granični teorem)**.** *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih jednako distribuiranih slučajnih vektora s očekivanjem $\boldsymbol{\mu}$ i konačnom kovarijacijskom matricom $\boldsymbol{\Sigma}$. Tada vrijedi*

$$\sqrt{n}\ \left(\overline{X}_n - \boldsymbol{\mu}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),\ \ n \to \infty. \tag{1.12}$$

Teorem 1.2.9 (Slutsky)**.** *Neka je $(X_n : n \in \mathbb{N})$ niz slučajnih vektora. Tada vrijedi sljedeće*

(i) Ako je $X_n \in \mathbb{R}^d$, $X_n \xrightarrow{D} X$ i ako je $f : \mathbb{R}^d \to \mathbb{R}^k$ takva da je $\mathbf{P}(X \in C(f)) = 1$, gdje je $C(f)$ skup na kojem je f neprekidna, tada

$$f(X_n) \xrightarrow{D} f(X).$$

(ii) Ako $X_n \xrightarrow{D} X$ i $(X_n - Y_n) \xrightarrow{P} 0$ tada

$$Y_n \xrightarrow{D} X.$$

(iii) Ako $X_n \in \mathbb{R}^d$, $Y_n \in \mathbb{R}^k$, $X_n \xrightarrow{D} X$ i $Y_n \xrightarrow{D} c$, gdje je c konstanta, tada

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} X \\ c \end{pmatrix}.$$

Definicija 1.2.10. *Neka su $(X_n : n \in \mathbb{N})$ i $(Y_n : n \in \mathbb{N})$ nizovi slučajnih vektora. Kažemo da su $(X_n)_{n\in\mathbb{N}}$ i $(Y_n)_{n\in\mathbb{N}}$ asimptotski ekvivalentni ako $(X_n - Y_n) \xrightarrow{\mathbf{P}} 0$.*

U nastavku pretpostavljamo da je $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ euklidska norma na $\mathbb{R}^d$.

Definicija 1.2.11. *Neka je $(X_n : n \in \mathbb{N})$ niz slučajnih vektora. Kažemo da je niz $(X_n)_{n\in\mathbb{N}}$ ograničen po vjerojatnosti ako za svaki $\varepsilon > 0$ postoje $M_\varepsilon > 0$ i $N_\varepsilon \in \mathbb{N}$ takvi da za svaki $n \in \mathbb{N}$, $n \geq N_\varepsilon$ vrijedi*

$$\mathbf{P}(\|X_n\| > M_\varepsilon) \leq \varepsilon.$$

Propozicija 1.2.12. *Neka je $(X_n : n \in \mathbb{N})$ niz slučajnih vektora. Ako $(X_n)_{n\in\mathbb{N}}$ konvergira po distribuciji prema slučajnom vektoru X, tada je $(X_n)_{n\in\mathbb{N}}$ ograničen po vjerojatnosti.*

Dokaz. Funkcija $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ je neprekidna, pa koristeći teorem 1.2.9, slijedi

$$Y_n := \|X_n\| \xrightarrow{D} \|X\| =: Y.$$

Neka je $\varepsilon > 0$. Kako je skup $\mathbb{R} \setminus C(F_Y)$ najviše prebrojiv, postoji $M_\varepsilon > 0$ dovoljno velik, takav da je $\mathbf{P}(Y > M_\varepsilon) < \frac{\varepsilon}{2}$ i $M_\varepsilon \in C(F_Y)$. Zbog $Y_n \xrightarrow{D} Y$ i $M_\varepsilon \in C(F_Y)$ slijedi da postoji $N_\varepsilon \in \mathbb{N}$, takav da za svaki $n \in \mathbb{N}$, $n \geq N_\varepsilon$ vrijedi

$$\mathbf{P}(Y_n > M_\varepsilon) \leq \mathbf{P}(Y > M_\varepsilon) + \frac{\varepsilon}{2} \leq \varepsilon.$$

Dakle, $(X_n)_{n\in\mathbb{N}}$ je ograničen po vjerojatnosti. □

Propozicija 1.2.13. *Neka su* $(X_n : n \in \mathbb{N})$ *i* $(Y_n : n \in \mathbb{N})$ *nizovi slučajnih vektora iste dimenzije. Ako su* $(X_n)_{n\in\mathbb{N}}$ *i* $(Y_n)_{n\in\mathbb{N}}$ *ograničeni po vjerojatnosti, tada je i niz* $(X_n + Y_n)_{n\in\mathbb{N}}$ *ograničen po vjerojatnosti.*

Dokaz. Neka je $\varepsilon > 0$. Dani nizovi su ograničeni po vjerojatnosti, pa postoje $M_1 > 0$, $M_2 > 0$, $N_1 \in \mathbb{N}$ i $N_2 \in \mathbb{N}$, takvi da za svaki $n \in \mathbb{N}$ vrijedi:

$$n \geq N_1 \implies \mathbf{P}(\|X_n\| > M_1) \leq \frac{\varepsilon}{2}, \tag{1.13}$$

$$n \geq N_2 \implies \mathbf{P}(\|Y_n\| > M_2) \leq \frac{\varepsilon}{2}. \tag{1.14}$$

Stavimo $M = 2\max\{M_1, M_2\}$ i $N = \max\{N_1, N_2\}$. Tada za $n \in \mathbb{N}$, $n \geq N$ vrijedi

$$\begin{aligned}
\mathbf{P}(\|X_n + Y_n\| > M) &\leq \mathbf{P}\left(\|X_n\| > \frac{M}{2}\right) + \mathbf{P}\left(\|Y_n\| > \frac{M}{2}\right) \\
&\leq \mathbf{P}(\|X_n\| > M_1) + \mathbf{P}(\|Y_n\| > M_2) \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad (1.15)
\end{aligned}$$

Dakle, niz $(X_n + Y_n)_{n\in\mathbb{N}}$ je ograničen po vjerojatnosti. □

Poglavlje 2

Uniformni jaki zakon velikih brojeva

U ovom poglavlju pretpostavljamo da imamo niz $(X_n : n \in \mathbb{N})$ nezavisnih i jednako distribuiranih slučajnih varijabli na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$, $\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$, te da je $x \mapsto U(x, \theta) \in \mathbb{R}$ izmjeriva funkcija za sve θ iz parametarskog prostora $\Theta \subset \mathbb{R}^m$. Fiksirajmo $\theta_0 \in \Theta$. Statistika od interesa nam je $\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta)$. Ako pretpostavimo da

$$\mu(\theta) := \mathbf{E}_{\theta_0}(U(X, \theta)) = \int U(x, \theta)\, dF(x|\theta_0)$$

postoji i konačno je za sve $\theta \in \Theta$, tada po jakom zakonu velikih brojeva slijedi

$$(\forall \theta \in \Theta)\ \ \frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) \xrightarrow{g.s.} \mu(\theta)\ ,\ \ n \to \infty. \tag{2.1}$$

Sada želimo pojačati prethodnu tvrdnju tako da je konvergencija uniformna po θ u smislu da

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta) \right| \xrightarrow{g.s.} 0\ ,\ \ n \to \infty. \tag{2.2}$$

Da bi pokazali kako (2.2) može biti od koristi pretpostavimo da je $(\hat{\theta}_n : n \in \mathbb{N})$ niz takav da $\hat{\theta}_n \xrightarrow{g.s.} \theta_0$, $n \to \infty$, te da je funkcija μ neprekidna. Htjeli bi dobiti da tada vrijedi

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) \xrightarrow{g.s.} \mu(\theta_0)\ ,\ \ n \to \infty. \tag{2.3}$$

Općenito (2.1) nije dovoljno za pokazati (2.3), ali koristeći (2.2) tvrdnja se lako pokaže:

$$\begin{aligned} \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\theta_0) \right| &\leq \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\hat{\theta}_n) \right| + \left| \mu(\hat{\theta}_n) - \mu(\theta_0) \right| \\ &\leq \sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta) \right| + \left| \mu(\hat{\theta}_n) - \mu(\theta_0) \right|, \qquad (2.4) \end{aligned}$$

pa kako je μ neprekidna i $\hat{\theta}_n \xrightarrow{g.s.} \theta_0$ te koristeći (2.2) slijedi (2.3).
Sljedeći teorem daje uvjete uz koje vrijedi (2.2).

Teorem 2.0.1. *Fiksirajmo $\theta_0 \in \Theta$ i pretpostavimo da vrijedi sljedeće:*

(a) *Θ je kompaktan skup,*

(b) *$\theta \mapsto U(x,\theta)$ je neprekidna za svaki x,*

(c) *postoji izmjeriva funkcija $x \mapsto K(x)$ takva da je $\mathbf{E}_{\theta_0}|K(X)| < \infty$ i $|U(x,\theta)| \leq K(x)$ za sve x i θ.*

Tada je:

(i) $\limsup\limits_{n\to\infty} \sup\limits_{\theta\in\Theta} \frac{1}{n}\sum\limits_{i=1}^{n} U(X_i,\theta) \leq \sup\limits_{\theta\in\Theta} \mu(\theta) \quad (\mathbf{P}_{\theta_0}\text{- } g.s.)$

(ii) $\lim\limits_{n\to\infty} \sup\limits_{\theta\in\Theta} \left| \frac{1}{n}\sum\limits_{i=1}^{n} U(X_i,\theta) - \mu(\theta) \right| = 0 \quad (\mathbf{P}_{\theta_0}\text{- } g.s.)$

Dokaz.

(i) Za $\rho > 0$ definirajmo

$$\varphi(x,\theta,\rho) = \sup_{|\theta'-\theta|<\rho} U(x,\theta').$$

Tada je $|\varphi(x,\theta,\rho)| \leq K(x)$ i $\lim\limits_{\rho\to 0} \varphi(x,\theta,\rho) = U(x,\theta)$, jer je $\theta \mapsto U(x,\theta)$ neprekidna. Primjenom Lebesgueovog teorema o dominiranoj konvergenciji slijedi

$$\lim_{\rho\to 0} \int \varphi\,(x,\theta,\rho)\; dF(x|\theta_0) = \int U\,(x,\theta)\; dF(x|\theta_0) = \mu(\theta). \tag{2.5}$$

Neka je $\varepsilon > 0$. Iz (2.5) slijedi da za svaki $\theta \in \Theta$ postoji $\rho_\theta > 0$ takav da je

$$\int \varphi(x,\theta,\rho_\theta)\, dF(x|\theta_0) < \mu(\theta) + \varepsilon. \tag{2.6}$$

Kugle $B(\theta,\rho_\theta) = \{\theta' : |\theta'-\theta| < \rho_\theta\}$, $\theta \in \Theta$, pokrivaju Θ. Kako je skup Θ kompaktan postoji konačan podpokrivač, odnosno postoje $m \in \mathbb{N}$ i $\theta_j \in \Theta$, $j \in \{1,2,\ldots,m\}$ takvi da je $\Theta \subset \bigcup_{j=1}^{m} B(\theta_j,\rho_{\theta_j})$. Dakle, za svaki $\theta \in \Theta$ postoji $j \in \{1,2,\ldots,m\}$ takav da je $\theta \in B(\theta_j,\rho_{\theta_j})$, pa iz definicije od φ slijedi da je $U(x,\theta) \leq \varphi(x,\theta_j,\rho_{\theta_j})$ za svaki x. Dakle, vrijedi

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta) \leq \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i,\theta_j,\rho_{\theta_j}) \leq \sup_{1\leq j\leq m} \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i,\theta_j,\rho_{\theta_j}),$$

odakle slijedi

$$\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}U(X_i,\theta)\leq\sup_{1\leq j\leq m}\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i,\theta_j,\rho_{\theta_j}).\tag{2.7}$$

Sada primjenom jakog zakona velikih brojeva na niz $(\varphi(X_n,\theta_j,\rho_{\theta_j}))_{n\in\mathbb{N}}$ i koristeći (2.6) slijedi

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i,\theta_j,\rho_{\theta_j})\leq\mu(\theta_j)+\varepsilon\ (g.s.)\ ,\quad \text{za } j=1,2,\ldots,m.\tag{2.8}$$

Iz (2.8) zaključujemo da vrijedi

$$\begin{aligned}\limsup_{n\to\infty}\sup_{1\leq j\leq m}\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i,\theta_j,\rho_{\theta_j}) &\leq \sup_{1\leq j\leq m}\mu(\theta_j)+\varepsilon\ \ (g.s.)\\ &\leq \sup_{\theta\in\Theta}\mu(\theta)+\varepsilon,\end{aligned}$$

pa zbog (2.7) imamo

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}U(X_i,\theta)\leq\sup_{\theta\in\Theta}\mu(\theta)+\varepsilon\ \ (g.s.).\tag{2.9}$$

Kako (2.9) vrijedi za svaki $\varepsilon>0$ slijedi tražena tvrdnja.

(ii) Kako je $|U(x,\theta)|\leq K(x)$ koristeći Lebesgueov teorem o dominiranoj konvergenciji imamo

$$\lim_{\theta'\to\theta}\mu(\theta')=\lim_{\theta'\to\theta}\int U(x,\theta')\,dF(x|\theta_0)=\int U(x,\theta)\,dF(x|\theta_0)=\mu(\theta).$$

Dakle, μ je neprekidna funkcija, pa je onda i $\tilde{U}(x,\theta):=U(x,\theta)-\mu(\theta)$ neprekidna po θ za svaki x. Nadalje, vrijedi $|\tilde{U}(x,\theta)|\leq K(x)+\mathbf{E}_{\theta_0}(K(X))$, pa primjenom prethodno dokazanog na funkcije $\tilde{U}$ i $-\tilde{U}$ dobijemo

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\tilde{U}(X_i,\theta)\leq 0\ (g.s.)\ \text{ i }\ \limsup_{n\to\infty}\sup_{\theta\in\Theta}-\frac{1}{n}\sum_{i=1}^{n}\tilde{U}(X_i,\theta)\leq 0\ (g.s.).\tag{2.10}$$

Tražena tvrdnja sada slijedi iz (2.10) koristeći da za proizvoljnu funkciju f vrijedi:

$$0\leq\sup_x\left|f(x)\right|=\max\{\sup_x f(x),\ \sup_x -f(x)\}.\tag{2.11}$$

Naime, koristeći (2.11) dobijemo

$$0 \leq \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{U}(X_i, \theta) \right| = \max \left\{ \sup_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \tilde{U}(X_i, \theta),\ \sup_{\theta\in\Theta} -\frac{1}{n} \sum_{i=1}^{n} \tilde{U}(X_i, \theta) \right\},$$

odakle koristeći (2.10) slijedi

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{U}(X_i, \theta) \right| = 0 \ \ (g.s.) \implies \lim_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{U}(X_i, \theta) \right| = 0 \ \ (g.s.).$$

$\square$

Poglavlje 3

Konzistentnost procjenitelja maksimalne vjerodostojnosti

U ovom poglavlju definiramo procjenitelj maksimalne vjerodostojnosti i pokazujemo na primjerima da procjenitelj maksimalne vjerodostojnosti općenito ne mora postojati, te da čak i ako postoji, ne mora nužno biti jako konzistentan. Na kraju dokazujemo teorem koji daje uvjete uz koje niz procjenitelja maksimalne vjerodostojnosti postoji i konzistentan je.

Definicija 3.0.1. *Neka je* $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$ *realizacija slučajnog uzorka* $(X_1, X_2, \ldots, X_n)$ *iz modela* $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$*. Funkcija vjerodostojnosti je funkcija* $L_n : \Theta \to \mathbb{R}$ *definirana sa*

$$L_n(\theta) = L_n(\theta|\mathbf{x}_n) := \prod_{i=1}^{n} f(x_i|\theta).$$

Primijetimo da u diskretnim modelima za funkciju vjerodostojnosti vrijedi

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \mathbf{P}_\theta(X_i = x_i) = \mathbf{P}_\theta(X_1 = x_1, \ldots, X_n = x_n).$$

Dakle, u tom slučaju je funkcija vjerodostojnosti zapravo vjerojatnost da je realizacija slučajnog uzorka $(X_1, X_2, \ldots, X_n)$ jednaka $\mathbf{x}_n$, uz pretpostavku da je θ prava vrijednost parametra. Stoga se čini prirodno parametar θ procjenjivati procjeniteljem $\hat{\theta}_n$ koji će maksimizirati funkciju vjerodostojnosti. Upravo tako ćemo i definirati procjenitelj maksimalne vjerodostojnosti u općenitom slučaju.

Definicija 3.0.2. *Procjenitelj maksimalne vjerodostojnosti je funkcija* $\hat{\theta}_n := \hat{\theta}_n(X_1, \ldots, X_n)$ *takva da je*

$$L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta).$$

Napomenimo da je traženje točke maksimuma funkcije vjerodostojnosti L_n ekvivalentno traženju točke maksimuma funkcije $l_n(\theta) = \log(L_n(\theta))$. Funkciju l_n nazivamo log-vjerodostojnost. U sljedećem primjeru ilustriramo procjenu parametra metodom maksimalne vjerodostojnosti.

Primjer 3.0.3. *(i) Neka je $(x_1, x_2, \ldots, x_n)$ realizacija slučajnog uzorka $(X_1, X_2, \ldots, X_n)$ iz eksponencijalnog modela s parametrom $\lambda > 0$, tj. $f(x|\lambda) = \lambda e^{-\lambda x}\mathbf{1}_{\langle 0,+\infty\rangle}(x)$ i $\Theta = \langle 0, +\infty\rangle$. Funkcija vjerodostojnosti je*

$$L_n(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum\limits_{i=1}^{n} x_i}.$$

Promatramo log-vjerodostojnost

$$l_n(\lambda) = \log(L_n(\lambda)) = n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i. \tag{3.1}$$

Dakle, tražimo točku maksimuma funkcije l_n na skupu $\langle 0, +\infty\rangle$. Deriviranjem funkcije l_n i izjednačavanjem derivacije s 0 dobivamo

$$\dot{l}_n = 0 \implies \frac{n}{\lambda} - \sum_{i=0}^{n} x_i = 0 \implies \hat{\lambda} = \frac{n}{\sum\limits_{i=1}^{n} x_i}.$$

Pogledajmo drugu derivaciju

$$\ddot{l}_n(\lambda) = -\frac{n}{\lambda^2} < 0 \text{ za sve } \lambda \in \langle 0, +\infty\rangle.$$

Dakle, točka $\hat{\lambda}$ je točka globalnog maksimuma funkcije l_n, pa je procjenitelj maksimalne vjerodostojnosti za λ jednak $\hat{\lambda}_n = \frac{1}{\overline{X}_n}$. Primijetimo da je u ovom slučaju niz procjenitelja maksimalne vjerodostojnosti jako konzistentan. Naime, iz jakog zakona velikih brojeva slijedi

$$\overline{X}_n \xrightarrow{\mathbf{P}_\lambda\text{-}g.s.} \mathbf{E}_\lambda(X_1) = \frac{1}{\lambda} \implies \hat{\lambda}_n = \frac{1}{\overline{X}_n} \xrightarrow{\mathbf{P}_\lambda\text{-}g.s.} \lambda.$$

(ii) Neka je $(x_1, x_2, \ldots, x_n)$ realizacija slučajnog uzorka $(X_1, X_2, \ldots, X_n)$ iz uniformnog modela $U(0, \theta)$, $\theta \in \langle 0, +\infty\rangle$, tj. $f(x|\theta) = \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x)$ i $\Theta = \langle 0, +\infty\rangle$. Funkcija vjerodostojnosti je

$$L_n(\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\mathbf{1}_{[x_{(n)},+\infty\rangle}(\theta),$$

gdje je $x_{(n)} = \max\{x_1, x_2, \ldots, x_n\}$. *Kako je* $\theta \mapsto \frac{1}{\theta^n}$ *padajuća funkcija na* $\langle 0, +\infty\rangle$, *slijedi da je procjenitelj maksimalne vjerodostojnosti* $\hat{\theta}_n = X_{(n)}$. *U ovom primjeru se isto lako pokaže da je niz procjenitelja maksimalne vjerodostojnosti jako konzistentan. Naime, po propoziciji 1.2.2 dovoljno je pokazati da za svaki* $\varepsilon > 0$ *vrijedi*

$$\lim_{n\to\infty} \mathbf{P}_\theta\left(\bigcup_{k=n}^{\infty}\{|X_{(k)} - \theta| \geq \varepsilon\}\right) = 0, \tag{3.2}$$

jer iz toga slijedi

$$\hat{\theta}_n = X_{(n)} \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} \theta.$$

Neka je $\varepsilon > 0$. *Za svaki* $k \in \mathbb{N}$ *je*

$$\begin{aligned} \mathbf{P}_\theta\left(|X_{(k)} - \theta| \geq \varepsilon\right) &= \mathbf{P}_\theta\left(X_{(k)} \geq \theta + \varepsilon\right) + \mathbf{P}_\theta\left(X_{(k)} \leq \theta - \varepsilon\right) \\ &= \mathbf{P}_\theta\left(X_{(k)} \leq \theta - \varepsilon\right) \\ &= \mathbf{P}_\theta(X_1 \leq \theta - \varepsilon, \ldots, X_k \leq \theta - \varepsilon) \\ &= \left(\frac{\theta - \varepsilon}{\theta}\right)^k. \end{aligned} \tag{3.3}$$

Pokažimo da vrijedi (3.2). *Koristeći* (3.3) *slijedi:*

$$\begin{aligned} \mathbf{P}_\theta\left(\bigcup_{k=n}^{\infty}\{|X_{(k)} - \theta| \geq \varepsilon\}\right) &\leq \sum_{k=n}^{\infty} \mathbf{P}_\theta\left(|X_{(k)} - \theta| \geq \varepsilon\right) \\ &= \sum_{k=n}^{\infty} \mathbf{P}_\theta(X_{(k)} \leq \theta - \varepsilon) \\ &= \sum_{k=n}^{\infty}\left(\frac{\theta - \varepsilon}{\theta}\right)^k \xrightarrow{n\to\infty} 0. \end{aligned} \tag{3.4}$$

U prethodnom primjeru procjenitelj maksimalne vjerodostojnosti je postojao i niz procjenitelja je bio jako konzistentan, međutim to općenito ne mora biti tako. Sljedeći primjer daje situaciju u kojoj procjenitelj maksimalne vjerodostojnosti ne postoji jer funkcija vjerodostojnosti nije ograničena na parametarskom skupu.

Primjer 3.0.4. *Neka je* $(x_1, x_2, \ldots, x_n)$ *realizacija slučajnog uzorka* $(X_1, X_2, \ldots, X_n)$ *iz modela* $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$, *gdje je*

$$f(x|\mu, \sigma^2) = \frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + \frac{1}{2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ \theta = (\mu, \sigma^2)\ i\ \Theta = \mathbb{R}\times\langle 0, +\infty\rangle.$$

Funkcija vjerodostojnosti je

$$L_n(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2).$$

Ako stavimo $\mu = x_1$, vidimo da $L_n(x_1, \sigma^2)$ može biti po volji veliko za dovoljno male $\sigma^2 > 0$. Naime, prvi član u produktu raste prema $+\infty$ kada σ^2 ide prema 0, a ostali članovi u produktu su ograničeni i pozitivni, pa slijedi da $L_n(x_1, \sigma^2) \nearrow +\infty$ kada $\sigma^2 \searrow 0$. Dakle, funkcija vjerodostojnosti nije ograničena odozgo, pa procjenitelj maksimalne vjerodostojnosti u ovom primjeru ne postoji.

Napomenimo da procjenitelj maksimalne vjerodostojnosti sigurno postoji ako je parametarski prostor Θ kompaktan i ako je $\theta \mapsto f(x|\theta)$ neprekidna za svaki x. Sljedeći primjer daje situaciju u kojoj procjenitelj maksimalne vjerodostojnosti postoji, ali niz procjenitelja maksimalne vjerodostojnosti nije jako konzistentan.

Primjer 3.0.5. *Promatramo slučajne uzorke $(X_1, X_2, \dots, X_n)$, $n \in \mathbb{N}$ iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$, gdje je $\Theta = [0, 1]$ i $f(\cdot|\theta)$ gustoća na $[-1, 1]$ zadana sa*

$$f(x|\theta) = \frac{(1-\theta)}{\delta(\theta)}\left(1 - \frac{|x-\theta|}{\delta(\theta)}\right)\mathbf{1}_{[\theta-\delta(\theta),\theta+\delta(\theta)]}(x) + \frac{\theta}{2}\mathbf{1}_{[-1,1]}(x),$$

gdje je $\delta(\theta) = (1-\theta)e^{-(1-\theta)^{-4}+1}$ i $f(x|1) = \frac{1}{2}\mathbf{1}_{[-1,1]}(x)$. Kako je Θ kompaktan i funkcija $\theta \mapsto f(x|\theta)$ je neprekidna, procjenitelj maksimalne vjerodostojnosti postoji za svaki $n \in \mathbb{N}$. Neka je $(\hat{\theta}_n)_{n\in\mathbb{N}}$ niz procjenitelja maksimalne vjerodostojnosti. Dakle, za svaki $n \in \mathbb{N}$, $\hat{\theta}_n$ je vrijednost u kojoj funkcija

$$l_n(\theta) = \log(L_n(\theta)) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$

postiže maksimum. Pokazat ćemo da za svaki $\theta \in \Theta$, $\hat{\theta}_n \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} 1$. Neka je $\alpha \in \langle 0, 1\rangle$. Kako je za $\theta < 1$

$$f(x|\theta) \le \frac{(1-\theta)}{\delta(\theta)} + \frac{\theta}{2} < \frac{1}{\delta(\theta)} + \frac{1}{2},$$

slijedi da je

$$\max_{0\le\theta\le\alpha} \frac{1}{n} l_n(\theta) \le \log\left(\frac{1}{\delta(\alpha)} + \frac{1}{2}\right).$$

Dakle, dovoljno je pokazati da za svaki $\theta_0 \in \Theta$ vrijedi

$$\max_{0\le\theta\le 1} \frac{1}{n} l_n(\theta) \xrightarrow{\mathbf{P}_{\theta_0}\text{-}g.s.} +\infty, \tag{3.5}$$

jer tada za svaki $\alpha \in \langle 0,1\rangle$ *za dovoljno velike* $n \in \mathbb{N}$ *vrijedi* $\alpha < \hat{\theta}_n \leq 1$*, odakle slijedi* $\hat{\theta}_n \xrightarrow{\mathbf{P}_{\theta_0}\text{-}g.s.} 1$*. Neka je* $M_n = \max\{X_1, X_2, \dots, X_n\}$*. Pokažimo da za svaki* $\theta_0 \in \Theta$ *vrijedi*

$$n^{\frac{1}{4}}(1 - M_n) \xrightarrow{\mathbf{P}_{\theta_0}\text{-}g.s.} 0. \tag{3.6}$$

Neka je $\theta_0 \in \Theta$*. Za dovoljno male* $\varepsilon > 0$ *je*

$$\mathbf{P}_{\theta_0}(X_1 \leq 1 - \varepsilon) \leq \mathbf{P}_0(X_1 \leq 1 - \varepsilon),$$

pa slijedi:

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}_{\theta_0}\left(n^{\frac{1}{4}}(1 - M_n) \geq \varepsilon\right) &= \sum_{n=1}^{\infty} \mathbf{P}_{\theta_0}\left(M_n \leq 1 - \frac{\varepsilon}{n^{\frac{1}{4}}}\right) \\
&= \sum_{n=1}^{\infty} \left(\mathbf{P}_{\theta_0}\left(X_1 \leq 1 - \frac{\varepsilon}{n^{\frac{1}{4}}}\right)\right)^n \\
&\leq \sum_{n=1}^{\infty} \left(\mathbf{P}_{0}\left(X_1 \leq 1 - \frac{\varepsilon}{n^{\frac{1}{4}}}\right)\right)^n \\
&= \sum_{n=1}^{\infty} \left(1 - \frac{\varepsilon^2}{2\sqrt{n}}\right)^n \\
&\leq \sum_{n=1}^{\infty} \left(e^{-\frac{\varepsilon^2}{2\sqrt{n}}}\right)^n = \sum_{n=1}^{\infty} e^{-\frac{\varepsilon^2\sqrt{n}}{2}} < \infty.
\end{aligned} \tag{3.7}$$

Red $\sum_{n=1}^{\infty} e^{-\frac{\varepsilon^2\sqrt{n}}{2}}$ *konvergira, jer je za dovoljno velike* $n \in \mathbb{N}$, $e^{-\frac{\varepsilon^2\sqrt{n}}{2}} \leq \frac{1}{n^2}$ *i red* $\sum_{n=1}^{\infty} \frac{1}{n^2}$ *konvergira. Iz* (3.7) *koristeći propoziciju 1.2.1 slijedi* $\mathbf{P}_{\theta_0}\left(n^{\frac{1}{4}}(1 - M_n) \geq \varepsilon \text{ b.m.p.}\right) = 0$*, odakle slijedi* (3.6)*. Sada vrijedi*

$$\max_{0\leq\theta\leq 1} \frac{1}{n} l_n(\theta) \geq \frac{1}{n} l_n(M_n) \geq \frac{n-1}{n} \log\left(\frac{M_n}{2}\right) + \frac{1}{n}\log\left(\frac{1 - M_n}{\delta(M_n)}\right),$$

odakle slijedi

$$\liminf_{n\to\infty} \max_{0\leq\theta\leq 1} \frac{1}{n} l_n(\theta) \geq \liminf_{n\to\infty} \frac{1}{n}\log\left(\frac{1 - M_n}{\delta(M_n)}\right) - \log(2) \ \ (\mathbf{P}_{\theta_0}\text{-}g.s.). \tag{3.8}$$

Nadalje, zbog (3.6) *vrijedi*

$$\frac{1}{n}\log\left(\frac{1 - M_n}{\delta(M_n)}\right) = \frac{1}{n(1 - M_n)^4} - \frac{1}{n} \xrightarrow{\mathbf{P}_{\theta_0}\text{-}g.s.} +\infty. \tag{3.9}$$

Tvrdnja (3.5) *sada slijedi iz* (3.8) *i* (3.9)*.*

Definirajmo sada Kullback-Leibler informaciju i iskažimo rezultat koji će nam trebati za dokazivanje konzistentnosti procjenitelja maksimalne vjerodostojnosti. Također, u nastavku pretpostavljamo da za model koji promatramo $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ nosač gustoće $\mathrm{supp} f(\cdot|\theta) = \{x : f(x|\theta) > 0\}$ ne ovisi o $\theta \in \Theta$.

Definicija 3.0.6. *Neka su f_0 i f_1 gustoće obzirom na σ-konačnu mjeru μ. Kullback-Leibler informacija od f_0 obzirom na f_1 je broj $K(f_0, f_1) := \mathbf{E}_0 \log\left(\frac{f_0(X)}{f_1(X)}\right) = \int \log \frac{f_0(x)}{f_1(x)} f_0(x) d\mu(x)$.*

Dokaz sljedeće leme može se naći u [4].

Lema 3.0.7. *Neka su f_0 i f_1 gustoće obzirom na σ-konačnu mjeru μ. Tada je*

$$K(f_0, f_1) \geq 0,$$

pri čemu jednakost vrijedi ako i samo ako je $f_0 = f_1$ μ-g.s.

Iskažimo i dokažimo sada teorem koji daje uvjete uz koje imamo jaku konzistentnost niza procjenitelja maksimalne vjerodostojnosti.

Teorem 3.0.8. *Neka je za $n \in \mathbb{N}$, $(X_1, X_2, \ldots, X_n)$ slučajni uzorak iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ i $\theta_0 \in \Theta$ zadan. Definirajmo*

$$U(x, \theta) := \log \frac{f(x|\theta)}{f(x|\theta_0)}, \quad x \in supp f(\cdot|\theta).$$

Ako vrijedi:

(a) Θ je kompaktan,

(b) $f(x|\theta)$ je neprekidna po θ za svaki x,

(c) postoji izmjeriva funkcija $x \mapsto K(x)$ takva da je $\mathbf{E}_{\theta_0}|K(X)| < \infty$ i $|U(x, \theta)| \leq K(x)$ za sve x i θ.

Tada je svaki niz procjenitelja maksimalne vjerodostojnosti $(\hat{\theta}_n)_{n\in\mathbb{N}}$ jako konzistentan za θ_0, tj. $\hat{\theta}_n \xrightarrow{\mathbf{P}_{\theta_0}\text{-}g.s.} \theta_0$.

Dokaz. Neka je $\rho > 0$ i $S = \{\theta \in \Theta : |\theta - \theta_0| \geq \rho\}$. Tada je S kompaktan, pa iz teorema 2.0.1 slijedi

$$\limsup_{n\to\infty} \sup_{\theta\in S} \frac{1}{n} \sum_{i=1}^{n} U(X_i, \theta) \leq \sup_{\theta\in S} \mu(\theta), \quad \mathbf{P}_{\theta_0}\text{-}g.s.$$

gdje je $\mu(\theta) = \mathbf{E}_{\theta_0} U(X_1, \theta) = -K(f_{\theta_0}, f_\theta)$. Iz leme 3.0.7 slijedi da je $\mu(\theta) < 0$ za sve $\theta \in S$. Nadalje, primjenom Lebesgueovog teorema o dominiranoj konvergenciji slijedi da je μ neprekidna funkcija, pa postiže maksimum na kompaktnom skupu S. Neka je $\delta = \sup\limits_{\theta \in S} \mu(\theta)$. Tada je $\delta < 0$ i

$$\limsup_{n\to\infty} \sup_{\theta\in S} \frac{1}{n} \sum_{i=1}^{n} U(X_i, \theta) \leq \delta, \quad \mathbf{P}_{\theta_0}\text{-}g.s. \tag{3.10}$$

Iz (3.10) slijedi da postoji događaj Ω_ρ, $\mathbf{P}_{\theta_0}(\Omega_\rho) = 1$, takav da za svaki $\omega \in \Omega_\rho$ postoji $N(\omega) \in \mathbb{N}$ takav da za sve $n \in \mathbb{N}$, $n > N$ i $x_i = X_i(\omega)$ vrijedi

$$\sup_{\theta\in S} \frac{1}{n} \sum_{i=1}^{n} U(x_i, \theta) \leq \frac{\delta}{2} < 0.$$

Za niz procjenitelja maksimalne vjerodostojnosti $(\hat{\theta}_n)_{n\in\mathbb{N}}$ je

$$\frac{1}{n} \sum_{i=1}^{n} U(x_i, \hat{\theta}_n) = \sup_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} U(x_i, \theta) \geq 0,$$

gdje je desna strana veća ili jednaka od nula jer je za $\theta = \theta_0$ pripadna suma jednaka 0. Dakle, $\hat{\theta}_n \notin S$ za $n > N$, odnosno $|\hat{\theta}_n - \theta_0| < \rho$. Kako $\rho > 0$ možemo uzeti po volji mali, slijedi tvrdnja teorema. Naime, možemo staviti $\rho_k = \frac{1}{k}$, $k \in \mathbb{N}$, pa za svaki $k \in \mathbb{N}$ imamo događaj Ω_k takav da je $\mathbf{P}_{\theta_0}(\Omega_k) = 1$ i za $\omega \in \Omega_k$ postoji $N(\omega) \in \mathbb{N}$, takav da za sve $n > N$ vrijedi $|\hat{\theta}_n - \theta_0| < \frac{1}{k}$. Sada stavimo $\Omega' = \bigcap\limits_{k=1}^{\infty} \Omega_k$, pa je $\mathbf{P}_{\theta_0}(\Omega') = 1$ i za $\omega \in \Omega'$ je $\lim\limits_{n\to\infty} \hat{\theta}_n(X_1(\omega), \ldots, X_n(\omega)) = \theta_0$. □

Poglavlje 4

Asimptotska normalnost procjenitelja maksimalne vjerodostojnosti

U ovom poglavlju dokazujemo rezultat o asimptotskoj normalnosti niza procjenitelja maksimalne vjerodostojnosti. Preciznije dajemo uvjete uz koje postoji konzistentan i asimptotski normalan niz korijena jednadžbe vjerodostojnosti

$$\dot{l}_n(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i|\theta) = 0.$$

Taj rezultat nam neće uvijek osigurati asimptotsku normalnost niza procjenitelja maksimalne vjerodostojnosti, ali u nekim slučajevima hoće. Kao na primjer kada je procjenitelj maksimalne vjerodostojnosti jedini korijen jednadžbe vjerodostojnosti, što je česta situacija u primjeni. Kako želimo promatrati derivacije funkcije $\theta \mapsto f(x|\theta)$, trebamo dodatnu pretpostavku o diferencijabilnosti. U nastavku pretpostavljamo da $\frac{\partial^2}{\partial\theta^2} f(x|\theta)$ postoji i neprekidna je za svaki x, te pretpostavljamo da je $\Theta \subset \mathbb{R}^k$ otvoren. Za početak uvedimo neke dodatne oznake i pojmove.
Za funkciju $f : \mathbb{R}^k \to \mathbb{R}$ definiramo njenu derivaciju

$$\dot{f}(x) = \frac{\partial}{\partial x} f(x) = \left(\frac{\partial}{\partial x_1} f(x), \dots, \frac{\partial}{\partial x_k} f(x) \right).$$

Za vektorsku funkciju $g : \mathbb{R}^k \to \mathbb{R}^d$,

$$g = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_d \end{bmatrix}$$

definiramo

$$\dot{g}(x) = \frac{\partial}{\partial x} g(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} g_1(x) & \cdots & \frac{\partial}{\partial x_k} g_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} g_d(x) & \cdots & \frac{\partial}{\partial x_k} g_d(x) \end{bmatrix}.$$

Također, za funkciju $f : \mathbb{R}^k \to \mathbb{R}$ definiramo drugu derivaciju

$$\ddot{f}(x) = \frac{\partial^2}{\partial x^2} f(x) = \frac{\partial}{\partial x} \dot{f}(x)^T = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \cdots & \frac{\partial^2}{\partial x_k \partial x_1} f(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_k} f(x) & \cdots & \frac{\partial^2}{\partial x_k^2} f(x) \end{bmatrix}.$$

Integral vektorske funkcije $g : \mathbb{R}^k \to \mathbb{R}^d$ obzirom na mjeru μ definiramo preko komponentnih funkcija:

$$\int g \, d\mu = \int \begin{bmatrix} g_1 \\ \vdots \\ g_d \end{bmatrix} d\mu = \begin{bmatrix} \int g_1 \, d\mu \\ \vdots \\ \int g_d \, d\mu \end{bmatrix}.$$

Definirajmo funkciju Ψ sa

$$\Psi(x, \theta) = \left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^T \in \mathbb{R}^k.$$

Fisherova informacija je funkcija $\mathcal{F} : \Theta \to \mathbb{R}^{k \times k}$ definirana sa

$$\mathcal{F}(\theta) = \mathbf{E}_\theta \Psi(X, \theta) \Psi(X, \theta)^T.$$

Uz pretpostavku da je $\int \frac{\partial}{\partial \theta} f(x|\theta) d\mu(x) = 0$ imamo

$$\mathbf{E}_\theta \Psi(X, \theta) = \int \frac{(\frac{\partial}{\partial \theta} f(x|\theta))^T}{f(x|\theta)} f(x|\theta) \, d\mu(x) = \int \left(\frac{\partial}{\partial \theta} f(x|\theta) \right)^T d\mu(x) = 0.$$

Dakle, u tom slučaju je Fisherova informacija zapravo kovarijacijska matrica od $\Psi(X, \theta)$,

$$\mathcal{F}(\theta) = \mathrm{Var}_\theta(\Psi(X, \theta)).$$

Ako vrijedi $\int \frac{\partial^2}{\partial \theta^2} f(x|\theta) \, d\mu(x) = 0$, tada je

$$\begin{aligned} \mathbf{E}_\theta \dot{\Psi}(X, \theta) &= \int \left[\frac{\partial}{\partial \theta} \frac{\left(\frac{\partial}{\partial \theta} f(x|\theta) \right)^T}{f(x|\theta)} \right] f(x|\theta) \, d\mu(x) \\ &= \int \frac{f(x|\theta) \frac{\partial^2}{\partial \theta^2} f(x|\theta) - (\frac{\partial}{\partial \theta} f(x|\theta))^T \frac{\partial}{\partial \theta} f(x|\theta)}{f(x|\theta)^2} f(x|\theta) \, d\mu(x) \\ &= \mathbf{0} - \int \Psi(x, \theta) \Psi(x, \theta)^T f(x|\theta) \, d\mu(x) \\ &= -\mathcal{F}(\theta). \end{aligned}$$

Dakle, u tom slučaju za Fisherovu informaciju vrijedi

$$\mathcal{F}(\theta) = -\mathbf{E}_\theta \dot{\Psi}(X, \theta). \tag{4.1}$$

Ilustrirajmo na primjeru prethodno uvedene pojmove.

Primjer 4.0.1. *Promatramo normalni model $\mathcal{N}(\mu, \sigma^2)$, tj. gustoća je zadana sa*

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Tada je

$$\log f(x|\mu, \sigma^2) = -\log(\sqrt{2\pi\sigma^2}) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Izračunajmo sada Ψ, $\dot{\Psi}$ i Fisherovu informaciju $\mathcal{F}$.

$$\Psi(x, (\mu, \sigma^2)) = \begin{bmatrix} \frac{\partial}{\partial\mu} \log f(x|\mu, \sigma^2) \\ \frac{\partial}{\partial\sigma^2} \log f(x|\mu, \sigma^2) \end{bmatrix} = \begin{bmatrix} \frac{(x-\mu)}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{bmatrix}$$

$$\dot{\Psi}(x, (\mu, \sigma^2)) = \begin{bmatrix} -\frac{1}{\sigma^2} & -\frac{(x-\mu)}{\sigma^4} \\ -\frac{(x-\mu)}{\sigma^4} & \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6} \end{bmatrix}.$$

Kako u ovom primjeru vrijedi $\int \frac{\partial^2}{\partial\theta^2} f(x|\theta)\, dx = 0$, Fisherovu informaciju možemo izračunati iz (4.1)*. Koristeći da je $\mathbf{E}_\theta(X-\mu) = 0$ i $\mathbf{E}_\theta(X-\mu)^2 = \sigma^2$, slijedi*

$$\mathcal{F}(\mu, \sigma^2) = -\mathbf{E}_\theta \dot{\Psi}(X, (\mu, \sigma^2)) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}.$$

Pogledajmo još jedan primjer za diskretni model.

Primjer 4.0.2. *Promatramo Poissonov model $P(\theta)$, tj. gustoća je zadana sa*

$$f(x|\theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x \in \mathbb{N}_0.$$

Tada je

$$\log f(x|\theta) = x \log(\theta) - \theta - \log(x!), \quad x \in \mathbb{N}_0.$$

Odredimo Ψ, $\dot{\Psi}$ i Fisherovu informaciju $\mathcal{F}$.

$$\Psi(x, \theta) = \frac{\partial}{\partial\theta} \log f(x|\theta) = \frac{x}{\theta} - 1$$

$$\dot{\Psi}(x, \theta) = -\frac{x}{\theta^2}.$$

Pokažimo da i u ovom primjeru vrijedi $\int \frac{\partial^2}{\partial\theta^2} f(x|\theta)\, d\mu(x) = 0$*,*

$$\begin{aligned}\int \frac{\partial^2}{\partial\theta^2} f(x|\theta)\, d\mu(x) = \sum_{k=0}^{\infty} \frac{\partial^2}{\partial\theta^2} f(k|\theta) &= \sum_{k=0}^{\infty} \frac{\theta^2 + k^2 - (2\theta+1)k}{\theta^2} \cdot f(k|\theta) \\ &= \frac{1}{\theta^2}\mathbf{E}_\theta \left[\theta^2 + X^2 - (2\theta+1)X\right] \\ &= \frac{1}{\theta^2}\left(\theta^2 + \theta^2 + \theta - (2\theta+1)\theta\right) = 0.\end{aligned}$$

U prethodnom računu smo koristili da je $\mathbf{E}_\theta X = \theta$ *i* $Var_\theta(X) = \theta$*, također integral je zapravo suma, jer je u ovom slučaju* μ *brojeća mjera. Sada možemo izračunati* $\mathcal{F}$ *koristeći* (4.1)*:*

$$\mathcal{F}(\theta) = -\mathbf{E}_\theta \dot{\Psi}(X,\theta) = \frac{1}{\theta}.$$

Prije nego iskažemo i dokažemo glavni rezultat ovog poglavlja podsjetimo se dvaju rezultata iz diferencijalnog računa funkcija više varijabli.

Teorem 4.0.3. *Ako je* $f : \mathbb{R}^d \to \mathbb{R}^k$ *takva da* $\dot{f}$ *postoji i neprekidna je na kugli* $B(x_0, r) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$*, tada za* $t \in \mathbb{R}^d$*,* $|t| < r$ *vrijedi*

$$f(x_0 + t) = f(x_0) + \int_0^1 \dot{f}(x_0 + ut)\, du \cdot t.$$

Dokaz. Definirajmo funkciju $h : \mathbb{R} \to \mathbb{R}^k$, $h(u) = f(x_0 + ut)$. Tada je

$$\dot{h}(u) = \dot{f}(x_0 + ut) \cdot t,$$

odakle slijedi

$$\begin{aligned}f(x_0 + t) - f(x_0) = h(1) - h(0) = \int_0^1 \dot{h}(u)\, du &= \int_0^1 \dot{f}(x_0 + ut) \cdot t\, du \\ &= \int_0^1 \dot{f}(x_0 + ut)\, du \cdot t. \qquad (4.2)\end{aligned}$$

□

Teorem 4.0.4. *Ako je* $f : \mathbb{R}^d \to \mathbb{R}$ *takva da* $\ddot{f}$ *postoji i neprekidna je na kugli* $B(x_0, r) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$*, tada za* $t \in \mathbb{R}^d$*,* $|t| < r$ *vrijedi*

$$f(x_0 + t) = f(x_0) + \dot{f}(x_0) \cdot t + t^T \cdot \int_0^1 \int_0^1 v \ddot{f}(x_0 + uvt)\, du\, dv \cdot t.$$

Dokaz. Definirajmo funkciju $h : \mathbb{R} \to \mathbb{R}$, $h(u) = f(x_0 + ut)$. Tada je

$$\dot{h}(u) = \dot{f}(x_0 + ut) \cdot t \ \text{ i } \ \ddot{h}(u) = t^T \cdot \ddot{f}(x_0 + ut) \cdot t. \tag{4.3}$$

Koristeći metodu parcijalne integracije slijedi

$$\begin{aligned} h(1) - h(0) = \int_0^1 \dot{h}(v)\, dv &= (v-1) \cdot \dot{h}(v) \Big|_0^1 - \int_0^1 (v-1) \cdot \ddot{h}(v)\, dv \\ &= \dot{h}(0) - \int_0^1 (v-1) \cdot \ddot{h}(v)\, dv. \end{aligned} \tag{4.4}$$

Još jednom primjenom parcijalne integracije slijedi

$$\begin{aligned} \int_0^1 (v-1) \cdot \ddot{h}(v)\, dv &= (v-1) \int_0^v \ddot{h}(s)\, ds \Big|_0^1 - \int_0^1 \int_0^v \ddot{h}(s)\, ds\, dv \\ &= -\int_0^1 \int_0^1 v \cdot \ddot{h}(uv)\, du\, dv. \end{aligned} \tag{4.5}$$

Koristeći (4.3), (4.4) i (4.5) slijedi tražena tvrdnja. □

Sljedeći teorem je najavljeni rezultat koji daje uvjete uz koje postoji jako konzistentan niz korijena jednadžbe vjerodostojnosti te koji je uz to i asimptotski normalan.

Teorem 4.0.5. *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s gustoćom $f(x|\theta)$ (obzirom na mjeru μ) te neka je $\theta_0 \in \Theta$ prava vrijednost parametra. Ako su zadovoljeni sljedeći uvjeti:*

(a) $\Theta \subset \mathbb{R}^k$ *je otvoren,*

(b) druge parcijalne derivacije od $f(x|\theta)$ obzirom na θ postoje i neprekidne su za svaki x te vrijedi

$$\int \frac{\partial^i}{\partial \theta^i} f(x|\theta)\, d\mu(x) = 0 \ \ za \ \ i = 1, 2,$$

(c) postoji funkcija $K(x)$ takva da je $\mathbf{E}_{\theta_0} K(X) < \infty$ i svaka komponenta funkcije $|\dot{\Psi}(x, \theta)|$ je uniformno omeđena sa $K(x)$ na nekoj okolini od θ_0,

(d) $\mathcal{F}(\theta_0) = -\mathbf{E}_{\theta_0} \dot{\Psi}(X, \theta_0)$ je pozitivno definitna matrica.

Tada postoji jako konzistentan niz $(\hat{\theta}_n)_{n \in \mathbb{N}}$ korijena jednadžbe vjerodostojnosti za koji vrijedi:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathcal{F}(\theta_0)^{-1}).$$

Dokaz. Dokaz provodimo u dva koraka. Prvo dokazujemo egzistenciju konzistentnih korijena jednadžbe vjerodostojnosti, a potom asimptotsku normalnost.

1. *Postojanje konzistentnih korijena.*
   Neka je $\delta > 0$ takav da je $S_\delta = \{\theta : |\theta - \theta_0| \leq \delta\}$ sadržan u okolini od θ_0 na kojoj su sve komponente od $|\dot{\Psi}(x, \theta)|$ uniformno omeđene sa $K(x)$. Skup S_δ je kompaktan i funkcija $\theta \mapsto f(x|\theta)$ je neprekidna. Kako neprekidna funkcija na kompaktnom skupu postiže maksimum, za svaki $n \in \mathbb{N}$ postoji $\hat{\theta}_n$ takav da vrijedi

   $$L_n(\hat{\theta}_n) = \max_{\theta \in S_\delta} L_n(\theta). \tag{4.6}$$

   Ako pokažemo da su zadovoljene sve pretpostavke teorema 3.0.8, tada primjenom tog teorema na skup S_δ slijedi jaka konzistentnost niza $(\hat{\theta}_n)_{n\in\mathbb{N}}$. Zbog konzistentnosti niza $(\hat{\theta}_n)_{n\in\mathbb{N}}$, za dovoljno velike $n \in \mathbb{N}$ tada vrijedi da je $\hat{\theta}_n \in \mathrm{Int}(S_\delta)$. Dakle, zbog (4.6) slijedi da je $\hat{\theta}_n$ korijen jednadžbe vjerodostojnosti, tj. vrijedi $\dot{l}_n(\hat{\theta}_n) = 0$. Dakle, preostaje pokazati da su zadovoljene sve pretpostavke teorema 3.0.8. Uvjeti (a) i (b) su očito zadovoljeni pa pokažimo još da vrijedi i uvjet (c). Primjenom teorema 4.0.4 na funkciju $U(x, \theta) = \log \frac{f(x|\theta)}{f(x|\theta_0)}$ imamo

   $$\begin{aligned} U(x,\theta) =& U(x,\theta_0) + \Psi(x,\theta_0)^T(\theta-\theta_0) \\ &+ (\theta-\theta_0)^T \left(\int_0^1 \int_0^1 \lambda \dot{\Psi}(x, \theta_0 + \lambda u(\theta - \theta_0))\, d\lambda du \right) (\theta - \theta_0). \end{aligned}$$

   Funkcija $x \mapsto \Psi(x, \theta_0)$ je integrabilna, $U(x, \theta_0) = 0$ i zadnji član u gornjem izrazu je ograničen integrabilnom funkcijom jer su sve komponente od $\dot{\Psi}$ ograničene integrabilnom funkcijom K. Dakle, postoji integrabilna funkcija $x \mapsto \tilde{K}(x)$ takva da je $|U(x, \theta)| \leq \tilde{K}(x)$ za sve x i sve $\theta \in S_\delta$.

2. Asimptotska normalnost
   Za log-vjerodostojnost vrijedi $\dot{l}_n(\theta)^T = \sum_{i=1}^n \Psi(X_i, \theta)$. Koristeći teorem 4.0.3 dobijemo

   $$\dot{l}_n(\theta)^T = \dot{l}_n(\theta_0)^T + \int_0^1 \sum_{i=1}^n \dot{\Psi}(X_i, \theta_0 + \lambda(\theta - \theta_0))\, d\lambda\, (\theta - \theta_0). \tag{4.7}$$

   Neka je $(\hat{\theta}_n)_{n\in\mathbb{N}}$ jako konzistentan niz korijena jednadžbe vjerodostojnosti, dakle za dovoljno velike $n \in \mathbb{N}$ je $\dot{l}_n(\hat{\theta}_n) = 0$ (u prvom dijelu dokaza je pokazano da takav niz postoji). Uvrštavanjem $\theta = \hat{\theta}_n$ u (4.7) i dijeljenjem s $\sqrt{n}$ slijedi

   $$\frac{1}{\sqrt{n}} \dot{l}_n(\theta_0)^T = B_n \sqrt{n}(\hat{\theta}_n - \theta_0),$$

gdje je

$$B_n = -\int_0^1 \frac{1}{n}\sum_{i=1}^n \dot{\Psi}(X_i, \theta_0 + \lambda(\hat{\theta}_n - \theta_0))\, d\lambda.$$

Kako je $\mathbf{E}_{\theta_0}\Psi(X,\theta_0) = 0$ i $Var_{\theta_0}\Psi(X,\theta_0) = \mathcal{F}(\theta_0)$, primjenom centralnog graničnog teorema slijedi

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)^T = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \Psi(X_i,\theta_0)\right) \xrightarrow{D} Z \sim \mathcal{N}(\mathbf{0}, \mathcal{F}(\theta_0)).$$

Ako pokažemo da $B_n \xrightarrow{g.s.} \mathcal{F}(\theta_0)$, tada za dovoljno velike $n \in \mathbb{N}$ postoji B_n^{-1}, jer je $\mathcal{F}(\theta_0)$ regularna i preslikavanje $A \mapsto \det(A)$ je neprekidno. Također, kako je preslikavanje $A \mapsto A^{-1}$ neprekidno slijedi $B_n^{-1} \xrightarrow{g.s.} \mathcal{F}(\theta_0)^{-1}$, pa primjenom teorema 1.2.9 (Slutsky) slijedi tražena tvrdnja:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = B_n^{-1}\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)^T \xrightarrow{D} \mathcal{F}(\theta_0)^{-1}Z \sim \mathcal{N}(\mathbf{0}, \mathcal{F}(\theta_0)^{-1}).$$

Neka je $\varepsilon > 0$. Primjenom Lebesgueovog teorema o dominiranoj konvergenciji lako se pokaže da je funkcija $\theta \mapsto \mathbf{E}_{\theta_0}\dot{\Psi}(X,\theta)$ neprekidna u θ_0, pa postoji $\delta > 0$ takav da vrijedi:

$$|\theta - \theta_0| < \delta \implies |\mathbf{E}_{\theta_0}\dot{\Psi}(X,\theta) + \mathcal{F}(\theta_0)| < \varepsilon.$$

Pritom je δ izabran dovoljno mali, tako da je skup $S_\delta = \{\theta : |\theta - \theta_0| \leq \delta\}$ sadržan u okolini od θ_0 na kojoj su sve komponente od $|\dot{\Psi}(x,\theta)|$ uniformno omeđene sa $K(x)$. Koristeći uniformi jaki zakon velikih brojeva (teorem 2.0.1) zaključujemo da postoji $n_0 \in \mathbb{N}$ takav da vrijedi:

$$n \geq n_0 \implies \sup_{\theta\in S_\delta}\left|\frac{1}{n}\sum_{i=1}^n \dot{\Psi}(X_i,\theta) - \mathbf{E}_{\theta_0}\dot{\Psi}(X,\theta)\right| \leq \varepsilon.$$

Preciznije postoji događaj vjerojatnosti jedan na kojemu vrijedi prethodno i n_0 ovisi o realizaciji niza $X_1, X_2, \ldots$, ali u nastavku radi jednostavnosti to nećemo posebno naglašavati. Zbog jake konzistentnosti niza $(\hat{\theta}_n)_{n\in N}$ postoji $n_1 \in \mathbb{N}$ takav da za $n \in \mathbb{N}$, $n \geq n_1$ vrijedi $|\hat{\theta}_n - \theta_0| < \delta$. Sada za svaki $n \geq \max\{n_0, n_1\}$ imamo:

$$\begin{aligned}
|B_n - \mathcal{F}(\theta_0)| &\leq \int_0^1 \left|\frac{1}{n}\sum_{i=1}^n \dot{\Psi}\Big(X_i, \theta_0 + \lambda(\hat{\theta}_n - \theta_0)\Big) + \mathcal{F}(\theta_0)\right| d\lambda \\
&\leq \int_0^1 \sup_{\theta\in S_\delta}\left[\left|\frac{1}{n}\sum_{i=1}^n \dot{\Psi}(X_i,\theta) - \mathbf{E}_{\theta_0}\dot{\Psi}(X,\theta)\right| + \left|\mathbf{E}_{\theta_0}\dot{\Psi}(X,\theta) + \mathcal{F}(\theta_0)\right|\right] d\lambda \\
&\leq 2\varepsilon.
\end{aligned}$$

Dakle, vrijedi $B_n \xrightarrow{g.s.} \mathcal{F}(\theta_0)$. $\square$

Napomena 4.0.6. *Često se kaže da prethodni teorem pokazuje da je procjenitelj maksimalne vjerodostojnosti konzistentan i asimptotski normalan, međutim to nije sasvim točno. Sve što teorem tvrdi je da uz zadane uvjete postoji konzistentan niz korijena jednadžbe vjerodostojnosti koji je asimptotski normalan. Naime, uz zadane uvjete procjenitelj maksimalne vjerodostojnosti uopće ne mora postojati. Nadalje, čak i ako postoji i zadovoljava jednadžbu vjerodostojnosti ne mora nužno biti konzistentan. Također u slučaju da korijen jednadžbe vjerodostojnosti nije jedinstven, teorem nam ne odgovara na pitanje koji korijen izabrati.*

Primjer 4.0.7. *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s gustoćom*

$$f(x|\theta) = \frac{1}{\pi(1+(x-\theta)^2)}, \quad \theta \in \mathbb{R}.$$

Dakle, radi se o Cauchyjevoj distribuciji s parametrom θ. Neka je $\theta_0 \in \mathbb{R}$ prava vrijednost parametra. Može se pokazati da su zadovoljene sve pretpostavke teorema 4.0.5 i da je $\mathcal{F}(\theta_0) = \frac{1}{2}$. Dakle, po teoremu 4.0.5 slijedi da postoji konzistentan niz $(\hat{\theta}_n)_{n\in\mathbb{N}}$ korijena jednadžbe vjerodostojnosti takav da

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, 2).$$

Odredimo jednadžbu vjerodostojnosti. Funkcija log-vjerodostojnosti je

$$l_n(\theta) = -n\log(\pi) - \sum_{i=1}^{n} \log(1 + (X_i - \theta)^2).$$

Deriviranjem i izjednačavanjem derivacije s 0 *dolazimo do jednadžbe vjerodostojnosti*

$$\dot{l}_n(\theta) = 2\sum_{i=1}^{n} \frac{(X_i - \theta)}{1 + (X_i - \theta)^2} = 0. \tag{4.8}$$

Radi se o nelinearnoj jednadžbi koja se može svesti na traženje nultočki polinoma velikog stupnja (općenito stupnja $2n-1$, gdje je n veličina uzorka). Dakle, općenito će imati više korijena te ne znamo odrediti rješenja u eksplicitnom obliku, pa vidimo da uopće nije jasno koji korijen izabrati.

Pokažimo za kraj ovog poglavlja nekoliko primjera gdje teorem 4.0.5 uistinu garantira konzistentnost i asimptotsku normalnost niza procjenitelja maksimalne vjerodostojnosti.

Primjer 4.0.8. *Neka je $(x_1, x_2, \ldots, x_n)$ realizacija slučajnog uzorka $(X_1, X_2, \ldots, X_n)$ iz modela $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$, gdje je*

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad \theta = (\alpha,\beta), \quad \Theta = \langle 0, +\infty\rangle \times \langle 0, +\infty\rangle.$$

Funkcija log-vjerodostojnosti je

$$l_n(\alpha,\beta) = -n\log(\Gamma(\alpha)) + n\alpha\log(\beta) + (\alpha-1)\sum_{i=1}^n \log(x_i) - \beta\sum_{i=1}^n x_i.$$

Jednadžba vjerodostojnosti $\dot{l}_n(\alpha,\beta) = 0$ se svodi na sustav jednadžbi

$$\begin{aligned} \frac{\partial}{\partial\alpha} l_n(\alpha,\beta) &= -n\psi(\alpha) + n\log(\beta) + \sum_{i=1}^n \log(x_i) = 0 \\ \frac{\partial}{\partial\beta} l_n(\alpha,\beta) &= \frac{n\alpha}{\beta} - \sum_{i=1}^n x_i = 0, \end{aligned}$$

gdje je $\psi(\alpha) = \frac{\partial}{\partial\alpha}\log(\Gamma(\alpha)) = \frac{\dot{\Gamma}(\alpha)}{\Gamma(\alpha)}$. Iz druge jednadžbe slijedi

$$\beta = \frac{n\alpha}{\sum_{i=1}^n x_i} \tag{4.9}$$

pa uvrštavanjem u prvu jednadžbu, dobijemo jednadžbu

$$\psi(\alpha) - \log(\alpha) = \frac{1}{n}\sum_{i=1}^n \log(x_i) - \log\left(\frac{1}{n}\sum_{i=1}^n x_i\right). \tag{4.10}$$

Funkcija $x \mapsto \log(x)$ je strogo konkavna, pa je

$$\frac{1}{n}\sum_{i=1}^n \log(x_i) - \log\left(\frac{1}{n}\sum_{i=1}^n x_i\right) < 0.$$

Nadalje, funkcija $H(\alpha) := \psi(\alpha) - \log(\alpha)$, $H : \langle 0, +\infty\rangle \to \langle -\infty, 0\rangle$ je strogo rastuća i neprekidna, pa zaključujemo da postoji jedinstveno rješenje jednadžbe (4.10)*:*

$$\hat{\alpha} = H^{-1}\left(\frac{1}{n}\sum_{i=1}^n \log(x_i) - \log\left(\frac{1}{n}\sum_{i=1}^n x_i\right)\right).$$

Uvrštavanjem $\hat{\alpha}$ *u* (4.9) *dolazimo do jedinstvenog korijena jednadžbe vjerodostojnosti* $(\hat{\alpha}, \hat{\beta})$*. Da bi pokazali da je to točka globalnog maksimuma funkcije* l_n *na* Θ *dovoljno je pokazati da je matrica druge derivacije (hessian) od* l_n*,*

$$\ddot{l}_n(\alpha,\beta) = \begin{bmatrix} -n\psi'(\alpha) & \frac{n}{\beta} \\ \frac{n}{\beta} & -\frac{n\alpha}{\beta^2} \end{bmatrix},$$

negativno definitna na Θ*. Ako pokažemo da je* $-\ddot{l}_n(\alpha,\beta)$ *pozitivno definitna za sve* $(\alpha,\beta) \in \Theta$ *tada je* $\ddot{l}_n(\alpha,\beta)$ *negativno definitna za sve* $(\alpha,\beta) \in \Theta$*.*

$$-\ddot{l}_n(\alpha,\beta) = \begin{bmatrix} n\psi'(\alpha) & -\frac{n}{\beta} \\ -\frac{n}{\beta} & \frac{n\alpha}{\beta^2} \end{bmatrix}.$$

Kako je $n\psi'(\alpha) > 0$ *i* $\det(-\ddot{l}_n(\alpha,\beta)) = \frac{n^2}{\beta^2}(\alpha\psi'(\alpha) - 1) > 0$ *za svaki* $(\alpha,\beta) \in \Theta$*, koristeći Sylvesterov kriterij, slijedi da je* $-\ddot{l}_n(\alpha,\beta)$ *pozitivno definitna za svaki* $(\alpha,\beta) \in \Theta$*. Dakle, za svaki* $n \in \mathbb{N}$ *procjenitelj maksimalne vjerodostojnosti postoji i jedinstveni je korijen jednadžbe vjerodostojnosti. Ako pokažemo da su zadovoljene sve pretpostavke teorema 4.0.5, tada slijedi da je niz procjenitelja maksimalne vjerodostojnosti jako konzistentan i asimptotski normalan. Pretpostavka* (*a*) *je očito zadovoljena i direktnim računom se pokaže da vrijedi* (*b*)*. Neka je* $\theta_0 = (\alpha_0, \beta_0)$*. Izračunajmo* $\dot{\Psi}(\alpha,\beta)$ *i* $\mathcal{F}(\alpha_0,\beta_0)$*:*

$$\log f(x|\alpha,\beta) = -\log\Gamma(\alpha) + \alpha\log(\beta) + (\alpha-1)\log(x) - \beta x,$$

$$\Psi(x,(\alpha,\beta)) = \frac{\partial}{\partial\theta}\log f(x|\theta) = \begin{bmatrix} -\psi(\alpha) + \log(\beta) + \log(x) \\ \frac{\alpha}{\beta} - x \end{bmatrix},$$

$$\dot{\Psi}(x,(\alpha,\beta)) = \begin{bmatrix} -\psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{bmatrix},$$

$$\mathcal{F}(\alpha_0,\beta_0) = -\mathbf{E}_{\theta_0}\dot{\Psi}(X,(\alpha_0,\beta_0)) = \begin{bmatrix} \psi'(\alpha_0) & -\frac{1}{\beta_0} \\ -\frac{1}{\beta_0} & \frac{\alpha_0}{\beta_0^2} \end{bmatrix}.$$

Vidimo da $\dot{\Psi}(x,(\alpha,\beta))$ *ne ovisi o* x *i sve komponente su neprekidne po* θ*, pa slijedi da je zadovoljen uvjet* (*c*)*. Koristeći Sylvesterov kriterij slijedi da je matrica* $\mathcal{F}(\alpha_0,\beta_0)$ *pozitivno definitna, pa je zadovoljena i pretpostavka* (*d*)*. Dakle, zadovoljene su sve pretpostavke teorema 4.0.5 i pokazali smo da je procjenitelj maksimalne vjerodostojnosti jedini korijen jednadžbe vjerodostojnosti, pa zaključujemo da je niz procjenitelja maksimalne vjerodostojnosti jako konzistentan i asimptotski normalan.*

Jedna široka klasa modela u kojima se procjenitelj maksimalne vjerodostojnosti lijepo ponaša su eksponencijalne familije. Prethodni primjer također spada pod tu klasu, naime radilo se o gustoćama $\Gamma(\alpha, \beta)$ distribucije i taj model je 2-parametarska eksponencijalna familija. U sljedećem primjeru pokazujemo da u jednoparametarskoj eksponencijalnoj familiji procjenitelji maksimalne vjerodostojnosti postoje, konzistentni su i asimptotski normalni. Kako velik broj često korištenih modela spada pod eksponencijalne familije, ovaj rezultat donekle opravdava popularnost procjenitelja maksimalne vjerodostojnosti u primjenama.

Primjer 4.0.9 (jednoparametarska eksponencijalna familija)**.**
Promatramo model $\mathcal{P} = \{f(\cdot|\eta) : \eta \in \Theta = Int\,\Sigma\}$*, gdje su gustoće oblika*

$$f(x|\eta) = h(x) \cdot e^{\eta t(x) - A(\eta)}\ , \quad x \in \mathbb{R},$$

za neku nenegativnu izmjerivu funkciju $h : \mathbb{R} \to [0, +\infty\rangle$*, koja ima neprazan nosač i izmjerivu funkciju* $t : \mathbb{R} \to \mathbb{R}$*, takvu da je* $\mathbf{P}_\eta(t(X) = c) \neq 1$*, za sve* $c \in \mathbb{R}$ *i sve* $\eta \in \Theta$*. Skup* Σ *je zadan kao*

$$\Sigma = \left\{\eta \in \mathbb{R} : \int_{\mathbb{R}} e^{\eta t(x)} \cdot h(x)\, dx < +\infty\right\}.$$

Skup Σ *je konveksan, pa je* $\Theta = Int\,\Sigma$ *otvoren interval. Nadalje,*

$$A(\eta) = \log \int_{\mathbb{R}} e^{\eta t(x)} \cdot h(x)\, dx\ , \quad \eta \in \Theta.$$

Pokažimo da ovaj model zadovoljava sve uvjete teorema 4.0.5. Skup Θ *je otvoren interval, dakle vrijedi* (a)*. Može se pokazati (pogledati [9]) da je funkcija* $\eta \mapsto A(\eta)$ *klase* $C^\infty(\Theta)$ *i da vrijedi:*

$$A'(\eta) = \mathbf{E}_\eta(t(X)), \tag{4.11}$$
$$A''(\eta) = Var_\eta(t(X)). \tag{4.12}$$

Izračunajmo $\frac{\partial}{\partial \eta} f(x|\eta)$ *i* $\frac{\partial^2}{\partial \eta^2} f(x|\eta)$*:*

$$\frac{\partial}{\partial \eta} f(x|\eta) = (t(x) - A'(\eta)) \cdot e^{\eta t(x) - A(\eta)} \cdot h(x) = (t(x) - A'(\eta)) \cdot f(x|\eta) \tag{4.13}$$

$$\begin{aligned} \frac{\partial^2}{\partial \eta^2} f(x|\eta) &= -A''(\eta) \cdot f(x|\eta) + (t(x) - A'(\eta)) \cdot \frac{\partial}{\partial \eta} f(x|\eta) \\ &= -A''(\eta) \cdot f(x|\eta) + (t(x) - A'(\eta))^2 \cdot f(x|\eta). \end{aligned} \tag{4.14}$$

Koristeći (4.11), (4.12), (4.13) *i* (4.14) *slijedi:*

$$\begin{aligned}
\int \frac{\partial}{\partial\eta} f(x|\eta)\,dx &= \int (t(x) - A'(\eta)) \cdot f(x|\eta)\,dx = \mathbf{E}_\eta[t(X) - A'(\eta)] \\
&= \mathbf{E}_\eta[t(X)] - \mathbf{E}_\eta[A'(\eta)] = 0 \\
\int \frac{\partial^2}{\partial\eta^2} f(x|\eta)\,dx &= -\int A''(\eta) \cdot f(x|\eta)\,dx + \int (t(x) - A'(\eta))^2 \cdot f(x|\eta)\,dx \\
&= -A''(\eta) + \mathbf{E}_\eta[(t(X) - A'(\eta))^2] \\
&= -A''(\eta) + Var_\eta(t(X)) = 0.
\end{aligned}$$

Dakle, vrijedi uvjet (*b*)*. Izračunajmo* $\dot\Psi(x,\eta)$*:*

$$\begin{aligned}
\log f(x|\eta) &= \eta \cdot t(x) - A(\eta) + \log(h(x)) \\
\Psi(x,\eta) &= \frac{\partial}{\partial\eta} \log f(x|\eta) = t(x) - A'(\eta) \\
\dot\Psi(x,\eta) &= \frac{\partial^2}{\partial\eta^2} \log f(x|\eta) = -A''(\eta).
\end{aligned}$$

Neka je $\eta_0 \in \Theta$*. Skup* Θ *je otvoren interval, pa postoji* $\varepsilon > 0$ *takav da za* $I_{\eta_0} := [\eta_0 - \varepsilon, \eta_0 + \varepsilon]$ *vrijedi* $I_{\eta_0} \subset \Theta$*. Funkcija* $\eta \mapsto |A''(\eta)|$ *je neprekidna, pa poprima maksimum* $M_{\eta_0} := \max\limits_{\eta \in I_{\eta_0}} |A''(\eta)|$ *na kompaktnom skupu* I_{η_0}*. Sada za sve* $x \in \mathbb{R}$ *i* $\eta \in \langle \eta_0 - \varepsilon, \eta_0 + \varepsilon \rangle$ *vrijedi*

$$|\dot\Psi(x,\eta)| = |A''(\eta)| \le M_{\eta_0}.$$

Dakle, vrijedi uvjet (*c*)*. Preostaje još provjeriti uvjet* (*d*)*. Za Fisherovu informaciju vrijedi*

$$\mathcal{F}(\eta_0) = -\mathbf{E}_{\eta_0} \dot\Psi(X,\eta_0) = A''(\eta_0) = Var_{\eta_0}(t(X)) > 0,$$

pa je zadovoljen i uvjet (*d*)*. Da bi pokazali da je niz procjenitelja maksimalne vjerodostojnosti jako konzistentan i asimptotski normalan, dovoljno je pokazati da na događaju* $\mathbf{P}_{\eta_0}$*-vjerojatnosti jedan, za dovoljno velike* $n \in \mathbb{N}$*, postoji procjenitelj maksimalne vjerodostojnosti* $\hat\eta_n = \hat\eta_n(X_1, X_2, \ldots, X_n)$ *koji je ujedno i jedinstveni korijen jednadžbe vjerodostojnosti*

$$\dot l_n(\eta) = \frac{\partial}{\partial\eta} \sum_{i=1}^n \log f(X_i|\eta) = \sum_{i=1}^n t(X_i) - nA'(\eta) = 0. \tag{4.15}$$

Iz (4.15) *slijedi*

$$\dot l_n(\eta) = 0 \iff A'(\eta) = \frac{T_n}{n}, \tag{4.16}$$

gdje je $T_n = \sum_{i=1}^n t(X_i)$*. Kako je* $\mathbf{E}_{\eta_0} t(X_i) = A'(\eta_0)$*, primjenom jakog zakona velikih brojeva slijedi*

$$\frac{T_n}{n} \xrightarrow{\mathbf{P}_{\eta_0}\text{-}g.s.} A'(\eta_0). \tag{4.17}$$

Funkcija $\eta \mapsto A'(\eta)$ *je neprekidna i strogo rastuća na otvorenom intervalu* Θ*, jer za njenu derivaciju vrijedi*

$$A''(\eta) = Var_\eta(t(X_1)) > 0\ , \text{ za svaki } \eta \in \Theta.$$

Slika strogo rastuće neprekidne funkcije definirane na otvorenom intervalu je otvoren interval, pa je skup $I = A'(\Theta)$ *otvoren interval koji zadrži* $A'(\eta_0)$*. Kako je* $A'(\eta_0) \in I$ *i* $(\mathbf{P}_{\eta_0}\text{-}g.s.) \lim_{n\to\infty} \frac{T_n}{n} = A'(\eta_0)$*, na skupu* $\mathbf{P}_{\eta_0}$*-vjerojatnosti jedan postoji* $n_0 \in \mathbb{N}$ *takav da je za sve* $n \geq n_0$*,* $\frac{T_n}{n} \in I$*. Za svaki takav* $n \in \mathbb{N}$ *jednadžba* $A'(\eta) = \frac{T_n}{n}$ *ima jedinstveno rješenje* $\hat{\eta}_n \in \Theta$ *i to je po* (4.16) *jedinstveni korijen jednadžbe vjerodostojnosti. Budući da je za svaki* $\eta \in \Theta$*,* $\ddot{l}_n(\eta) = -nA''(\eta) < 0$*, zaključujemo da je* $\hat{\eta}_n$ *točka globalnog maksimuma od* l_n*, dakle,* $\hat{\eta}_n$ *je procjenitelj maksimalne vjerodostojnosti.*

Napomena 4.0.10. *Uočimo da možemo promatrati različite parametrizacije istih gustoća. Međutim, to nije problem jer se može pokazati da je procjenjivanje metodom maksimalne vjerodostojnosti invarijantno na funkcijske transformacije. Odnosno, ako je* $\hat{\theta}$ *procjenitelj maksimalne vjerodostojnosti za* θ *i* $\tau(\theta)$ *funkcija od* θ*, tada je* $\tau(\hat{\theta})$ *procjenitelj maksimalne vjerodostojnosti za* $\tau(\theta)$*.*

Poglavlje 5

Asimptotska efikasnost

U ovom poglavlju uvodimo pojam asimptotske efikasnosti i dajemo metodu kojom iz konzistentnog i asimptotski normalnog niza procjenitelja možemo dobiti asimptotski efikasan niz procjenitelja.

Definicija 5.0.1. *Neka je $(X_n : n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$. Niz procjenitelja $(\hat{\theta}_n)_{n\in\mathbb{N}}$ za θ, takav da*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma(\theta))$$

je asimptotski efikasan ako je $\Sigma(\theta) = \mathcal{F}(\theta)^{-1}$, za svaki $\theta \in \Theta$.

Direktno iz prethodne definicije slijedi da je uz uvjete teorema 4.0.5, niz konzistentnih korijena jednadžbe vjerodostojnosti iz tog teorema ujedno i asimptotski efikasan. Dakle, niz procjenitelja maksimalne vjerodostojnosti će često biti asimptotski efikasan, što je još jedno lijepo svojstvo tih procjenitelja i dodatni razlog njihove popularnosti. Međutim, kao što smo već vidjeli u primjeru 4.0.7 iz prethodnog poglavlja, nekada nije sasvim jednostavno odrediti korijene jednadžbe vjerodostojnosti. To nas motivira da u takvim situacijama pokušamo doći do nekog jednostavnijeg niza procjenitelja koji će biti asimptotski ekvivalentan nizu procjenitelja maksimalne vjerodostojnosti. U nastavku dajemo motivaciju i ideju za jednu takvu metodu "poboljšanja" asimptotski normalnog niza procjenitelja. Pretpostavimo da imamo niz procjenitelja $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ za θ, takav da

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \Sigma(\theta)).$$

Htjeli bi iz tog niza doći do niza procjenitelja $(\tilde{\theta}_n^{(1)})_{n\in\mathbb{N}}$ koji je asimptotski efikasan, a razlog tome je što će asimptotski efikasni nizovi procjenitelja često imati manju asimptotsku varijancu. Također, želimo da bude relativno lako izračunati te procjenitelje. Naime, u slučajevima kao što su primjeri 4.0.7 i 4.0.8 računanje nultočki jednadžbe vjerodostojnosti je iznimno komplicirano. Tada bi u takvim situacijama možda mogli doći do jednostavnijih asimptotski efikasnih procjenitelja poboljšavanjem procjenitelja dobivenih metodom momenata, koji su često jednostavni za odrediti i asimptotski normalni. Ilustrirajmo opisanu situaciju na jednom primjeru.

Primjer 5.0.2. *Neka je $X_1, X_2, \ldots X_n$ slučajni uzorak iz modela $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$, gdje su gustoće zadane kao*

$$f(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}.$$

Parametar θ u ovom slučaju predstavlja očekivanje i median. Koristeći centralni granični teorem slijedi da za niz uzoračkih sredina $(\overline{X}_n)_{n\in\mathbb{N}}$ vrijedi

$$\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{\pi^2}{3}).$$

Također, može se pokazati (pogledati [4]) da za niz uzoračkih medijana $(M_n)_{n\in\mathbb{N}}$ vrijedi

$$\sqrt{n}(M_n - \theta) \xrightarrow{D} \mathcal{N}(0, 4).$$

Fisherova informacija u ovom primjeru je $\mathcal{F}(\theta) = \frac{1}{3}$, za svaki $\theta \in \mathbb{R}$, pa je asimptotska varijanca asimptotski efikasnog niza procjenitelja jednaka 3, što je manje od 4 i $\frac{\pi^2}{3} \approx 3.2899$. Dakle, isplatilo bi se poboljšati nizove procjenitelja $(\overline{X}_n)_{n\in\mathbb{N}}$ i $(M_n)_{n\in\mathbb{N}}$ na način da dođemo do asimptotski efikasnih nizova.

Ideja je iskoristiti jako konzistentan i asimptotski efikasan niz korijena jednadžbe vjerodostojnosti, čiju egzistenciju garantira teorem 4.0.5. Naime, ako imamo neki jako konzistentan i asimptotski normalan niz procjenitelja $(\tilde{\theta}_n)_{n\in\mathbb{N}}$, možemo napraviti jednu iteraciju Newtonove metode za traženje nultočke primjenjenu na funkciju $\dot{l}_n$ s početnom procjenom $\tilde{\theta}_n$. Dakle, definiramo niz

$$\tilde{\theta}_n^{(1)} = \tilde{\theta}_n - \ddot{l}_n(\tilde{\theta}_n)^{-1}\dot{l}_n(\tilde{\theta}_n)^T.$$

Tada očekujemo da ćemo se na taj način dovoljno približiti asimptotski efikasnom nizu nultočki jednadžbe vjerodostojnosti i da će ta dva niza postati asimptotski ekvivalentna, pa će onda i niz $(\tilde{\theta}_n^{(1)})_{n\in\mathbb{N}}$ biti asimptotski efikasan. Također, kako uz određene pretpostavke niz $(\frac{1}{n}\ddot{l}_n(\tilde{\theta}_n))_{n\in\mathbb{N}}$ konvergira prema $-\mathcal{F}(\theta)$, očekujemo da će biti dobar i niz definiran kao

$$\theta_n^* = \tilde{\theta}_n + \frac{1}{n}\mathcal{F}(\tilde{\theta}_n)^{-1}\dot{l}_n(\tilde{\theta}_n)^T.$$

Sljedeći teorem opravdava opisanu metodu.

Teorem 5.0.3. *Neka je $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ jako konzistentan niz procjenitelja za θ takav da*

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma(\theta)),$$

gdje je $\Sigma(\theta)$ pozitivno definitna matrica za svaki $\theta \in \Theta$. Pretpostavimo da vrijede sve pretpostavke teorema 4.0.5 i da je funkcija $\theta \mapsto \mathcal{F}(\theta)$ neprekidna. Tada su nizovi definirani sa

$$\tilde{\theta}_n^{(1)} = \tilde{\theta}_n - \ddot{l}_n(\tilde{\theta}_n)^{-1}\dot{l}_n(\tilde{\theta}_n)^T \quad i \quad \theta_n^* = \tilde{\theta}_n + \frac{1}{n}\mathcal{F}(\tilde{\theta}_n)^{-1}\dot{l}_n(\tilde{\theta}_n)^T$$

asimptotski efikasni.

Dokaz. Neka je $(\hat{\theta}_n)_{n\in\mathbb{N}}$ jako konzistentan i asimptotski efikasan niz korijena jednadžbe vjerodostojnosti čiju egzistenciju garantira teorem 4.0.5. Koristeći teorem 4.0.3 dobijemo

$$\dot{l}_n(\tilde{\theta}_n)^T = \dot{l}_n(\hat{\theta}_n)^T + \int_0^1 \ddot{l}_n\left(\hat{\theta}_n + v(\tilde{\theta}_n - \hat{\theta}_n)\right) dv(\tilde{\theta}_n - \hat{\theta}_n). \tag{5.1}$$

Kako je $\dot{l}_n(\hat{\theta}_n) = 0$ za dovoljno velike $n \in \mathbb{N}$, iz (5.1) slijedi

$$\dot{l}_n(\tilde{\theta}_n)^T = \int_0^1 \ddot{l}_n\left(\hat{\theta}_n + v(\tilde{\theta}_n - \hat{\theta}_n)\right) dv(\tilde{\theta}_n - \hat{\theta}_n). \tag{5.2}$$

Nadalje, direktno iz definicije niza $(\tilde{\theta}_n^{(1)})_{n\in\mathbb{N}}$ vidimo da vrijedi

$$(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) = (\tilde{\theta}_n - \hat{\theta}_n) - \ddot{l}_n(\tilde{\theta}_n)^{-1}\dot{l}_n(\tilde{\theta}_n)^T. \tag{5.3}$$

Koristeći (5.2) i (5.3), slijedi

$$\sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) = \left[\mathbf{I} - \ddot{l}_n(\tilde{\theta}_n)^{-1}\int_0^1 \ddot{l}_n\left(\hat{\theta}_n + v(\tilde{\theta}_n - \hat{\theta}_n)\right) dv\right]\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n). \tag{5.4}$$

Pretpostavka je da su zadovoljeni uvjeti teorema 4.0.5 i da su nizovi $(\tilde{\theta}_n)_n$ i $(\hat{\theta}_n)_n$ jako konzistentni, pa primjenom uniformnog jakog zakona velikih brojeva (teorem 2.0.1) na sličan način kao u teoremu 4.0.5, slijedi

$$\frac{1}{n}\ddot{l}_n(\tilde{\theta}_n) \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} -\mathcal{F}(\theta) \ \ i \ \ \frac{1}{n}\int_0^1 \ddot{l}_n\left(\hat{\theta}_n + v(\tilde{\theta}_n - \hat{\theta}_n)\right) dv \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} -\mathcal{F}(\theta).$$

Dakle, vrijedi

$$A_n := \left[\mathbf{I} - \ddot{l}_n(\tilde{\theta}_n)^{-1}\int_0^1 \ddot{l}_n\left(\hat{\theta}_n + v(\tilde{\theta}_n - \hat{\theta}_n)\right) dv\right] \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} \mathbf{0}.$$

Nizovi $(\sqrt{n}(\tilde{\theta}_n - \theta))_{n\in\mathbb{N}}$ i $(\sqrt{n}(\hat{\theta}_n - \theta))_{n\in\mathbb{N}}$ su asimptotski normalni, pa po propoziciji 1.2.12 slijedi da su ograničeni po vjerojatnosti. Kako je $\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) = \sqrt{n}(\tilde{\theta}_n - \theta) - \sqrt{n}(\hat{\theta}_n - \theta)$, slijedi da je i niz $(\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n))_{n\in\mathbb{N}}$ također ograničen po vjerojatnosti, jer je suma dva niza koja su ograničena po vjerojatnosti. Uvedimo oznaku $Y_n = \sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n)$. U nastavku pokazujemo da $\sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) = A_n Y_n \xrightarrow{\mathbf{P}_\theta} \mathbf{0}$. Neka je $\|\cdot\|$ operatorska norma inducirana vektorskom normom $\|\cdot\|_2$. Zbog jednostavnosti u nastavku koristimo istu oznaku za matričnu normu i za vektorsku normu, ali to u ovom slučaju neće stvarati probleme. Trebamo pokazati da za svaki $\varepsilon > 0$ vrijedi

$$\lim_{n\to\infty} \mathbf{P}_\theta\left(\|A_n Y_n\| \geq \varepsilon\right) = 0.$$

Zbog konzistentnosti operatorske norme je $\|A_n Y_n\| \leq \|A_n\| \cdot \|Y_n\|$, pa je

$$\mathbf{P}_\theta\left(\|A_n Y_n\| \geq \varepsilon\right) \leq \mathbf{P}_\theta\left(\|A_n\| \cdot \|Y_n\| \geq \varepsilon\right).$$

Dakle, dovoljno je pokazati

$$\lim_{n\to\infty} \mathbf{P}_\theta\left(\|A_n\| \cdot \|Y_n\| \geq \varepsilon\right) = 0.$$

Neka je $\varepsilon_1 > 0$. Niz $(Y_n)_{n\in\mathbb{N}}$ je ograničen po vjerojatnosti, pa postoje $M > 0$ i $n_0 \in \mathbb{N}$ takvi da za svaki $n \geq n_0$ vrijedi

$$\mathbf{P}_\theta(\|Y_n\| > M) \leq \frac{\varepsilon_1}{2}. \tag{5.5}$$

Nadalje, niz $(A_n)_{n\in\mathbb{N}}$ konvergira gotovo sigurno, pa konvergira i po vjerojatnosti. Zato postoji $n_1 \in \mathbb{N}$, takav da za svaki $n \geq n_1$ vrijedi

$$\mathbf{P}_\theta\left(\|A_n\| \geq \frac{\varepsilon}{M}\right) < \frac{\varepsilon_1}{2}. \tag{5.6}$$

Koristeći (5.5) i (5.6), slijedi da za svaki $n \in \mathbb{N}$, $n \geq \max\{n_0, n_1\}$ vrijedi

$$\begin{aligned}
\mathbf{P}_\theta(\|A_n\| \cdot \|Y_n\| \geq \varepsilon) &= \mathbf{P}_\theta(\|A_n\| \cdot \|Y_n\| \geq \varepsilon,\ \|Y_n\| > M) + \mathbf{P}_\theta(\|A_n\| \cdot \|Y_n\| \geq \varepsilon,\ \|Y_n\| \leq M) \\
&\leq \mathbf{P}_\theta(\|Y_n\| > M) + \mathbf{P}_\theta(\|A_n\| \geq \frac{\varepsilon}{M}) < \varepsilon_1.
\end{aligned}$$

Dakle, niz $\sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n)$ konvergira po vjerojatnosti prema $\mathbf{0}$. Korištenjem teorema 1.2.9 (Slutsky) slijedi

$$\sqrt{n}(\tilde{\theta}_n^{(1)} - \theta) = \sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) + \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathcal{F}(\theta)^{-1}).$$

Na potpuno isti način se pokaže asimptotska efikasnost niza $(\theta_n^*)_{n\in\mathbb{N}}$. □

Za kraj pokažimo jednu primjenu prethodnog teorema.

Primjer 5.0.4. *Neka je $X_1, X_2, \ldots, X_n$ slučajni uzorak iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \mathbf{R}\}$, gdje su gustoće*

$$f(x|\theta) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

U primjeru 4.0.7 smo vidjeli da jednadžba vjerodostojnosti

$$\dot{l}_n(\theta) = 2\sum_{i=1}^{n} \frac{(X_i - \theta)}{1 + (X_i - \theta)^2} = 0$$

ima više korijena koje nije jednostavno odrediti te da nije jasno koji od njih treba odabrati. Dakle, određivanje procjenitelja maksimalne vjerodostojnosti nije sasvim jednostavno u ovom slučaju. Zbog toga u ovom primjeru konstruiramo asimptotski efikasan niz procjenitelja za θ koji je lako odrediti iz danog uzorka. Neka je $X_{(n;k)}$, $k \in \{1, \ldots, n\}$, k-ta po veličini slučajna varijabla u slučajnom uzorku $X_1, X_2, \ldots, X_n$. Dakle, vrijedi $X_{(n;1)} \leq X_{(n;2)} \leq \cdots \leq X_{(n;n)}$. Iako $X_{(n;k)}$ ovisi o $n \in \mathbb{N}$, u nastavku to nećemo posebno naglašavati i koristimo oznaku $X_{(k)} = X_{(n;k)}$. Neka je $m_n = X_{(\lceil \frac{n}{2} \rceil)}$, gdje je za $x \in \mathbb{R}$, $\lceil x \rceil$ najmanji cijeli broj veći ili jednak od x. Pokažimo da je $(m_n)_{n\in\mathbb{N}}$ jako konzistentan niz procjenitelja za θ. Za $\theta \in \mathbb{R}$, neka je $F = F(\cdot|\theta)$ funkcija distribucije od X_1 i $\hat{F}_n$ empirijska funkcija distribucije bazirana na uzorku $X_1, X_2, \ldots, X_n$, tj. vrijedi

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}. \tag{5.7}$$

Odredimo funkciju F:

$$\begin{aligned} F(x) = F(x|\theta) = \int_{-\infty}^{x} f(t|\theta)\,dt &= \int_{-\infty}^{x} \frac{1}{\pi(1 + (t - \theta)^2)}\,dt \\ &= \int_{-\infty}^{x-\theta} \frac{1}{\pi(1 + t^2)}\,dt \\ &= \frac{1}{\pi}\,\mathrm{arctg}(x - \theta) + \frac{1}{2}. \end{aligned} \tag{5.8}$$

Iz (5.8) *slijedi da je θ medijan distribucije F, odnosno da vrijedi*

$$F(\theta) = \frac{1}{2}\,. \tag{5.9}$$

Nadalje, koristeći definiciju od m_n i (5.7) *slijedi da je*

$$m_n = \min\left\{x \in \mathbb{R} \,\middle|\, \hat{F}_n(x) \geq \frac{1}{2}\right\}. \tag{5.10}$$

Neka je $\varepsilon > 0$ proizvoljan. Zbog (5.9) *te kako je funkcija F strogo rastuća slijedi da postoji $\delta > 0$ takav da je:*

$$F(\theta - \varepsilon) < \frac{1}{2} - \delta, \tag{5.11}$$

$$F(\theta + \varepsilon) > \frac{1}{2} + \delta. \tag{5.12}$$

Koristeći jaki zakon velikih brojeva slijedi

$$\forall x \in \mathbb{R}\,, \quad \hat{F}_n(x) \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} F(x),$$

pa na skupu $\mathbf{P}_\theta$-vjerojatnosti jedan postoji $N \in \mathbb{N}$, takav da za sve $n \in \mathbb{N}$, $n \geq N$ vrijedi

$$|\hat{F}_n(\theta - \varepsilon) - F(\theta - \varepsilon)| < \delta, \tag{5.13}$$

$$|\hat{F}_n(\theta + \varepsilon) - F(\theta + \varepsilon)| < \delta. \tag{5.14}$$

Za $n \geq N$ je $|m_n - \theta| \leq \varepsilon$, jer u slučaju da je $m_n < \theta - \varepsilon$, iz (5.10) *i* (5.13) *bi slijedilo*

$$F(\theta - \varepsilon) > \hat{F}_n(\theta - \varepsilon) - \delta \geq \hat{F}_n(m_n) - \delta \geq \frac{1}{2} - \delta,$$

što je kontradikcija, jer je δ izabran tako da vrijedi (5.11). *Isto tako kada bi vrijedilo $m_n > \theta + \varepsilon$, iz* (5.10) *i* (5.14) *bi slijedilo*

$$F(\theta + \varepsilon) < \hat{F}_n(\theta + \varepsilon) + \delta \leq \frac{1}{2} + \delta,$$

što je kontradikcija, jer je δ izabran tako da vrijedi (5.12). *Dakle, vrijedi $m_n \xrightarrow{\mathbf{P}_\theta\text{-}g.s.} \theta$, odnosno niz procjenitelja $(m_n)_{n\in\mathbb{N}}$ je jako konzistentan. Nadalje, može se pokazati (pogledati [4, str. 91]) da vrijedi*

$$\sqrt{n}(m_n - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{\pi^2}{4}).$$

Dakle, $(m_n)_{n\in\mathbb{N}}$ je jako konzistentan niz procjenitelja za θ takav da vrijedi

$$\sqrt{n}(m_n - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{\pi^2}{4}).$$

Također, zadovoljene su i pretpostavke teorema 4.0.5, pa prema teoremu 5.0.3 slijedi da su nizovi

$$m_n^* = m_n + \frac{1}{n}\mathcal{F}(m_n)^{-1}\dot{l}_n(m_n), \tag{5.15}$$

$$\tilde{m}_n^{(1)} = m_n - \ddot{l}_n(m_n)^{-1}\dot{l}_n(m_n) \tag{5.16}$$

asimptotski efikasni nizovi procjenitelja za θ. Kako je u ovom primjeru Fisherova informacija vrlo jednostavna, $\mathcal{F}(\theta) = \frac{1}{2}$, niz (5.15) *će biti jednostavniji od niza* (5.16). *Uvrštavanjem*

$$\mathcal{F}(m_n) = \frac{1}{2} \;\; i \;\; \dot{l}_n(m_n) = 2\sum_{i=1}^{n} \frac{(X_i - m_n)}{1 + (X_i - m_n)^2}$$

u (5.15) *slijedi*

$$m_n^* = m_n + \frac{4}{n}\sum_{i=1}^{n} \frac{(X_i - m_n)}{1 + (X_i - m_n)^2}.$$

Uočimo da je u prethodnom primjeru, slično kao i u primjeru 5.0.2, za svaki $\theta \in \Theta$ asimptotska varijanca asimptotski efikasnog niza procjenitelja manja od asimptotske varijance početnog niza procjenitelja. Postavlja se pitanje je li općenito asimptotska varijanca asimptotski efikasnog niza procjenitelja manja od asimptotske varijance bilo kojeg asimptotski normalnog niza procjenitelja, međutim to nažalost nije točno. Sljedeći primjer pokazuje da možemo naći asimptotski normalan niz procjenitelja koji može imati asimptotsku varijancu manju od $\mathcal{F}(\theta)^{-1}$. Takav niz procjenitelja nazivamo superefikasnim nizom procjenitelja.

Primjer 5.0.5. *Neka je $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$ slučajni uzorak iz normalnog modela $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. Sličnim računom kao u primjeru 4.0.1, slijedi da je za ovaj model $\mathcal{F}(\theta) = 1$, za svaki $\theta \in \mathbb{R}$. Definirajmo niz procjenitelja*

$$\hat{\theta}_n = \begin{cases} \overline{X}_n, & \left|\overline{X}_n\right| \geq n^{-\frac{1}{4}} \\ a\overline{X}_n, & \left|\overline{X}_n\right| < n^{-\frac{1}{4}}, \end{cases} \tag{5.17}$$

gdje je $a \in \mathbb{R}$, takav da je $0 < a < 1$. Neka je $\theta \in \mathbb{R}$, $\theta \neq 0$. Za proizvoljni $\varepsilon > 0$ imamo

$$\begin{aligned} \mathbf{P}_\theta\left(\sqrt{n}\left|\overline{X}_n - \hat{\theta}_n\right| \geq \varepsilon\right) &\leq \mathbf{P}_\theta\left(\overline{X}_n \neq \hat{\theta}_n\right) \\ &\leq \mathbf{P}_\theta\left(|\overline{X}_n| < n^{-\frac{1}{4}}\right) \\ &\leq \mathbf{P}_\theta\left(|\theta| - |\overline{X}_n - \theta| < n^{-\frac{1}{4}}\right) \\ &= \mathbf{P}_\theta\left(|\overline{X}_n - \theta| > |\theta| - n^{-\frac{1}{4}}\right) \\ &= \mathbf{P}_\theta\left(|\sqrt{n}(\overline{X}_n - \theta)| > \sqrt{n}|\theta| - n^{\frac{1}{4}}\right). \end{aligned} \tag{5.18}$$

Vrijedi $\sqrt{n}|\theta| - n^{\frac{1}{4}} \xrightarrow{n\to\infty} +\infty$, jer je $|\theta| > 0$. Nadalje, za svaki $n \in \mathbb{N}$ je $\sqrt{n}(\overline{X}_n - \theta) \sim \mathcal{N}(0, 1)$, pa slijedi

$$\mathbf{P}_\theta\left(\left|\sqrt{n}(\overline{X}_n - \theta)\right| > \sqrt{n}|\theta| - n^{\frac{1}{4}}\right) \xrightarrow{n\to\infty} 0. \tag{5.19}$$

Zbog (5.19) *iz* (5.18) *slijedi*

$$\sqrt{n}\left(\overline{X}_n - \hat{\theta}_n\right) \xrightarrow{\mathbf{P}_\theta} 0. \tag{5.20}$$

Koristeći teorem 1.2.9 (Slutsky) i (5.20) *slijedi*

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(\hat{\theta}_n - \overline{X}_n) + \sqrt{n}(\overline{X}_n - \theta) \xrightarrow{D} \mathcal{N}(0, 1).$$

Neka je $\theta = 0$. *Tada za svaki* $\varepsilon > 0$ *vrijedi*

$$\begin{aligned} \mathbf{P}_0\left(\sqrt{n}\left|\hat{\theta}_n - a\overline{X}_n\right| \geq \varepsilon\right) &\leq \mathbf{P}_0\left(\hat{\theta}_n \neq a\overline{X}_n\right) \\ &\leq \mathbf{P}_0\left(\left|\overline{X}_n\right| \geq n^{-\frac{1}{4}}\right) \\ &= \mathbf{P}_0\left(\left|\sqrt{n}\cdot\overline{X}_n\right| \geq n^{\frac{1}{4}}\right) \xrightarrow{n\to\infty} 0. \end{aligned} \tag{5.21}$$

Koristeći (5.21) *slijedi da* $\sqrt{n}(\hat{\theta}_n - a\overline{X}_n) \xrightarrow{\mathbf{P}_0} 0$, *pa primjenom teorema 1.2.9 (Slutsky) slijedi*

$$\sqrt{n}(\hat{\theta}_n - 0) = \sqrt{n}(\hat{\theta}_n - a\overline{X}_n) + a\sqrt{n}\cdot\overline{X}_n \xrightarrow{D} \mathcal{N}(0, a^2).$$

Dakle, pokazali smo da je niz $(\hat{\theta}_n)_{n\in\mathbb{N}}$ *asimptotski normalan, tj. vrijedi*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2(\theta)),$$

gdje je

$$\sigma^2(\theta) = \begin{cases} 1, & \theta \neq 0 \\ a^2, & \theta = 0 \end{cases}. \tag{5.22}$$

Kako je $a^2 < 1$, *slijedi da je* $\sigma^2(0) < \mathcal{F}(0)^{-1}$.

Bibliografija

[1] *Maximum Likelihood Estimates*, https://www.stat.purdue.edu/~dasgupta/ml.pdf, posjećena 04.08.2021.

[2] R. Durrett, *Probability: Theory and Examples*, https://www.math.ucla.edu/~biskup/275a.1.20f/PDFs/Durrett-v5.pdf, posjećena 03.08.2021.

[3] G. Casella E. L. Lehmann, *Theory of Point Estimation*, Springer, New York, 1998.

[4] T. S. Ferguson, *A Course in Large Sample Theory*, Chapman and Hall, London, 1996.

[5] Thomas S. Ferguson, *An Inconsistent Maximum Likelihood Estimate*, Journal of the American Statistical Association **77** (1982), 831–834, https://www.jstor.org/stable/2287314.

[6] C. J. Geyer, *The Wald Consistency Theorem*, https://www.stat.umn.edu/geyer/8112/notes/wald.pdf, posjećena 03.08.2021.

[7] D. Hunter, *Statistics 597A, Asymptotic Tools, Lecture notes: topics 27-28*, http://personal.psu.edu/drh20/asymp/fall2002/lectures/ln12.pdf, posjećena 03.08.2021.

[8] M. Huzak, *Matematička statistika, Predavanja poglavlje 3*, https://web.math.pmf.unizg.hr/nastava/ms/files/ms3v2.pdf, posjećena 03.08.2021.

[9] M. Huzak, *Matematička statistika, Predavanja poglavlje 4*, https://web.math.pmf.unizg.hr/nastava/ms/files/ms4v2.pdf, posjećena 03.08.2021.

[10] E. L. Lehmann, *Elements of Large-Sample Theory*, Springer, New York, 1999.

[11] S. Mardešić, *Matematička analiza u n-dimenzionalnom realnom prostoru*, Školska knjiga, Zagreb, 1974.

[12] D. Pati, *Maximum Likelihood Estimators*, https://ani.stat.fsu.edu/~debdeep/mle.pdf, posjećena 03.08.2021.

[13] D. Pollard, *Convergence of Stochastic Processes*, Springer, New York, 1984.

[14] N. Sarapa, *Teorija Vjerojatnosti*, Školska knjiga, Zagreb, 2002.

Sažetak

U ovom radu se bavimo asimptotskim svojstvima niza procjenitelja maksimalne vjerodostojnosti. Na početku uvodimo osnovne pojmove matematičke statistike te navodimo rezultate iz teorije vjerojatnosti koje koristimo u ostatku rada. U drugom poglavlju iskazujemo i dokazujemo jednu verziju uniformnog jakog zakona velikih brojeva. Glavni rezultat rada je konzistentnost i asimptotska normalnost niza procjenitelja maksimalne vjerodostojnosti i obrađen je kroz treće i četvrto poglavlje. Uz to u navedenim poglavljima ilustriramo potencijalne probleme procjenitelja maksimalne vjerodostojnosti, ali i lijepo ponašanje na velikoj klasi često korištenih modela, kao na primjer u jednoparametarskoj eksponencijalnoj familiji. Na kraju uvodimo pojam asimptotske efikasnosti i pokazujemo kako se uz određene pretpostavke može iz konzistentnog i asimptotski normalnog niza procjenitelja konstruirati asimptotski efikasan niz procjenitelja.

Summary

In this thesis, we deal with the asymptotic properties of a sequence of maximum likelihood estimators. At the beginning, we introduce the basic concepts of mathematical statistics and list the results from probability theory that we use in the rest of the thesis. In the second chapter we state and prove one version of the uniform strong law of large numbers. The main result of the thesis is the consistency and asymptotic normality of a sequence of maximum likelihood estimators and it is processed through the third and fourth chapters. In addition, in the above chapters we illustrate the potential problems of the maximum likelihood estimator, but also the nice behavior on a large class of frequently used models, such as in a one-parameter exponential family. Finally, we introduce the notion of asymptotic efficiency and show how, under certain assumptions, an asymptotically efficient sequence of estimators can be constructed from a consistent and asymptotically normal sequence of estimators.

Životopis

Rođen sam 18. veljače 1997. godine u Dubrovniku. Srednju školu završavam u Dubrovniku 2016. godine i nakon državne mature iste godine upisujem preddiplomski sveučilišni studij Matematika na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Godine 2019. stekao sam titulu sveučilišnog prvostupnika matematike, nakon čega u rujnu iste godine upisujem diplomski sveučilišni studij Matematička statistika na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Tijekom preddiplomskog i diplomskog studija bio sam demonstrator iz kolegija: Linearna algebra 1, Linearna algebra 2, Euklidski prostori, Kompleksna analiza i Statistika. Na kraju oba studija, odlukom Vijeća Matematičkog odsjeka Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, nagrađen sam za izniman uspjeh.