Asimptotska svojstva Bayesovog procjenitelja

Kiralj, Josip

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Josip Kiralj

ASIMPTOTSKA SVOJSTVA BAYESOVOG PROCJENITELJA

Diplomski rad

Voditelj rada:
Izv. prof. dr. sc. Miljenko Huzak

Zagreb, rujan 2018.

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

1. ______________________________, predsjednik
2. ______________________________, član
3. ______________________________, član

Povjerenstvo je rad ocijenilo ocjenom ________________.

Potpisi članova povjerenstva:

1. ______________________________
2. ______________________________
3. ______________________________

Zahvaljujem se izv. prof. dr. sc. Miljenku Huzaku na izdvojenom vremenu i pomoći pri izradi ovog diplomskog rada. Hvala mojoj obitelji, prijateljima i djevojci na podršci tijekom studija.

Sadržaj

Sadržaj **iv**

Uvod **1**

1 Osnovne pretpostavke **2**

2 Bayesov procjenitelj **7**

3 Konzistentnost aposteriorne distribucije **11**

4 Asimptotska normalnost **27**
4.1 Asimptotska normalnost aposteriorne distribucije 27
4.2 Asimptotska normalnost Bayesovog procjenitelja 39

Bibliografija **48**

Uvod

Statističkom zaključivanju možemo pristupiti na dva načina: klasičnim i bayesovskim. U klasičnom pristupu pretpostavljamo da su podaci realizacija slučajne varijable/vektora čija je distribucija indeksirana nekim parametrom θ. Taj parametar je nepoznat i fiksan te je cilj klasičnog pristupa pronaći vjerodostojnu procjenu parametra od interesa ili barem odrediti dovoljno dobro područje kojemu taj parametar pripada, odnosno ne pripada. U bayesovskom pristupu situacija je drugačija. Uz navedeno, pretpostavljamo da je parametar θ slučajna veličina sa poznatom, apriornom distribucijom koju zadajemo, a reprezentira naše prijašnje znanje o parametru. Uzimajući u obzir podatke, tu distribuciju mijenjamo te time dobivamo aposteriornu distribuciju koja nam opisuje parametar na temelju podataka i našeg apriornog znanja.
U ovom radu će biti riječ o asimptotskim rezultatima u Bayesovoj statistici. Rad je podijeljen u 4 poglavlja. U prvom poglavlju uvodimo osnovne pojmove i oznake korištene u daljnjem radu te ćemo pokazat na koji način se konstruira aposteriorna distribucija iz apriorne distribucije i danih podataka. U drugom poglavlju ćemo uvesti pojam Bayesovog procjenitelja, o čemu ovisi te na primjerima ilustrirati njegovu konstrukciju. U trećem poglavlju se bavimo pojmom konzistentnosti aposteriorne distribucije, pojmom koji na neki način opravdava bayesovski pristup sa gledišta klasične statistike. U zadnjem poglavlju se bavimo asimptotskim rezultatima analognim asimptotskoj normalnosti procjenitelja maksimalne vjerodostojnosti u regularnim statističkim modelima. Pokazujemo asimptotsku normalnost aposteriorne distribucije. Posljedično tome, dokazujemo asimptotsku ekvivalenciju Bayesovih procjenitelja sa procjeniteljem maksimalne vjerodostojnosti iz čega će zatim slijediti i njihova asimptotska normalnost.

Poglavlje 1

Osnovne pretpostavke

U ovom poglavlju ćemo uvesti osnovne pojmove bayesovske statistike i pretpostavke važne za razumijevanje daljnjeg rada.

Neka je $(S, \mathcal{S})$ izmjeriv prostor i Θ potpun, separabilan metrički prostor te $\mathcal{B}(\Theta)$ pripadna Borelova σ-algebra. Θ nazivamo parametarski prostor. Izmjerivu funkciju $\boldsymbol{\theta} : S \to \Theta$ ćemo zvati parametar. Svaku vjerojatnosnu mjeru Π na parametarskom prostoru $(\Theta, \mathcal{B}(\Theta))$ zovemo apriorna distribucija parametra $\boldsymbol{\theta}$. Neka je $(\mathbf{X}, \mathcal{A})$ izmjeriv prostor i $X : S \to \mathbf{X}$ slučajna varijabla. Označimo sa $\mathcal{P} = \{ P_\theta,\ \theta \in \Theta\}$ parametarsku familiju distribucija od X obzirom na parametarski prostor Θ. Pretpostavljamo da je parametrizacija injektivna, odnosno

$$\theta, \theta' \in \Theta, \theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}.$$

Promatramo $(X_n, n \in \mathbb{N})$ niz slučajnih varijabli koje poprimaju vrijednosti u $\mathbf{X}$ te neka su one nezavisne i jednako distribuirane obzirom na P_θ za svaki $\theta \in \Theta$. Sa P_θ^n ćemo označavati n-terostruki produkt vjerojatnosne mjere P_θ. Neka je P_θ^∞ mjera na $(\mathbf{X}^\infty, \mathcal{A}^\infty)$ zadana konačnodimenzionalnim distribucijama niza $(X_n, n \in \mathbb{N})$. Pretpostavljamo da je $\theta \mapsto P_\theta(A)$ slučajna varijabla na $(\Theta, \mathcal{B}(\Theta))$ za svako fiksno $A \in \mathcal{A}$. Uz dane pretpostavke postoji[1] jedinstvena vjerojatnost $\lambda_{n,\Pi}$ na $\mathcal{B}(\Theta) \times \mathcal{A}^n$ tako da vrijedi

$$\lambda_{n,\Pi}(B \times A) = \int_B P_\theta^n(A)\, d\Pi(\theta), \quad B \in \mathcal{B}(\Theta),\ A \in \mathcal{A}^n, \tag{1.1}$$

$$\lambda_{n,\Pi}(F) = \int_\Theta P_\theta^n(F_\theta)\, d\Pi(\theta), \quad F \in \mathcal{B}(\Theta) \times \mathcal{A}^n, \tag{1.2}$$

[1]Pogledati [9], teorem 10.15.

gdje je $F_\theta := \{x \in \mathbf{X}^n\ :\ (\theta, x) \in F\}$ prerez od F u $\theta \in \Theta$. Marginalna distribucija slučajnog vektora $\mathbf{X}_n := (X_1, X_2, \ldots, X_n)$ je tada dana izrazom

$$\lambda_n(A) := \int_\Theta P_\theta^n(A)\, d\Pi(\theta),\ \ A \in \mathcal{A}^n. \tag{1.3}$$

Analogno definiramo vjerojatnosnu mjeru λ_Π na $\mathcal{B}(\Theta) \times \mathcal{A}^\infty$ i marginalnu distribuciju vremenskog niza $(X_n, n \in \mathbb{N})$ koju ćemo označiti sa λ. Aposteriornom distribucijom parametra $\boldsymbol{\theta}$ za dano $X_1 = x_1, X_2 = x_2, \ldots X_n = x_n$ nazivamo svaku uvjetnu distribuciju parametra $\boldsymbol{\theta}$ uz dano $\mathbf{X}_n = \mathbf{x}_n$. Formalizirajmo sada aposteriornu distribuciju kao funkciju na $\mathcal{B}(\Theta) \times \mathbf{X}^n$.

Definicija 1.0.1.
Funkcija $\Pi(\cdot|\cdot) : \mathcal{B}(\Theta) \times \boldsymbol{X}^n \to [0, 1]$ *se naziva aposteriorna distribucija uz dano* $X_1, X_2, \ldots, X_n$ *ako vrijedi sljedeće:*

1. *Za svaki* $\boldsymbol{x}_n \in \boldsymbol{X}^n$, $\Pi(\cdot|\boldsymbol{x}_n)$ *je vjerojatnosna mjera na* $\mathcal{B}(\Theta)$,
2. *Za svaki* $B \in \mathcal{B}(\Theta)$, $\Pi(B|\cdot)$ *je* $\mathcal{A}^n$ *izmjeriva funkcija,*
3. *Za svaki* $B \in \mathcal{B}(\Theta)$, $A \in \mathcal{A}^n$ *vrijedi*

$$\lambda_{n,\Pi}(B \times A) = \int_A \Pi(B|\boldsymbol{x}_n)\, d\lambda_n(\boldsymbol{x}_n).$$

Budući da aposteriorna distribucija ovisi o realizacijama slučajnog vektora $\mathbf{X}_n$, smatrat ćemo ju slučajnom i označavati ju sa $\Pi(\cdot|\mathbf{X}_n)$. Pretpostavimo sada da je za svaki θ, P_θ apsolutno neprekidna mjera obzirom na neku σ-konačnu mjeru μ na $(\mathbf{X}, \mathcal{A})$ i stavimo $f_{X|\boldsymbol{\theta}}(x|\theta) := \frac{dP_\theta}{d\mu}(x)$. Tada će isto vrijediti za P_θ^n obzirom na mjeru μ^n pa stavimo $f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\theta) := \frac{dP_\theta^n}{d\mu^n}(\mathbf{x}_n)$. Nadalje, za statistički model koji promatramo pretpostavljamo određeni uvjet regularnosti, a to je da je nosač gustoće, $\operatorname{supp} f(\cdot|\theta) := \{x \in \mathbf{X} : f(x|\theta) > 0\}$ neovisan o $\theta \in \Theta$. Sljedeći teorem nam daje konstrukciju aposteriorne distribucije.

Teorem 1.0.2. *(Bayesov teorem)*
Neka je za svaki θ, P_θ *apsolutno neprekidna mjera obzirom na* σ*-konačnu mjeru* μ. *Tada je* $\Pi(\cdot|\boldsymbol{x}_n)$ *apsolutno neprekidna u odnosu na* Π *za gotove sve realizacije* $\boldsymbol{x}_n \in \boldsymbol{X}^n$ *te je Radon-Nikodymova derivacija dana sa*

$$\frac{d\Pi(\cdot|\boldsymbol{x}_n)}{d\Pi}(\theta|\boldsymbol{x}_n) = \frac{f_{X_n|\boldsymbol{\theta}}(\boldsymbol{x}_n|\theta)}{\int_\Theta f_{X_n|\boldsymbol{\theta}}(\boldsymbol{x}_n|\xi) d\Pi(\xi)}. \tag{1.4}$$

Nadalje,

$$\lambda_n\left(\boldsymbol{x}_n\ :\ \int_\Theta f_{X_n|\boldsymbol{\theta}}(\boldsymbol{x}_n|\xi)\, d\Pi(\xi) \in \{0, \infty\}\right) = 0. \tag{1.5}$$

Dokaz.
Dokažimo prvo tvrdnju (1.5). Neka su

$$C_0 := \left\{ \mathbf{x}_n \in \mathbf{X}^n \ : \ \int_\Theta f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi)\, d\Pi(\xi) = 0 \right\},$$

$$C_\infty := \left\{ \mathbf{x}_n \in \mathbf{X}^n \ : \ \int_\Theta f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi)\, d\Pi(\xi) = \infty \right\}.$$

Po definiciji od λ_n (1.3) i uz korištenje Fubinijevog teorema, vrijedi sljedeće:

$$\lambda_n(C_0) = \int_{C_0} \int_\Theta f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi)\, d\Pi(\xi)\, d\mu^n(\mathbf{x}_n) = \int_{C_0} 0\, d\mu^n(\mathbf{x}_n) = 0,$$

$$\lambda_n(C_\infty) = \int_{C_\infty} \int_\Theta f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi) d\Pi(\xi)\, d\mu^n(\mathbf{x}_n) = \int_{C_\infty} \infty\, d\mu^n(\mathbf{x}_n).$$

Kako je λ_n konačna mjera mora vrijediti $\mu^n(C_\infty) = 0$ pa imamo $\lambda_n(C_\infty) = 0$. Dakle, tvrdnja vrijedi. Pokažimo sada da je $\Pi(\cdot|\mathbf{x}_n)$ dana sa (1.4) aposteriorna distribucija od $\boldsymbol{\theta}$. Svojstvo 1. iz definicije 1.0.1 je očito, a svojstvo 2. slijedi iz Fubinijevog teorema. Pokažimo da je zadovoljeno svojstvo 3. Uzmimo proizvoljne $B \in \mathcal{B}(\Theta)$, $A \in \mathcal{A}^n$:

$$\begin{aligned}
\lambda_{n,\Pi}(B \times A) &\overset{(1.1)}{=} \int_B \int_A f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi) d\mu^n(\mathbf{x}_n)\, d\Pi(\xi) \\
&\overset{Fubini}{=} \int_A \int_B f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi) d\Pi(\xi)\, d\mu^n(\mathbf{x}_n) \\
&\overset{(1.4)}{=} \int_A \left(\Pi(B|\mathbf{x}_n) \int_\Theta f(\mathbf{x}_n|\xi)\, d\Pi(\xi) \right) d\mu^n(\mathbf{x}_n) \\
&= \int_\Theta \int_A \Pi(B|\mathbf{x}_n) f(\mathbf{x}_n|\xi)\, d\mu^n(\mathbf{x}_n)\, d\Pi(\xi). \qquad (1.6)
\end{aligned}$$

Za $A \in \mathcal{A}^n$ vrijedi:

$$\int_A d\lambda_n(\mathbf{x}_n) \overset{(1.3)}{=} \int_\Theta P^n_\xi(A)\, d\Pi(\xi) = \int_\Theta \int_A f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi)\ d\mu^n(\mathbf{x}_n)\, d\Pi(\xi). \qquad (1.7)$$

Tvrdnja (1.7) se Lebesgueovom indukcijom lako proširi na izmjerive funkcije. Kako je za svaki $B \in \mathcal{B}(\Theta)$, $\Pi(B|\cdot)$ nenegativna izmjeriva funkcija zaključujemo:

$$\int_A \Pi(B|\mathbf{x}_n)\, d\lambda_n(\mathbf{x}_n) = \int_\Theta \int_A \Pi(B|\mathbf{x}_n) f_{\mathbf{X}_n|\boldsymbol{\theta}}(\mathbf{x}_n|\xi)\ d\mu^n(\mathbf{x}_n)\, d\Pi(\xi)$$

$$\overset{(1.6)}{=} \lambda_{n,\Pi}(B \times A).$$

Time smo pokazali da je zadovoljeno svojstvo 3. iz definicije. □

Pokažimo sada na nekim primjerima računanje aposteriorne distribucije.

Primjer 1.0.3.

(a) *Neka je $\Theta = [0,1]$. Pretpostavimo da su $X_1, X_2, \dots, X_n$ nezavisne i Bernoullijeve slučajne varijable sa parametrom θ te neka je apriorna distribucija $\Pi \sim Beta(\alpha,\beta)$ i π gustoća od Π. Odredimo aposteriornu distribuciju. Označimo sa $r := \sum_{i=1}^n x_i$ i $\boldsymbol{x}_n = (x_1, x_2, \dots x_n)$.*

$$\pi(\theta|\boldsymbol{x}_n) = \frac{f(\boldsymbol{x}_n|\theta)\pi(\theta)}{\int_\Theta f(\boldsymbol{x}_n|\xi) d\Pi(\xi)} = \frac{f(\boldsymbol{x}_n|\theta)\pi(\theta)}{\int_0^1 f(\boldsymbol{x}_n|\xi)\pi(\xi) d\xi}$$

$$= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^r (1-\theta)^{n-r}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \xi^r (1-\xi)^{n-r}\, \xi^{\alpha-1}(1-\xi)^{\beta-1} d\xi}$$

$$= \frac{\theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1}}{B(\alpha+r, \beta+n-r)}.$$

Aposteriorna distribucija je beta distribucija sa parametrima $\alpha + r$ i $\beta + n - r$. Dakle, $\boldsymbol{\theta} \mid \boldsymbol{X}_n \sim Beta(\alpha + \sum_{i=1}^n X_i, \beta + n - \sum_{i=1}^n X_i)$.

(b) *Neka je $\Theta = \mathbb{R}$ i neka su $X_1, X_2, \dots, X_n$ nezavisne slučajne varijable koje imaju normalnu distribuciju sa očekivanjem θ i varijancom σ^2. Stavimo da je apriorna distribucija parametra $\boldsymbol{\theta} \sim N(\mu, \tau^2)$. Odredimo prvo nazivnik u izrazu aposteriorne gustoće:*

$$\int_\Theta f(\boldsymbol{x}_n|\theta)d\Pi(\theta) = \int_\mathbb{R} f(\boldsymbol{x}_n|\theta)\pi(\theta)\, d\theta$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n}\frac{1}{\sqrt{2\pi}\,\tau}\int_\mathbb{R} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2}e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}\, d\theta$$

$$= \frac{1}{(2\pi)^{(n+1)/2}\sigma^n\tau}\int_\mathbb{R} e^{\frac{-1}{2\sigma^2\tau^2}(\theta^2(n\tau^2+\sigma^2)-2\theta(\tau^2\sum_{i=1}^n x_i+\sigma^2\mu)+\sigma^2\mu^2+\tau^2\sum_{i=1}^n x_i^2)}d\theta$$

$$= \frac{1}{(2\pi)^{(n+1)/2}\sigma^n\tau}e^{(\sigma^2\mu^2+\tau^2\sum_{i=1}^n x_i^2)}\int_\mathbb{R} e^{\frac{-(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\theta^2-2\theta\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)}d\theta$$

$$= \frac{e^{(\sigma^2\mu^2+\tau^2\sum_{i=1}^n x_i^2)}}{(2\pi)^{(n+1)/2}\sigma^n\tau}e^{\frac{(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)^2}\int_\mathbb{R} e^{\frac{-(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\theta-\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)^2}d\theta$$

$$= \frac{e^{(\sigma^2\mu^2+\tau^2\sum_{i=1}^n x_i^2)}}{(2\pi)^{(n+1)/2}\sigma^n\tau}e^{\frac{(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)^2}\frac{\sqrt{2\pi}\sigma\tau}{\sqrt{n\tau^2+\sigma^2}}.$$

Brojnik je jednak:

$$\frac{1}{(2\pi)^{(n+1)/2}\sigma^n\tau}e^{(\sigma^2\mu^2+\tau^2\sum_{i=1}^n x_i^2)}e^{\frac{-(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\theta^2-2\theta\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)}.$$

Uvrštavanjem gornjih izraza u formulu za aposteriornun gustoću imamo:

$$\frac{\sqrt{n\tau^2+\sigma^2}}{\sqrt{2\pi}\sigma\tau}e^{\frac{-(n\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\theta-\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^n x_i+\frac{\mu}{\tau^2})\right)^2}.$$

Vidimo da je aposteriorna distribucija normalna sa očekivanjem $\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}(\frac{n}{\sigma^2}\overline{x}_n+\frac{\mu}{\tau^2})$ *i varijancom* $\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}$.

Napomena 1.0.4.
U primjeru 1.0.3. *b*) *su apriorna i aposteriorna distribucija normalne, dakle pripadaju istoj familiji distribucija. Općenito, apriorne distribucije za koje je aposteriorna iz iste familije distribucija kao i apriorna zovemo konjugiranim apriornim distribucijama.*

Poglavlje 2

Bayesov procjenitelj

U ovom poglavlju pretpostavljamo da je $\Theta \subset \mathbb{R}^k$. Neka je $g : \Theta \to \mathbb{R}^k$ izmjerivo preslikavanje. Procjeniteljem od $g(\theta)$ nazivamo svaku izmjerivu funkciju $\delta : \mathbf{X}^n \to \mathbb{R}^k$. Da bi imalo smisla procjenjivati $g(\theta)$ sa δ pretpostavljamo da je slika procjenitelja $\delta(\mathbf{X}^n)$ sadržana u $g(\Theta)$. Nadalje, zanima nas koliko smo dobro procijenili $g(\theta)$. Za to ćemo koristiti funkciju gubitka.

Definicija 2.0.1.
Funkcija $L : \Theta \times \mathbb{R}^k \to \mathbb{R}$ koja je nenegativna i za koju vrijedi $L(\theta, g(\theta)) = 0$ zovemo funkcijom gubitka za $g(\theta)$.

Uvodimo funkciju rizika:

$$R(\theta, \delta) = \int L(\theta, \delta(\mathbf{x}_n))\ dP_\theta^n(\mathbf{x}_n).$$

Kako je u bayesovkoj statistici parametar $\boldsymbol{\theta}$ slučajna varijabla, računamo aposteriorni rizik:

$$E(R(\boldsymbol{\theta}, \delta)) = \int R(\theta, \delta)\ d\Pi(\theta) = \int\int L(\theta, \delta(\mathbf{x}_n))\ dP_\theta^n(\mathbf{x}_n)\ d\Pi(\theta) \tag{2.1}$$

$$= \int\int L(\theta, \delta(\mathbf{x}_n))\ d\Pi(\theta \mid \mathbf{x}_n)\, d\lambda_n(\mathbf{x}_n). \tag{2.2}$$

Sada možemo definirati Bayesov procjenitelj obzirom na funkciju gubitka L i apriornu distribuciju Π.

Definicija 2.0.2.
Procjenitelj $\delta(X_n)$ koji minimizira aposteriorni rizik (2.1) *nazivamo Bayesov procjenitelj od $\boldsymbol{\theta}$ obzirom na funkciju gubitka L i apriornu distribuciju* Π.

Reći ćemo da je procjenitelj δ konačnog rizika ako je $R(\theta, \delta) < \infty$ $\Pi - g.s.$ Sljedeći teorem nam daje dovoljne uvjete za postojanje Bayesovog procjenitelja te kako ga izračunati.

Teorem 2.0.3.
Neka je $g(\boldsymbol{\theta})$ slučajna varijabla te Π apriorna distribucija od $\boldsymbol{\theta}$. Pretpostavimo da za funkcija gubitka $L(\theta, d)$ vrijedi sljedeće:

1. *Postoji procjenitelj δ sa konačnim rizikom;*

2. *Za gotovo svaki $\boldsymbol{x}_n$ postoji $\delta_\Pi(\boldsymbol{x}_n)$ koji minimizira po $\delta(\boldsymbol{x}_n)$ funkcional*

$$E(L(\boldsymbol{\theta}, \delta(\boldsymbol{x}_n)) \mid \boldsymbol{X}_n = \boldsymbol{x}_n) := \int L(\theta, \delta(\boldsymbol{x}_n)) \; d\Pi(\theta \mid \boldsymbol{x}_n). \tag{2.3}$$

Tada je $\delta_\Pi(\boldsymbol{X}_n)$ Bayesov procjenitelj.

Prije dokaza teorema napomenimo da je za $\psi(\mathbf{x}_n) = E(L(\boldsymbol{\theta}, \delta(\mathbf{x}_n)) \mid \mathbf{X}_n = \mathbf{x}_n)$, $\psi(\mathbf{X}_n)$ verzija uvjetnog očekivanja za $L(\Theta, \delta(\mathbf{X}_n))$ uz danu $\sigma-$algebru generiranu sa $\mathbf{X}_n$. Funkciju $\psi(\mathbf{X}_n)$ ćemo označavati sa $L(\Theta, \delta(\mathbf{X}_n) \mid \mathbf{X}_n)$.

Dokaz.
Neka je δ proizvoljan procjenitelj sa konačnim rizikom. Zbog pretpostavke 2. i nenegativnosti funkcije gubitka je i δ_Π konačnog rizika. Nadalje, imamo

$$E(L(\boldsymbol{\theta}, \delta(\mathbf{x}_n)) \mid \mathbf{X}_n = \mathbf{x}_n) \geq E(L(\boldsymbol{\theta}, \delta_\Pi(\mathbf{x}_n)) \mid \mathbf{X}_n = \mathbf{x}_n) \quad \lambda_n - g.s.$$

Kako su izrazi s lijeve i desne strane nenegativni, integriranjem s obzirom na mjeru λ_n i korištenjem monotonosti integrala slijedi tvrdnja. □

Navedimo neke primjere funkcije gubitka te odredimo pripadne Bayesove procjenitelje.

Primjer 2.0.4.

(a) *Neka je $\Theta \subset \mathbb{R}$ i $L(\theta, d) = (d - g(\theta))^2$. $L(\theta, d)$ zovemo funkcija kvadratnog gubitka. Pretpostavimo da postoji procjenitelj konačnog rizika. Radi kraćeg zapisa stavimo $\boldsymbol{x}_n$ umjesto $\boldsymbol{X}_n = \boldsymbol{x}_n$. Imamo sljedeće:*

$$E(L(\boldsymbol{\theta}, \delta(\boldsymbol{x}_n)) \mid \boldsymbol{x}_n) = E\Big((g(\boldsymbol{\theta}) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n) + E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n) - \delta(\boldsymbol{x}_n))^2 \mid \boldsymbol{x}_n\Big)$$

$$= E\Big((g(\boldsymbol{\theta}) - E\,(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))^2 \mid \boldsymbol{x}_n\Big) + E\Big((\delta(\boldsymbol{x}_n) - E\,(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))^2 \mid \boldsymbol{x}_n\Big)$$

$$+ 2\,E\Big((g(\boldsymbol{\theta}) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))(\delta(\boldsymbol{x}_n) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n)) \mid \boldsymbol{x}_n\Big)$$

$$= \Big(E((g(\boldsymbol{\theta}) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))(\delta(\boldsymbol{x}_n) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n)) \mid \boldsymbol{x}_n) \;= 0\Big)$$

$$= E((g(\boldsymbol{\theta}) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))^2 \mid \boldsymbol{x}_n) + E((\delta(\boldsymbol{x}_n) - E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n))^2 \mid \boldsymbol{x}_n).$$

Vidimo da je regresijska funkcija $E(L(\boldsymbol{\theta}, \delta(\boldsymbol{x}_n)) \mid \boldsymbol{x}_n)$ minimizirana sa regresijskom funkcijom $\delta(\boldsymbol{x}_n) = E(g(\boldsymbol{\theta}) \mid \boldsymbol{x}_n)$. Primjenom teorema 2.0.3 *slijedi da je Bayesov procjenitelj za $g(\theta)$ obzirom na kvadratnu funkciju gubitka $E(g(\boldsymbol{\theta}) \mid \boldsymbol{X}_n)$.*

(b) *Uzmimo sada da je $L(\theta, d) = |d - g(\theta)|$, funkcija apsolutnog gubitka. Isto tako, pretpostavimo da postoji procjenitelj konačnog rizika. Neka je θ_M medijan uvjetne distribucije od $\boldsymbol{\theta}$ uz dano $\boldsymbol{X}_n = \boldsymbol{x}_n$. Za $c \geq \theta_M$ imamo:*

$$E(|g(\boldsymbol{\theta}) - c| \mid \boldsymbol{x}_n) - E(|g(\boldsymbol{\theta}) - \theta_M| \mid \boldsymbol{x}_n) = (c - \theta_M)\big(P(g(\boldsymbol{\theta}) \leq \theta_M \mid \boldsymbol{x}_n) - P(g(\boldsymbol{\theta}) > \theta_M \mid \boldsymbol{x}_n)\big)$$

$$+ 2E\big((c - g(\boldsymbol{\theta}))\mathbb{1}_{\{\theta_M < g(\boldsymbol{\theta}) < c\}} \mid \boldsymbol{x}_n\big).$$

Zbog svojstva medijana vrijedi $P(g(\boldsymbol{\theta}) \leq \theta_M \mid \boldsymbol{x}_n) \geq 1/2$ pa je gornji izraz veći ili jednak 0. Slično, za $c \leq \theta_M$ imamo:

$$E(|g(\boldsymbol{\theta}) - c| \mid \boldsymbol{x}_n) - E(|g(\boldsymbol{\theta}) - \theta_M| \mid \boldsymbol{x}_n) = (\theta_M - c)\big(P(g(\boldsymbol{\theta}) \geq \theta_M \mid \boldsymbol{x}_n) - P(g(\boldsymbol{\theta}) < \theta_M \mid \boldsymbol{x}_n)\big)$$

$$+ 2E\big((g(\boldsymbol{\theta}) - c)\mathbb{1}_{\{c < g(\boldsymbol{\theta}) < \theta_M\}} \mid \boldsymbol{x}_n\big).$$

Analogno zaključujemo da je i ovaj izraz veći ili jednak nula. Sada za svaki $c \in \Theta$ vrijedi:

$$E(|g(\boldsymbol{\theta}) - c| \mid \boldsymbol{x}_n) \geq E(|g(\boldsymbol{\theta}) - \theta_M| \mid \boldsymbol{x}_n).$$

Dakle, $\theta_M(\boldsymbol{X}_n)$ je Bayesov procjenitelj za funkciju apsolutnog gubitka.

Sada kada znamo kako izgledaju Bayesovi procjenitelji za pojedine funkcije gubitka, izračunajmo na primjerima Bayesove procjenitelje.

Primjer 2.0.5.
Neka su pretpostavke i oznake jednake kao u Primjeru 1.0.3. *a*). *Imamo* $\boldsymbol{\theta} \mid \boldsymbol{X}_n \sim Beta(\alpha + \sum_{i=1}^n X_i, \beta + n - \sum_{i=1}^n X_i)$. *Odredimo Bayesov procjenitelj u odnosu na funkciju kvadratnog gubitka. Po prethodnom primjeru znamo da je procjenitelj jednak* $E(\boldsymbol{\theta} \mid \boldsymbol{X}_n)$ $(r = \sum_{k=1}^n x_k)$*:*

$$E(\boldsymbol{\theta} \mid \boldsymbol{x}_n) = \int_0^1 \theta \, \frac{\theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1}}{B(\alpha+r, \beta+n-r)} \, d\theta$$

$$= \frac{B(\alpha+r+1, \beta+n-r)}{B(\alpha+r, \beta+n-r)} = \frac{\Gamma(\alpha+r+1)\Gamma(\beta+n-r)\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)\Gamma(\alpha+\beta+n+1)}$$

$$= \frac{\alpha+r}{\alpha+\beta+n}.$$

Dakle, $E(\boldsymbol{\theta} \mid \boldsymbol{X}_n) = \frac{\alpha}{\alpha+\beta+n} + \frac{\sum_{i=1}^n X_i}{\alpha+\beta+n}$.

Primjer 2.0.6.
U primjeru 1.0.3 *b*) *smo izračunali aposteriornu distribuciju u slučaju kada apriorna distribucija i uzorak prate normalnu distribuciju te smo dobili da je i aposteriorna distribucija normalna. Točnije, uz iste oznake kao u primjer* 1.0.3 *imamo*

$$\boldsymbol{\theta} \mid \boldsymbol{X}_n \sim N\Big(\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2} \big(\frac{n}{\sigma^2}\overline{X}_n + \frac{\mu}{\tau^2}\big), \, \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2} \Big).$$

Kako je riječ o normalnoj distribuciji, medijan je jedak očekivanju pa je Bayesov procjenitelj obzirom na obje funkcije gubitka jednak

$$\frac{\tau^2\sigma 2}{n\tau^2+\sigma^2}\big(\frac{n}{\sigma^2}\overline{X}_n + \frac{\mu}{\tau^2}\big).$$

Poglavlje 3

Konzistentnost aposteriorne distribucije

Za razliku od klasične statistike u kojoj je parametar od interesa fiksan i nepoznat, u Bayesovoj statistici nam je parametar slučajna veličina. Zanima nas možemo li nekako opravdati bayesovski pristup sa gledišta klasične statistike, odnosno u slučaju da je parametar stvarno fiksan, a ne slučajna veličina. Jedan način je preko pojma konzistentnosti aposteriorne distribucije koji ćemo sada definirati.

Definicija 3.0.1.
Neka je za svaki $n \in \mathbb{N}$, $\Pi(\cdot|\mathbf{X}_n)$ aposteriorna distribucija uz dane $X_1, X_2, \ldots, X_n$. Kažemo da je niz $(\Pi(\cdot|\mathbf{X}_n))_n$ aposteriornih distribucija konzistentan u θ_0 ako za svaku okolinu $U \ni \theta_0$ vrijedi:

$$\Pi(U|\mathbf{X}_n) \to 1 \;\; P_{\theta_0}^{\infty} - g.s.$$

Konzistentnost nam garantira da će se aposteriorna distribucija sa povećanjem uzorka koncentrirati oko prave vrijednosti parametra. Intuitivno, aposteriorna distribucija će učiti iz podataka i koncentrirati težinu oko pravog parametra θ_0.
Pogledajmo na primjeru kako izgleda konzistentnost niza aposteriornih distribucija.

Primjer 3.0.2.
Uzmimo $X_1, X_2, \ldots, X_n$ i apriornu distribuciju parametra kao iz primjera 1.0.3. b). Znamo da je aposteriorna distribucija $\Pi(\cdot|\boldsymbol{X}_n) \sim Beta(\alpha + \sum_{i=1}^n X_i, \beta + n - \sum_{i=1}^n X_i)$. Pokažimo da je niz aposteriornih distribucija konzistentan u θ_0, $\theta_0 \in (0,1)$. Uvjetno očekivanje je dano sa:

$$E(\boldsymbol{\theta} \mid \boldsymbol{X}_n) = \frac{\alpha}{\alpha + \beta + n} + \frac{n}{\alpha + \beta + n}\bar{\mathbf{X}}_n.$$

Po jakom zakonu velikih brojeva gornji izraz konvergira ka θ_0 $P_{\theta_0}^{\infty}$ g.s. Izračunajmo uvjetnu varijancu ($r := \sum_{i=1}^n x_i$):

$$E(\boldsymbol{\theta}^2 \mid \boldsymbol{x}_n) = \int_0^1 \theta^2 \, \frac{\theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1}}{B(\alpha+r,\beta+n-r)} \, d\theta$$

$$= \frac{B(\alpha+r+2,\beta+n-r)}{B(\alpha+r,\beta+n-r)} = \frac{\Gamma(\alpha+r+2)\Gamma(\beta+n-r)\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)\Gamma(\alpha+\beta+n+2)}$$

$$= \frac{(\alpha+r+1)(\alpha+r)}{(\alpha+\beta+n+1)(\alpha+\beta+n)}.$$

$$Var(\boldsymbol{\theta}|\boldsymbol{x}_n) = E(\boldsymbol{\theta}^2 \mid \boldsymbol{X}_n) - E(\boldsymbol{\theta} \mid \boldsymbol{X}_n)^2 = \frac{(\alpha + \sum_{i=1}^n X_i)(\beta + n - \sum_{i=1}^n X_i)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}.$$

Vidimo da uvjetna varijanca konvergira k 0 $P_{\theta_0}^\infty$-g.s. Neka je sada $\epsilon > 0$ i $U = (\theta_0 - \epsilon, \theta_0 + \epsilon)$ okolina od θ_0. Vrijedi sljedeća nejednakost:

$$\Pi\Big(|\boldsymbol{\theta} - \theta_0| \geq \epsilon \mid \mathbf{X}_n\Big) \leq \Pi\Big(|\boldsymbol{\theta} - E(\boldsymbol{\theta} \mid \mathbf{X}_n)| \geq \epsilon/2 \mid \mathbf{X}_n\Big) + \Pi\Big(|E(\boldsymbol{\theta} \mid \mathbf{X}_n) - \theta_0| \geq \epsilon/2 \mid \mathbf{X}_n\Big).$$

Sada iz kovergencije uvjetnog očekivanja i primjenom Čebiševljeve nejednakosti iz gornje nejednakosti slijedi tvrdnja o aposteriornoj konzistentnosti.

Kako smo konzistentnost definirali za aposteriornu distribuciju koja ovisi o apriornoj distribuciji, možemo se pitati kako će biti asimptotski odnos između dvije konzistentne aposteriorne distribucije dobivene iz različitih apriornih distribucija. Sljedeći rezultat nam govori da će uz određene uvjete za veliki uzorak naše zaključivanje biti neosjetljivo na izbor apriorne distribucije

Teorem 3.0.3.
Neka je θ_0 sadržan u interioru parametarskog skupa Θ. Označimo sa π_1, π_2 apriorne gustoće obzirom na mjeru μ i pretpostavimo da su pozitivne i neprekidne u θ_0. Ako su pripadne aposteriorne gustoće $\pi_1(\theta \mid \mathbf{X}_n)$ i $\pi_2(\theta \mid \mathbf{X}_n)$ konzistentne u θ_0 tada vrijedi

$$\lim_{n\to\infty} \int |\pi_1(\theta \mid \mathbf{X}_n) - \pi_2(\theta \mid \mathbf{X}_n)| \, d\mu(\theta) = 0 \quad P_{\theta_0}^\infty - \ g.s.$$

Dokaz.
Pokazat ćemo da vrijedi sljedeće:

$$\lim_{n\to\infty} \int_\Theta \pi_2(\theta \mid \mathbf{X}_n) \left| 1 - \frac{\pi_1(\theta \mid \mathbf{X}_n)}{\pi_2(\theta \mid \mathbf{X}_n)} \right| d\mu(\theta) = 0 \quad P_{\theta_0}^\infty - g.s.$$

Neka su $\delta > 0$, $\eta > 0$, $\epsilon > 0$. Po pretpostavci su π_1 i π_2 neprekidne u θ_0 pa postoji otvorena okolina U od θ_0 tako da za svaki θ sadržan u U vrijedi:

$$\left|\frac{\pi_1(\theta)}{\pi_2(\theta)} - \frac{\pi_1(\theta_0)}{\pi_2(\theta_0)}\right| < \delta \quad i \quad |\pi_i(\theta) - \pi_i(\theta_0)| < \delta \quad i = 1, 2. \tag{3.1}$$

Kako su $\pi_1(\theta \mid \mathbf{X}_n)$, $\pi_2(\theta \mid \mathbf{X}_n)$ konzistente u θ_0, imamo

$$\Pi_i(U \mid \mathbf{X}_n) \to 1 \;\; P_{\theta_0}^{\infty} - g.s \qquad i = 1, 2.$$

Označimo sa $\mathbf{x}_n$ realizaciju slučajnog uzorka $\mathbf{X}_n$. Zbog konzistentnosti postoji n_0 takav da za svaki $n \geq n_0$ vrijedi:

$$\Pi_i(U \mid \mathbf{x}_n) \geq 1 - \eta \quad i = 1, 2. \tag{3.2}$$

Prisjetimo se, formula za aposteriornu gustoću je $\pi(\theta \mid \mathbf{x}_n) = \frac{f(\mathbf{x}_n|\theta)\,\pi(\theta)}{\int_\Theta f(\mathbf{x}_n|\xi)\,\pi(\xi)\;d\mu(\xi)}$ pa je:

$$\frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)} = \frac{\pi_1(\theta)}{\pi_2(\theta)} \frac{\int_\Theta f(\mathbf{x}_n|\xi)\,\pi_2(\xi)\;d\mu(\xi)}{\int_\Theta f(\mathbf{x}_n|\xi)\,\pi_1(\xi)\;d\mu(\xi)}. \tag{3.3}$$

Korištenjem nejednakosti (3.1) i (3.2) za $n \geq n_0$ i $\theta \in U$ slijedi:

$$\begin{aligned}
&\left(\frac{\pi_1(\theta_0)}{\pi_2(\theta_0)} - \delta\right)(1-\eta)\frac{\int_U f(\mathbf{x}_n|\xi)\,\pi_2(\xi)\;d\mu(\xi)}{\int_U f(\mathbf{x}_n|\xi)\,\pi_1(\xi)\;d\mu(\xi)} \\
&< \frac{\pi_1(\theta)}{\pi_2(\theta)}(1-\eta) \cdot \frac{\int_U f(\mathbf{x}_n|\xi)\,\pi_2(\xi)\;d\mu(\xi)}{\int_U f(\mathbf{x}_n|\xi)\,\pi_1(\xi)\;d\mu(\xi)} \\
&= \frac{\pi_1(\theta)}{\pi_2(\theta)}(1-\eta) \cdot \frac{\Pi_2(U|\mathbf{x}_n)\int_\Theta f(\mathbf{x}_n|\xi)\,\pi_2(\xi)\;d\mu(\xi)}{\Pi_1(U|\mathbf{x}_n)\int_\Theta f(\mathbf{x}_n|\xi)\,\pi_1(\xi)\;d\mu(\xi)} \\
&= \frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_1(\theta|\mathbf{x}_n)}(1-\eta)\frac{\Pi_2(U|\mathbf{x}_n)}{\Pi_1(U|\mathbf{x}_n)} \\
&\leq \frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)}.
\end{aligned} \tag{3.4}$$

Prva nejednakost slijedi iz (3.1), a sljedeće dvije jednakosti primjenom (3.3). Nejednakost (3.4) je posljedica relacije (3.2). Koristeći slične argumente prethodno opisanima dobivamo gornju ogradu za $\frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)}$:

$$\frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)} \leq \left(\frac{\pi_1(\theta_0)}{\pi_2(\theta_0)} + \delta\right)(1-\eta)^{-1} \frac{\int_U f(\mathbf{x}_n|\xi)\, \pi_2(\xi)\ d\mu(\xi)}{\int_U f(\mathbf{x}_n|\xi)\, \pi_1(\xi)\ d\mu(\xi)}. \tag{3.5}$$

Iz neprekidnost apriornih gustoća zaključujemo:

$$\begin{aligned}(\pi_i(\theta_0) - \delta) \int_U f(\mathbf{x}_n|\xi)\ d\mu(\xi) \leq & \int_U f(\mathbf{x}_n|\xi) \pi_i(\xi)\ d\mu(\xi) \\ \leq & (\pi_i(\theta_0) + \delta) \int_U f(\mathbf{x}_n|\xi)\ d\mu(\xi) \quad i = 1, 2. \end{aligned} \tag{3.6}$$

Koristeći (3.6) zaključujemo

$$\left(\frac{\pi_2(\theta_0) - \delta}{\pi_1(\theta_0) + \delta}\right) \leq \frac{\int_U f(\mathbf{x}_n|\xi)\, \pi_2(\xi)\ d\mu(\xi)}{\int_U f(\mathbf{x}_n|\xi)\, \pi_1(\xi)\ d\mu(\xi)} \leq \left(\frac{\pi_2(\theta_0) + \delta}{\pi_1(\theta_0) - \delta}\right).$$

Primjenom dobivenog rezultata na nejednakosti (3.4) i (3.5) vezane za $\frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)}$ za $\theta \in U$ imamo:

$$\begin{aligned}\left(\frac{\pi_1(\theta_0)}{\pi_2(\theta_0)} + \delta\right)(1-\eta)\left(\frac{\pi_2(\theta_0) - \delta}{\pi_1(\theta_0) + \delta}\right) \leq & \frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)} \\ \leq & \left(\frac{\pi_1(\theta_0)}{\pi_2(\theta_0)} + \delta\right)(1-\eta)^{-1}\left(\frac{\pi_2(\theta_0) + \delta}{\pi_1(\theta_0) - \delta}\right). \end{aligned} \tag{3.7}$$

Primijetimo da za $\delta \to 0$ i $\eta \to 0$ lijeva i desna strana od (3.7) teže ka 1. Dakle, za zadani $\epsilon > 0$ i dovoljno male δ i η vrijedi:

$$\left|\frac{\pi_1(\theta|\mathbf{x}_n)}{\pi_2(\theta|\mathbf{x}_n)} - 1\right| < \epsilon. \tag{3.8}$$

Konačno, za $n \geq n_0$ vrijedi:

$$\int |\pi_1(\theta \mid \mathbf{X}_n) - \pi_2(\theta \mid \mathbf{X}_n)| \; d\mu(\theta)$$

$$= \int_U \pi_2(\theta \mid \mathbf{X}_n) \left| 1 - \frac{\pi_1(\theta \mid \mathbf{X}_n)}{\pi_2(\theta \mid \mathbf{X}_n)} \right| d\mu(\theta) + \int_{U^c} |\pi_1(\theta \mid \mathbf{X}_n) - \pi_2(\theta \mid \mathbf{X}_n)| \; d\mu(\theta)$$

$$\leq \epsilon \int_U \pi_2(\theta \mid \mathbf{X}_n) \, d\mu(\theta) + \int_{U^c} \pi_1(\theta \mid \mathbf{X}_n) \; d\mu(\theta) + \int_{U^c} \pi_2(\theta \mid \mathbf{X}_n) \; d\mu(\theta)$$

$$\leq \epsilon + 2\eta.$$

Prva nejednakost slijedi (3.8), a posljednja nejednakost slijedi iz (3.2). □

Jedan od najvažnijih rezultata vezanih za konzistentnost je Doobov teorem. Teorem nam govori da će uz određene uvjete, za proizvoljan odabir apriorne distribucije, aposteriorna distribucija biti konzistenta na skupu $\Theta_0 \subset \Theta$ čija je vjerojatnost $\Pi(\Theta_0) = 1$. Prije iskaza i dokaza Doobovog teorema iskažimo rezultate koji će biti korišteni u dokazu. Dokazi sljedeća dva teorema mogu se naći u [1] i [8].

Teorem 3.0.4.
Pretpostavimo da je $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Definirajmo $X_n := E[X|\mathcal{F}_n]$, $n \in \mathbb{N}$, gdje je $\{\mathcal{F}_n, n \in \mathbb{N}\}$ filtracija i stavimo $\mathcal{F}_\infty = \sigma(\bigcup_{n=0}^{\infty} \mathcal{F}_n)$. Tada je

$$\lim_{n\to\infty} X_n = E(X \mid \mathcal{F}_\infty) \; g.s. \; i \;\; u \; \mathcal{L}^1(\Omega, \mathcal{F}, P).$$

Teorem 3.0.5.
Neka su $\boldsymbol{U}$ i $\boldsymbol{V}$ potpuni, separabilni metrički prostori i $E_1 \subset \boldsymbol{U}$, $E_2 \subset \boldsymbol{V}$, gdje je E_1 Borelov skup. Ako je $\phi : \boldsymbol{U} \to \boldsymbol{V}$ izmjerivo preslikavanje takvo da je $\phi|_{E_1}$ bijekcija i $\phi(E_1) = E_2$, tada je E_2 Borelov skup i $\phi|_{E_1}^{-1} : E_2 \to \boldsymbol{U}$ izmjeriva funkcija.

Teorem 3.0.6. *(Doob)*
Neka su Θ i $\boldsymbol{X}$ potpuni,separabilni metrički prostori te $\mathcal{B}(\Theta)$ i $\mathcal{A}$ pripadne Borelove σ-algebre. Zatim, neka je Π apriorna i $(\Pi(\cdot|\mathbf{X}_n))_n$ niz aposteriornih distribucija. Tada postoji $\Theta_0 \subset \Theta$, $\Pi(\Theta_0) = 1$ takav da je $(\Pi(\cdot|\mathbf{X}_n))_n$ konzistentan u θ za svaki $\theta \in \Theta_0$.

Dokaz.
Neka je λ_Π vjerojatnosna mjera $\mathcal{B}(\Theta) \times \mathcal{A}^\infty$ definirana u poglavlju 1. Za proizvoljni $C \subset \Theta$ iz Teorema 3.0.4 za $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$, $X_n = E(\mathbb{1}_C \mid \mathcal{F}_n)$ slijedi

$$\Pi(C \mid X_1, X_2, \ldots, X_n) \xrightarrow{n\to\infty} E(\mathbb{1}_C \mid X_1, X_2, \ldots) =: f \; \lambda_\Pi \; g.s.$$

Napomenimo samo da, obzirom da je riječ o produktnoj vjerojatnosnoj mjeri, funkcije $\mathbb{1}_C$ i f promatramo kao funkcije dvije varijable (θ, x). Dakle, $\mathbb{1}_C$ je zapravo $\mathbb{1}_{C\times\mathbf{X}^\infty}$, a $f(x)$ je $f(\theta, x)$. Neka je $\mathcal{A}_0$ prebrojiva algebra koja generira $\mathcal{A}$. Definirajmo skup

$$E := \left\{(\theta, x) \in \Theta \times \mathbf{X}^\infty : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(x_i) = P_\theta(A), \forall A \in \mathcal{A}_0\right\}.$$

Za fiksan $A \in \mathcal{A}_0$ je skup $E^A = \{(\theta, x) \in \Theta \times \mathbf{X}^\infty : \lim\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(x_i) = P_\theta(A)\}$ izmjeriv kao praslika razlike dviju izmjerivih funkcija pa je E izmjeriv skup kao presjek prebrojivo mnogo izmjerivih skupova. Za svaki $\theta \in \Theta$ označimo sa E_θ prerez skupa E po θ. Nadalje, vrijede sljedeće dvije tvrdnje:

(i) $\forall\theta, P_\theta^\infty(E_\theta) = 1$,

(ii) $\theta \neq \theta' \Rightarrow E_\theta \cap E_{\theta'} = \emptyset$.

Prva tvrdnja slijedi iz činjenice da je za fiksan A po jakom zakonu velikih brojeva $P_\theta^\infty(E_\theta^A) = 1$, pa je $P_\theta^\infty(E_\theta) = P_\theta^\infty(\bigcap\limits_{A\in\mathcal{A}_0} E_\theta^A) = 1$. Ako je $\theta \neq \theta'$ tada je po pretpostavci $P_\theta \neq P_{\theta'}$ pa postoji $B \in \mathcal{A}_0$ takav da $P_\theta(B) \neq P_{\theta'}(B)$ (u suprotnom bi se podudarale na π-sustavu $\mathcal{A}_0$ koji generira $\mathcal{A}$ pa time i na cijeloj $\mathcal{A}$). Dakle, (ii) vrijedi. Promotrimo sljedeću funkciju:

$$f^*(x) = \begin{cases} 1, & x \in \cup_{\theta\in C} E_\theta \\ 0, & x \notin \cup_{\theta\in C} E_\theta. \end{cases}$$

Pokažimo da je f^* izmjeriva funkcija. Za to je dovoljno pokazati da je $\cup_{\theta\in C} E_\theta$ izmjeriv skup. Definirajmo preslikavanje $\phi : E \to \mathbf{X}$ sa $\phi(\theta, x) = x$. Primjenom teorema 3.0.5 na funkciju ϕ slijedi da je $(\phi^{-1})^{-1}(C\times\mathbf{X}^\infty \cap E) = \cup_{\theta\in C} E_\theta$ izmjeriv skup. Dakle, f^* je izmjeriva funkcija. Korištenjem svojstava (i) i (ii) zaključujemo :

1. $\theta \in C \Rightarrow f^* = 1 \;\; P_\theta^\infty - g.s.$
2. $\theta \notin C \Rightarrow f^* = 0 \;\; P_\theta^\infty - g.s.$

Stoga je za svaki θ $f^* = \mathbb{1}_C(\theta) f^* \;\; P_\theta^\infty - g.s.$ Pokažimo sada da je f^* verzija od $E(\mathbb{1}_C \mid X_1, X_2, \ldots)$. Za proizvoljan $B \in \mathcal{A}^\infty$ imamo:

$$E(f^* \mathbb{1}_B) = \int \mathbb{1}_B f^* d\lambda_\pi = \int_\Theta \mathbb{1}_B(x)\mathbb{1}_C(\theta) f^*(x) dP_\theta^\infty(x) d\Pi(\theta)$$

$$= \int_\Theta \mathbb{1}_C(\theta) P_\theta^\infty(B) d\Pi(\theta)$$

$$= \lambda_\Pi(C \times B) = E(\mathbb{1}_{C\times \mathbf{X}^\infty}\, \mathbb{1}_{\Theta\times B}).$$

Kako je f verzija od $E(\mathbb{1}_C \mid X_1, X_2, \ldots)$, slijedi da je $f = f^*\ \lambda_\Pi - g.s.$, odnosno ako je $N := \{f \neq f^*\}$, tada je

$$\lambda_\Pi(N) = \int_\Theta P_\theta^\infty(N_\theta)\, d\Pi(\theta) = 0$$

pa slijedi da je $P_\theta^\infty(N_\theta) = 0\ \ \Pi - g.s.$ Dakle, postoji Θ_0 , $\Pi(\Theta_0) = 1$ takav da je $f = f^*\ \ P_\theta^\infty - g.s.$ za svaki $\theta \in \Theta_0$. Time smo pokazali da za svaki $\theta \in C \cap \Theta_0$ vrijedi

$$\Pi(C \mid X_1, X_2, \ldots, X_n) \xrightarrow{n\to\infty} 1\ \ P_\theta^\infty - g.s.$$

Označimo sada sa $\mathcal{U} = \{U_1, U_2, \ldots\}$ prebrojivu bazu topološkog prostora Θ. Po prethodnom dokazanom za svaki $i \in \mathbb{N}$ i $C = U_i \in \mathcal{U}$ postoji Θ_{0_i} takav da je $\Pi(\Theta_{0_i}) = 1$ i za svaki $\theta \in \Theta_{0_i} \cap U_i$ vrijedi $f = 1\ P_\theta^\infty$ g.s.
Sada za $\Theta_0 := \bigcap_i \Theta_{0_i}$ vrijedi $\Pi(\Theta_0) = 1$. Za proizvoljan $\theta_1 \in \Theta_0$ i proizvoljnu okolinu $B \ni \theta_1$ postoji $U_i \in \mathcal{U}$ takav da je $\theta_1 \in U_i \subset B$. Kako je $\mathbb{1}_B \geq \mathbb{1}_{U_i}$ i $\theta_1 \in \Theta_0 \subset \Theta_{0_i}$ slijedi

$$E(\mathbb{1}_B \mid X_1, X_2, \ldots, X_n) \geq E(\mathbb{1}_{U_i} \mid X_1, X_2, \ldots, X_n) \to 1\ \ P_{\theta_1}^\infty - g.s.$$

Dakle,

$$\Pi(B \mid X_1, X_2, \ldots, X_n) \to 1\ \ P_{\theta_1}^\infty - g.s.$$

□

U poglavlju 2. bavili smo se Bayesovim procjeniteljima, odnosno pitanjem kako procjeniti slučajnu varijablu, parametar $\boldsymbol{\theta}$, na temelju zadane apriorne distribucije i funkcije gubitka. U klasičnoj statistici se isto postavlja pitanje kako procjeniti parametar, koji je u ovom slučaju fiksan i nepoznat, na temelju realizacije slučajnog uzorka. Jedan pristup procjenjivanja parametra je metoda maksimalne vjerodostojnosti, koju ćemo sada opisati. Prvo definirajmo funkciju parametra θ koju ćemo zvati funkcijom vjerodostojnosti.

Definicija 3.0.7.
Neka je $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$ realizacija slučajnog uzorka $(X_1, X_2, \ldots, X_n)$. Funkcija vjerodostojnosti je funkcija $L_n : \Theta \to \mathbb{R}$ definirana sa:

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

Obrazložimo malo zašto nam je od interesa ovakva funkcija. Pretpostavimo da je $X_1, X_2, \ldots, X_n$ slučajan uzorak koji prati neku diskretnu razdiobu. Tada funkciju vjerodostojnosti za danu realizaciju slučajnog uzorka možemo zapisati na sljedeći način:

$$L_n(\theta) = \prod_{k=1}^{n} f(x_k|\theta) = \prod_{k=1}^{n} P_\theta(X_k = x_k) = P_\theta(\mathbf{X}_n = \mathbf{x}_n).$$

Dakle, tom slučaju nam je funkcija vjerodostojnosti zapravo vjerojatnost da je realizacija slučajnog uzorka $\mathbf{X}_n$ jednaka $\mathbf{x}_n$ uz pretpostavku da je θ prava vrijednost parametra. To možemo protumačiti kao mjeru koliko je "vjerodostojno" da je upravo θ prava vrijednost parametra koji određuje distribuciju slučajnog uzorka. Stoga je prirodno parametar procjenjivati sa procjeniteljem $\hat{\theta}_n$ koji će maksimizirati funkciju vjerodostojnosti. Takav procjenitelj ćemo zvati procjeniteljem maksimalne vjerodostojnosti.

Definicija 3.0.8.
Izmjerivu funkciju $\hat{\theta}_n(X_1, X_2, \ldots, X_n) =: \hat{\theta}_n$ nazivamo procjeniteljem maksimalne vjerodostojnosti ako vrijedi:

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

Napomenimo da takva funkcija općenito ne mora postojati. Jedan od primjera kada postoji je u slučaju kada je Θ kompaktan skup i kada je $\theta \mapsto f(x|\theta))$ neprekidna za svaki x. Ilustrirajmo procjenu parametra metodom maksimalne vjerodostojnosti na jednom primjeru.

Primjer 3.0.9.
Neka je $X_1, X_2, \ldots, X_n \sim U(0, \theta)$ niz nezavisnih jednako distribuiranih slučajnih varijabli i neka je $\Theta = (0, 1)$. Definirajmo $X_{(n)} := \max(X_1, X_2, \ldots, X_n)$, $X_{(1)} =: \min(X_1, X_2, \ldots, X_n)$. Funkcija vjerodostojnosti za realizaciju slučajnog uzorka $(x_1, x_2, \ldots, x_n)$ je dana sa:

$$L_n(\theta) = \prod_{k=1}^{n} f(x_k, |\theta) = \frac{1}{\theta^n}. \quad 0 \leq x_{(1)} \leq x_{(n)} \leq \theta.$$

Računanje maksimuma funkcije $L_n(\theta)$ je ekvivalentno računanju maksimuma log funkcije vjerodostojnosti $l_n(\theta) := \log L_n(\theta)$.

$$l_n(\theta) = \log(\theta^{-n}) = -n \log(\theta).$$

Budući da je $\dot{l}_n(\theta) = -n\frac{1}{\theta} < 0$, funkcija $l_n(\theta)$ je padajuća za $\theta \geq x_{(n)}$ pa zaključujemo da poprima maksimum u $x_{(n)}$. Dakle, procjenitelj maksimalne vjerodostojnosti za θ je $\hat{\theta}_n = X_{(n)}$.

Sada kada imamo procjenu parametra, zanima nas kako će se ta procjena ponašati sa povećanjem veličine uzorka. U klasičnoj statistici također imamo pojam konzistentnosti koji nam govori o asimptotskom ponašanju procjenitelja. Naime, htjeli bi da nam s povećanjem veličine uzorka naš procjenitelj na neki način bude što bliži pravoj vrijednosti parametra. Nama će od interesa biti jaka konzistentnost.

Definicija 3.0.10.
Niz procjenitelja $(T_n)_n$ je jako konzistentan za parametar θ ako vrijedi:

$$T_n \xrightarrow{n\to\infty} \theta \quad P_\theta^\infty - g.s.$$

U nastavku ćemo opisati pod kojim uvjetima nam ja istodobno garantirana jaka konzistentnost procjenitelja maksimalne vjerodostojnosti i konzistentnost aposteriorne distribucije. Za početak iskažimo teorem o uniformnom jakom zakonu velikih brojeva koji ćemo koristiti u dokazivanju daljnih teorema.

Teorem 3.0.11.
Neka je $x \mapsto U(x,\theta)$ izmjeriva funkcija za svaki $\theta \in \Theta$. Fiksirajmo $\theta_0 \in \Theta$ i stavimo $\tau(\theta) := E_{\theta_0}(U(X,\theta))$. Pretpostavimo da vrijedi sljedeće:

(i) Θ je kompaktan skup,

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

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

Tada je :

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{k=1}^{n} U(x_k,\theta) - \tau(\theta) \right| = 0 \quad P_{\theta_0}^\infty - g.s.$$

Definirajmo sada Kullback-Leibler informaciju i rezultat vezan za taj pojam koji će nam uz prethodni teorem biti ključan za dokazivanje konzistentnosti procjenitelja.

Definicija 3.0.12.
Neka su f_0, f_1 gustoće obzirom na σ-konačnu mjeru μ. Broj $K(f_0,f_1) := E_0 \log(\frac{f_0(X)}{f_1(X)}) = \int \log \frac{f_0(x)}{f_1(x)} f_0(x)\, d\mu(x)$ zovemo Kullback-Leibler informacijom od f_0 obzirom na f_1.

Dokaz prethodnog teorema i naredne leme može se naći u [4].

Lema 3.0.13.
Neka su f_0, f_1 gustoće obzirom na σ-konačnu mjeru μ. Tada je

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

Jednakost vrijedi ako i samo ako je $f_0 = f_1$ $\mu - g.s.$

.

Definicija 3.0.14.
Neka je $(S, \mathcal{F}, P)$ vjerojatnosni prostor. Najmanji zatvoreni skup $B \in \mathcal{F}$ takav da je $P(B) = 1$ nazivamo nosačem vjerojatnosne mjere P i označavamo sa $\mathrm{supp}(P)$.

Sada iskažimo i dokažimo rezultat koji nam govori uz koje uvjete ćemo imati jaku konzistentnost niza procjenitelja maksimalne vjerodostojnosti i konzistentnost apriorne distribucije.

Teorem 3.0.15.
Neka je Θ kompaktan metrički prostor, $\theta_0 \in \Theta$ zadan i

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

Ako $U(x, \theta)$ zadovoljava uvjete teorema 3.0.11, tada vrijedi sljedeće:

(i) svaki niz procjenitelja maksimalne vjerodostojnosti $(\hat{\theta}_n)_n$ je jako konzistentan za θ_0,

(ii) ako je Π apriorna distribucija na Θ i ako je $\theta_0 \in$ $\mathrm{supp}(\Theta)$, tada je aposteriorna distribucija konzistentna u θ_0.

Dokaz.

(i) Neka je $x \in \mathbf{X}^\infty$. Pokažimo da za proizvoljnu otvorenu okolinu S od θ_0 postoji $n(x) \in \mathbb{N}$ takav da je za svaki $n \geq n(x)$, $\hat{\theta}_n \in S$. Kako je $\theta \mapsto U(z, \theta)$ neprekidna za svaki $z \in \mathbf{X}$ i dominirana nekom integrabilnom funkcijom zbog uvjeta (iii) teorema 3.0.11, primjenom teorema o dominiranoj konvergenciji lako se pokaže da je $\tau(\theta) = E_{\theta_0}(U(X, \theta))$ neprekidna funkcija. Iz leme 3.0.13 slijedi da je $\tau(\theta) = -K(\theta_0, \theta) < 0$ za sve $\theta \neq \theta_0$. S^c je zatvoren podskup kompaktnog skupa Θ pa je kompaktan te $\tau(\theta)$ postiže maksimum na S^c, $\sup_{\theta \in S^c} \tau(\theta) < 0$. Po teoremu 3.0.11 za $0 < \epsilon < |\sup_{\theta \in S^c} \tau(\theta)|$ postoji $n(x) \in \mathbb{N}$ tako da je za $n \geq n(x)$

$$\sup_{\theta \in S^c} \left| \frac{1}{n} \sum_{k=1}^{n} U(x_k, \theta) - \tau(\theta) \right| < \epsilon.$$

Kako je ϵ proizvoljno malen, za $\theta \in S^c$ je $\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta) \leq 0$. Štoviše, zbog uvjeta na ϵ, $\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta)$ ne može biti nula za neki $\theta \in S^c$. Naime, kada bi postojao $\theta_1 \in S^c$ takav da je $\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta_1) = 0$ imali bi

$$\inf_{\theta\in S^c} |\tau(\theta)| \leq |\tau(\theta_1)| \leq \sup_{\theta\in S^c}\left|\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta) - \tau(\theta)\right| < \epsilon.$$

Zbog negativnosti funkcije $\tau(\theta)$ vrijedi sljedeće:

$$|\sup_{\theta\in S^c}\tau(\theta)| = |-\inf_{\theta\in S^c}(-\tau(\theta))| = |-\inf_{\theta\in S^c}|\tau(\theta)|\,| = \inf_{\theta\in S^c}|\tau(\theta)|\,.$$

Odnosno, vrijedi

$$|\sup_{\theta\in S^c}\tau(\theta)| = \inf_{\theta\in S^c}|\tau(\theta)| < \epsilon$$

što je u kontradikciji sa izborom ϵ. S druge strane, $(\hat{\theta}_n)_n$ je niz procjenitelja maksimalne vjerodostojnost pa je $\frac{1}{n}\sum_{k=1}^{n} U(x_k,\hat{\theta}_n) \geq 0$. Dakle, $\hat{\theta}_n \in S$ za svaki $n \geq n(x)$.

(ii) Neka je U okolina od θ_0 i $\delta > 0$ proizvoljan. Označimo sa U_δ kuglu radijusa δ sa središtem u θ_0. Uvedimo sljedeće oznake:

$$A_1 := \inf_{\theta\in\bar{U}_\delta} \tau(\theta) < 0 \quad , \quad A_2 := \sup_{\theta\in U^c} \tau(\theta) < 0.$$

Kako je $\tau(\theta) \leq 0$ te poprima vrijednost 0 u θ_0, zbog neprekidnosti vrijedi:

$$\delta \searrow 0 \;\Rightarrow\; \inf_{\theta\in\bar{U}_\delta} \tau(\theta) \nearrow 0.$$

Dakle, možemo izabrati dovoljno malen $\delta > 0$ tako da vrijedi $U_\delta \subset U$ i $|A_1| < |A_2|$. Uzmimo sada $\epsilon > 0$ tako da je $A_1 - \epsilon > A_2 + \epsilon$. Primjenom teorema 3.0.11 na skup $U^c \cup \bar{U}_\delta$ slijedi da postoji $n(x)$ tako da za $n \geq n(x)$ vrijedi:

$$\left|\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta) - \tau(\theta)\right| < \epsilon \quad \forall\theta \in U^c \cup \bar{U}_\delta.$$

Iz prethodne nejednakosti zaključujemo da vrijedi:

$$\tau(\theta) - \epsilon < \frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta) < \epsilon + \tau(\theta)$$

pa vrijede sljedeće nejednakosti:

$$A_1 - \epsilon < \frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta), \quad \forall\theta \in \bar{U}_\delta \tag{3.9}$$

$$\frac{1}{n}\sum_{k=1}^{n} U(x_k,\theta) < \epsilon + A_2, \quad \forall\theta \in U^c. \tag{3.10}$$

Primjenom (3.9) i (3.10) te zbog činjenice da je $\bar{U}_\delta \subset U$ imamo

$$\Pi(U \mid x_1, x_2, \dots, x_n) = \frac{\int_U f(\mathbf{x}_n|\theta)\ d\Pi(\theta)}{\int_\Theta f(\mathbf{x}_n|\theta)\ d\Pi(\theta)} \left(\cdot \frac{e^{-\log(f(\mathbf{x}_n|\theta_0))}}{e^{-\log(f(\mathbf{x}_n|\theta_0))}} \right)$$

$$= \frac{\int_U e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta)}{\int_U e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta) + \int_{U^c} e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta)}$$

$$\geq \frac{\int_{\bar{U}_\delta} e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta)}{\int_{\bar{U}_\delta} e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta) + \int_{U^c} e^{\sum_{k=1}^n U(x_k,\theta)} d\Pi(\theta)}$$

$$\geq \frac{\Pi(\bar{U}_\delta)\ e^{n(A_1-\epsilon)}}{\Pi(\bar{U}_\delta) e^{n(A_1-\epsilon)} + \Pi(U^c) e^{n(A_2+\epsilon)}}$$

$$= \left(1 + \frac{\Pi(U^c) e^{n(A_2+\epsilon)}}{\Pi(\bar{U}_\delta)\ e^{n(A_1-\epsilon)}} \right)^{-1}.$$

Budući da je $A_2 - A_1 + 2\epsilon < 0$, desna strana teži ka 1 za $n \to \infty$ te je stoga

$$\lim_{n\to\infty} \Pi(U \mid x_1, x_2, \dots, x_n) = 1.$$

Time smo dokazali konzistentost aposteriorne distribucije u parametru θ_0.

□

Pokazali smo da uz određene pretpostavke imamo istovremeno jaku konzistentnost procjenitelja maksimalne vjerodostojnosti i konzistentnost aposteriorne distribucije na cijelom nosaču od Π. No, to ne mora uvijek vrijediti. Slijedi primjer u kojemo aposteriorna distribucija nije konzistentna na cijelom nosaču, ali je procjenitelj maksimalne vjerodostojnosti jako konzistentnan.

Primjer 3.0.16.

Stavimo $\Theta = (0, 1] \cup (2, 3)$ *i neka je* $X_1, X_2, \dots X_n \overset{njd}{\sim} U(0, \theta)$. *Dakle, imamo slučajan uzorak koji prati uniformnu distribuciju. Pretpostavimo da je stvarna vrijednost parametra* $\theta_0 = 1$. *Apriorna gustoća* π *zadana je na* $(0, 1)$ *sa* $\pi(\theta) = e^{-1/(\theta-\theta_0)}$. *Budući da je* $\int_0^1 \pi(\theta)\, d(\theta) \approx 0.089 < 1$, *postoji apriorna gustoća koja je tako zadana na* $(0, 1)$ *i pozitivna na* $\Theta \setminus (0, 1)$.

Kako bi pokazali da aposteriorna distribucija nije konzistentna u $\theta_0 = 1$ *dovoljno je pokazati:*

$$\Pi((2,3)) \mid \mathbf{X}_n) \xrightarrow{P_{\theta_0}^{\infty}} 1. \tag{3.11}$$

Objasnimo zašto je to dovoljno. Za svaku okolinu U *koja sadrži* $\theta_0 = 1$ *i koja je sadržana u komplementu od* $(2,3)$ *vrijedi:*

$$\Pi(U \mid \mathbf{X}_n) \xrightarrow{P_{\theta_0}^{\infty}} 0,$$

pa po definicije konzistentnosti zaključujemo da aposteriorna distribucija nije konzistentna u θ_0*. Dokažimo sada* (3.11)*. Stavimo* $X_{(n)} := \max(X_1, X_2, \ldots, X_n)$*. Slijedi:*

$$\Pi((2,3)) \mid \mathbf{X}_n) = \frac{\int_2^3 \theta^{-n} \prod_{k=1}^n \mathbb{1}_{(0,\theta)}(X_k)\pi(\theta)\, d\theta}{\int_\Theta \theta^{-n} \prod_{k=1}^n \mathbb{1}_{(0,\theta)}(X_k)\pi(\theta)\, d\theta}$$

$$= \frac{\int_2^3 \theta^{-n} \mathbb{1}_{(0,\theta)}(X_{(n)})\pi(\theta)\, d\theta}{\int_0^1 \theta^{-n} \mathbb{1}_{(0,\theta)}(X_{(n)})\pi(\theta)\, d\theta + \int_2^3 \theta^{-n} \mathbb{1}_{(0,\theta)}(X_{(n)})\pi(\theta)\, d\theta}.$$

Znamo da je $0 \le X_{(n)} \le 1$ $P_{\theta_0}^{\infty} - g.s.$ *pa je brojnik u gornjem izrazu jednak* $\int_2^3 \theta^{-n}\pi(\theta)\, d\theta$*. Nadalje, ako je* $\theta < X_{(n)}$ *onda je* $\mathbb{1}_{(0,\theta)}(X_{(n)})$ *jednako nula pa gornji izraz možemo zapisati na sljedeći način:*

$$\frac{\int_2^3 \theta^{-n}\pi(\theta)\, d\theta}{\int_{X_{(n)}}^1 \theta^{-n}\pi(\theta)\, d\theta + \int_2^3 \theta^{-n}\pi(\theta)\, d\theta} = \frac{1}{1 + \frac{\int_{X_{(n)}}^1 \theta^{-n}\pi(\theta)\, d\theta}{\int_2^3 \theta^{-n}\pi(\theta)\, d\theta}}.$$

Radi jednostavnijeg zapisa stavimo $I_n^1 := \int_{X_{(n)}}^1 \theta^{-n}\pi(\theta)\, d\theta$ *i* $I_n^2 := \int_2^3 \theta^{-n}\pi(\theta)\, d\theta$*. Vrijedi sljedeće:*

$$I_n^1 \le \pi(X_{(n)}) \int_{X_{(n)}}^1 \theta^{-n}\, d\theta = \pi(X_{(n)}) \frac{1 - X_{(n)}^{-n+1}}{-n+1} = \pi(X_{(n)}) \frac{1}{n-1} \frac{1 - X_{(n)}^{n-1}}{X_{(n)}^{n-1}}.$$

Prva nejednakost slijedi iz činjenice da je $\pi(\theta)$ *padajuća na* $(0,1)$*. Logaritmirajmo gornju nejednakost i pomnožimo sa* $\frac{1}{n}$*:*

$$\frac{1}{n}\log(I_n^1) \le -\frac{\log(n-1)}{n} - \frac{n-1}{n}\log X_{(n)} + \frac{1}{n}\log(1 - X_{(n)}^{n-1}) + \frac{1}{n}\log \pi(X_{(n)}). \tag{3.12}$$

Prvi izraz sa desne strane teži u nulu. Za $\epsilon > 0$ *vrijedi:*

$$P^{\infty}_{\theta_0}(|X_{(n)} - 1| \geq \epsilon) = P^{\infty}_{\theta_0}(X_{(n)} \geq 1+\epsilon) + P^{\infty}_{\theta_0}(X_{(n)} \leq 1-\epsilon) = \begin{cases} \frac{(\theta-\epsilon)^n}{\theta^n}, & \epsilon < \theta \\ 0, & \epsilon \geq \theta. \end{cases}$$

Kako je za $\epsilon < \theta$, $\frac{(\theta-\epsilon)^n}{\theta} < 1$ *imamo* $\lim_{n\to\infty} P^{\infty}_{\theta_0}(|X_{(n)} - 1| \geq \epsilon) = 0$. $X_{(n)} \xrightarrow{P^{\infty}_{\theta_0}} 1$ *pa drugi izraz u* (3.12) *teži ka* $\log(1) = 0$ *po vjerojatnosti. Pokažimo da vrijedi:*

$$\frac{1}{n}\log(1 - X_{(n)}^{n-1}) \xrightarrow{P^{\infty}_{\theta_0}} 0.$$

Neka je $\epsilon > 0$:

$$\begin{aligned} P^{\infty}_{\theta_0}(|\frac{1}{n}\log(1 - X_{(n)}^{n-1})| \geq \epsilon) &= P^{\infty}_{\theta_0}(|\log(1 - X_{(n)}^{n-1})| \geq n\epsilon) \\ &= P^{\infty}_{\theta_0}(-\log(1 - X_{(n)}^{n-1}) \geq n\epsilon) \\ &= P^{\infty}_{\theta_0}(1 - X_{(n)}^{n-1} \leq e^{-n\epsilon}) \\ &= 1 - P^{\infty}_{\theta_0}(X_{(n)}^{n-1} < 1 - e^{-n\epsilon}) \\ &= 1 - P^{\infty}_{\theta_0}(X_{(n)} < (1 - e^{-n\epsilon})^{\frac{1}{n-1}}) \\ &= 1 - P^{\infty}_{\theta_0}(X_1 < (1 - e^{-n\epsilon})^{\frac{1}{n-1}}, \ldots, X_n < (1 - e^{-n\epsilon})^{\frac{1}{n-1}}) \\ &= 1 - P^{\infty}_{\theta_0}(X_1 < (1 - e^{-n\epsilon})^{\frac{1}{n-1}})^n \\ &= 1 - (1 - e^{-n\epsilon})^{\frac{n}{n-1}}. \end{aligned}$$

Druga jednakost je zbog ($0 \leq X_{(n)} \leq 1$ $P^{\infty}_{\theta_0}$*−g.s.). Tvrdnja slijedi jer je* $\lim_{n\to\infty}(1-e^{-n\epsilon})^{\frac{n}{n-1}} = 1$. *Konačno, za zadnji član u relaciji* (3.12)*, koristeći činjenicu da je na* $(0, 1)$ *funkcija gustoće* $\pi(\theta) = e^{-1/(\theta-\theta_0)}$*, slijedi:*

$$\frac{1}{n}\log\pi(X_{(n)}) = \frac{1}{n}\frac{-1}{X_{(n)} - \theta_0}.$$

Za proizvoljni $M > 0$ *je*

$$P^{\infty}_{\theta_0}(\frac{1}{n}\frac{-1}{X_{(n)} - \theta_0} \leq -M) = P^{\infty}_{\theta_0}(\frac{1}{nM} + \theta_0 \geq X_{(n)}) = 1 \quad (X_{(n)} \leq \theta_0 \ P^{\infty}_{\theta_0} - g.s.).$$

Time smo pokazali da je $\frac{1}{n}\log\pi(X_{(n)}) \xrightarrow{P^{\infty}_{\theta_0}} -\infty$. *Dakle,* $\frac{1}{n}\log I_n^1 \xrightarrow{P^{\infty}_{\theta_0}} -\infty$.
Za I_n^2 *vrijedi:*

$$\frac{1}{3^n}\Pi(2,3) \leq \int_2^3 \theta^{-n}\pi(\theta)\,d\theta \leq \frac{1}{2^n}\Pi(2,3).$$

Logaritmirajmo gornji izraz i podijelimo ga sa n. Dobivamo sljedeće nejednakosti :

$$\begin{aligned}-(\log\ 3) \leq -(\log\ 3) + \frac{1}{n}\log(\Pi(2,3)) &\leq \frac{1}{n}\ \log\ I_n^2\\ &\leq -(\log\ 2) + \frac{1}{n}\log(\Pi(2,3))\\ &\leq -(\log\ 2) + \Pi(2,3).\end{aligned}$$

Pokazali smo da je niz $(\frac{1}{n}\log\ I_n^2)_n$ *uniformno i apsolutno omeđen. Zbog toga vrijedi*

$$\frac{\log(I_n^1)}{\log(I_n^2)} \xrightarrow{P_{\theta_0}^\infty} -\infty\ .$$

Imamo:

$$\log\frac{I_n^1}{I_n^2} = \log\ I_n^1 - \log\ I_n^2 = \frac{\log\ I_n^1 - \log\ I_n^2}{\log\ I_n^2}\log\ I_n^2 = (\frac{\log(I_n^1)}{\log(I_n^2)} - 1)\log\ I_n^2.$$

Iz gornje jednakosti slijedi $\log\frac{I_n^1}{I_n^2} \xrightarrow{P_{\theta_0}^\infty} -\infty$*, a to je ekvivalentno* $\frac{I_n^1}{I_n^2} \xrightarrow{P_{\theta_0}^\infty} 0$*. Konačno, sada imamo :*

$$\Pi((2,3)\mid \mathbf{X_n}) = \frac{1}{1+\frac{\int_{X_{(n)}}^1 \theta^{-n}\pi(\theta)\,d\theta}{\int_2^3\theta^{-n}\pi(\theta)\,d\theta}} \xrightarrow{P_{\theta_0}^\infty} 1.$$

Time smo pokazali da niz aposteriorni distribucija nije konzistentan u $\theta_0 = 1$*. S druge strane, pokažimo da je* $(\hat{\theta}_n)_n = (X_{(n)})_n$ *jako konzistentan za* θ_0*. Ekvivalentno je pokazati*[1] *da za svaki* $\epsilon > 0$ *vrijedi*

$$\lim_{n\to\infty} P_{\theta_0}^\infty(\bigcup_{k=n}^{\infty}\{|X_{(k)} - \theta_0| \geq \epsilon\}) = 0. \tag{3.13}$$

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

$$\begin{aligned}P_{\theta_0}^\infty(|X_{(k)} - \theta_0| \geq \epsilon) =& P_{\theta_0}^\infty(X_{(k)} \geq \epsilon + \theta_0) + P_{\theta_0}^\infty(X_{(k)} \leq \theta_0 - \epsilon)\\ \\ =&\Big(X_{(n)} \leq \theta_0\ \ P_{\theta_0}^\infty - g.s\ \ \Rightarrow P_{\theta_0}^\infty(X_{(k)} \geq \epsilon + \theta_0) = 0\Big)\\ \\ =&P_{\theta_0}^\infty(X_{(k)} \leq \theta_0 - \epsilon).\end{aligned}$$

[1]Pogledati [9], propozicija 10.16.

Odredimo funkciju distribucije od $X_{(n)}$:

$$
\begin{aligned}
F_{X_{(n)}}(t) = P_{\theta_0}^{\infty}(X_{(n)} \le t) &= P_{\theta_0}^{\infty}(\max(X_1, X_2, \dots, X_n) \le t) \\
&= P_{\theta_0}^{\infty}(X_1 \le t,\ X_2 \le t,\ \dots, X_n \le t) \\
&= \prod_{k=1}^{n} P_{\theta_0}^{\infty}(X_k \le t) = \begin{cases} (\frac{t}{\theta_0})^n & X_{(n)} \in [0, \theta_0] \\ 0 & \textit{inače} \end{cases} .
\end{aligned}
$$

Pokažimo da (3.13) *vrijedi:*

$$
\begin{aligned}
P_{\theta_0}^{\infty}\Big(\bigcup_{k=n}^{\infty} \{|X_{(k)} - \theta_0| \ge \epsilon\} \Big) &\le \sum_{k=n}^{\infty} P_{\theta_0}^{\infty}(|X_{(k)} - \theta_0| \ge \epsilon) \\
&= \sum_{k=n}^{\infty} P_{\theta_0}^{\infty}(X_{(k)} \le \theta_0 - \epsilon) \\
&= \sum_{k=n}^{\infty} \left(\frac{\theta_0 - \epsilon}{\theta_0} \right)^k \xrightarrow{n \to \infty} 0.
\end{aligned}
$$

Konvergencija u zadnjem retku vrijedi jer je $\sum_{k=n}^{\infty} \left(\frac{\theta_0 - \epsilon}{\theta_0}\right)^k$ ostatak konvergentnog reda. Zaključujemo da je $(\hat{\theta}_n)_n$ jako konzistentan niz procjenitelja za $\theta_0 = 1$.

Napomena 3.0.17.
U primjeru 3.0.16 *naveli smo primjer u kojemu je procjenitelj maksimalne vjerodostojnosti konzistentan, a aposteriorna distribucija nije konzistentna na cijelom nosaču. Za primjer kada je aposteriorna distribucija konzistentna na cijelom nosaču, a procjenitelj maksimalne vjerodostojnosti nije konzistentan pogledati [5], stranica 29.*

Poglavlje 4

Asimptotska normalnost

4.1 Asimptotska normalnost aposteriorne distribucije

Jedan od standardnih rezultata u klasičnoj statistici je asimptotska normalnost procjenitelja maksimalne vjerodostojnosti. U ovom poglavlju ćemo se prisjetiti tog rezultata. Točnije, govorit ćemo o asimptotskoj normalnosti konzistentnih korijena jednadžbe vjerodostojnosti

$$\sum_{k=1}^{n} \frac{\partial}{\partial\theta} \log f(x_k|\theta) = 0.$$

Iz tog rezultata će slijediti asimptotska normalnost procjenitelja maksimalne vjerodostojnosti u slučaju da je $\theta \mapsto f(x|\theta)$ diferencijabilna i niz procjenitelja konzistentan za neki $\theta_0 \in \operatorname{Int}\Theta$ jer će tada niz procjenitelja maksimalne vjerodostojnosti biti i rješenje jednadžbe vjerodostojnosti. Nadalje, pokazat ćemo da vrijede analogni rezultati vezani za aposteriornu distribuciju te bayesovske procjenitelje obzirom na kvadratnu i apsolutnu funkciju gubitka.

Uvedimo sada neke osnovne pojmove i iskažimo teoreme potrebne za dokaz asimptotske normalnosti konzistentnih korijena jednadžbe vjerodostojnosti. U nastavku pretpostavljamo da je Θ otvoren i konveksan podskup of $\mathbb{R}^k$.

Definirajmo funkciju $\Psi : \mathbf{X} \times \Theta \to \mathbb{R}^k$ sa:

$$\Psi(x,\theta) = (\frac{\partial}{\partial\theta} \log f(x|\theta))^\top.$$

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

$$\mathcal{F}(\theta) = E_\theta \Psi(X,\theta)\Psi(X,\theta)^\top.$$

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

$$E_\theta \Psi(X,\theta) = \int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} f(x|\theta)\, d\mu(x) = \int \frac{\partial}{\partial\theta} f(x|\theta)\, d\mu(x) = 0.$$

Zaključujemo da je u tom slučaju Fisherova informacija zapravo matrica kovarijance od Ψ:

$$\mathcal{F}(\theta) = Var_\theta(\Psi(X,\theta)).$$

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

$$E_\theta(\dot{\Psi}(X,\theta)) = \int \left[\frac{\partial}{\partial\theta} \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} \right] f(x|\theta)\, d\mu(x)$$

$$\begin{aligned}
&= \int \frac{f(x|\theta)\ \frac{\partial^2}{\partial\theta^2} f(x|\theta) - \frac{\partial}{\partial\theta} f(x|\theta)^\top \frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)^2}\, f(x|\theta)\, d\mu(x) \\
&= \mathbf{0} - \int \Psi(x,\theta)\Psi(x,\theta)^\top f(x|\theta)\, d\mu(x) \\
&= -\mathcal{F}(\theta).
\end{aligned}$$

Dokazi sljedeća dva teorema mogu se naći u [4] i [10].

Teorem 4.1.1. *(Slutsky)*
Neka su $(X_n, n \in \mathbb{N})$ i $(Y_n, n \in \mathbb{N})$ nizovi slučajnih varijabli takvi da $X_n \xrightarrow{\mathcal{D}} X$ i $Y_n \xrightarrow{P} c$, gdje je c konstanta. Tada vrijedi sljedeće:

(i) $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$,

(ii) $X_n \cdot Y_n \xrightarrow{\mathcal{D}} cX$

(iii) ako je $c \neq 0$, onda $\frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X}{c}$.

Teorem 4.1.2. *(Scheffé)*
Neka je $(f_n,\ n \in \mathbb{N},)$ niz funkcija gustoće (obzirom na neku mjeru μ) i f funkcija gustoće. Vrijedi sljedeće:

$$\lim_{n\to\infty} f_n = f \ \ \mu - g.s. \Rightarrow \int |f_n(x) - f(x)|\, d\mu(x) \to\ 0.$$

Navedimo jednu posljedicu teorema 4.1.2.

Korolar 4.1.3.
Neka je $(S, \mathcal{F}, P)$ vjerojatnosni prostor te $(X_n)_n$ i X slučajne varijable na zadanom prostoru. Označimo sa $(f_n)_n$ i f pripadne funkcije gustoće (obzirom na neku mjeru μ). Ako je $\lim_{n\to\infty} f_n(x) = f(x)\ \ \mu - g.s.$, tada vrijedi:

$$\sup_{A\in\mathcal{F}} |P(X_n \in A) - P(X \in A)| \to 0. \tag{4.1}$$

Dakle, korolar 4.1.3 nam govori da točkovna konvergencija niza gustoća povlači konvergenciju slučajnih varijabli u smislu relacije (4.1.). Upravo ovakav tip konvergencije ćemo susresti u teoremu vezanom za asimptotsku normalnost aposteriorne distribucije. Primijetimo da (4.1.) povlači konvergenciju po distribuciji (obrat ne vrijedi općenito). Dakle, pokazat ćemo jaču tvrdnju u odnosu na rezultat o asimptotskoj normalnost procjenitelja maksimalne vjerodostojnosti koji nam garantira samo konvergenciju u distribuciji.

Teorem 4.1.4.
Neka su $X_1, X_2, \ldots, X_n$ nezavisne i jednako distribuirane slučajne varijable sa gustoćom $f(x|\theta)$ (obzirom na mjeru μ) te neka je θ_0 stvarna vrijednost parametra. Pretpostavimo da su zadovoljeni sljedeći uvjeti:

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

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

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

(3) $\mathcal{F}(\theta_0) = -E_{\theta_0}\dot{\Psi}(x,\theta_0)$ je pozitivno definitna matrica.

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

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

Dokaz.
Prvo ćemo dokazati egzistenciju konzistentnih korijena, a potom asimptotsku normalnost.

1. *Postojanje konzistentnih korijena*
   Neka je $\delta > 0$ takav da je $S_\delta := \{\theta : d(\theta,\theta_0) \leq \delta\}$ sadržan u otvorenoj okolini od θ_0 iz uvjeta (2). Primijetimo da je S_δ kompaktan skup. Zbog uvjeta (1) je funkcija $\theta \mapsto f(x|\theta)$ neprekidna za svaki x. Neprekidna funkcija na kompaktnom skupu postiže maksimum, dakle postoji $\hat{\theta}_n$ takav da vrijedi

$$L_n(\hat{\theta}_n) = \max_{\theta\in S_\delta} L_n(\theta). \tag{4.2}$$

   Jaka konzistentnost tako konstruiranog niza će slijediti primjenom teorema 3.0.15 na skup S_δ nakon što pokažemo da su zadovoljene sve pretpostavke teorema 3.0.11. Uvjeti $(i) - (ii)$ teorema 3.0.11 su automatski zadovoljeni. Budući da je S_δ konveksan skup, primjenom Taylorovog teorema srednje vrijednosti na funkciju $U(x,\theta) := \log \frac{f(x|\theta)}{f(x|\theta_0)}$ imamo:

$$\begin{aligned} U(x,\theta) =& U(x,\theta_0) + \Psi(x,\theta_0)^\top(\theta-\theta_0) \\ &+ (\theta-\theta_0)^\top \left(\int_0^1 \int_0^1 \lambda\dot{\Psi}(x,\theta+\lambda\eta(\theta-\theta_0))\ d\lambda d\eta\right)(\theta-\theta_0). \end{aligned} \tag{4.3}$$

Funkcija $x \mapsto \Psi(x,\theta_0)$ je integrabila, $U(x,\theta_0) = 0$ i zadnji član u gornjoj jednakosti je ograničen zbog pretpostavke (2) pa je zadovoljen uvjet (*iii*). Time je dokazana jaka konzistentnost.
Zbog konzistentnosti niza $(\hat{\theta}_n)_n$ za okolinu Int $(S_\delta) \ni \theta_0$ postoji n_0 tako da je za svaki $n \geq n_0$, $\hat{\theta}_n \in$ Int (S_δ). Sada zbog (4.2) slijedi da je za svaki $n \geq n_0$, $\hat{\theta}_n$ korijen jednadžbe vjerodostojnosti, odnosno vrijedi:

$$\frac{\partial}{\partial\theta}\log L_n(\hat{\theta}_n) = \sum_{k=1}^n \frac{\partial}{\partial\theta}\log f(x_k|\hat{\theta}_n) = 0.$$

2. *Asimptotska normalnost*
Za funkciju log-vjerodostojnosti $l_n(\theta) := \log L_n(\theta)$ vrijedi $\dot{l}_n(\theta) = \sum_{k=1}^n \Psi(X_k,\theta)$. Možemo ju zbog Taylorovog teorema zapisati na sljedeći način :

$$\dot{l}_n(\theta) = \dot{l}_n(\theta_0) + \int_0^1 \sum_{k=1}^n \dot{\Psi}(X_k,\theta_0+\lambda(\theta-\theta_0))\,d\lambda\,(\theta-\theta_0).$$

Neka je $(\hat{\theta}_n)_n$ jako konzistentan niz korijena jednadžbe vjerodostojnosti, dakle za dovoljno veliki n je $\dot{l}_n(\hat{\theta}_n) = 0$. Stavimo $\theta = \hat{\theta}_n$, uvrstimo u gornji izraz te ga potom podijelimo sa $\sqrt{n}$:

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

gdje je $B_n := -\int_0^1 \frac{1}{n}\sum_{k=1}^n \dot{\Psi}(X_k,\theta_0+\lambda(\hat{\theta}_n-\theta_0))\,d\lambda$.
Kako je $E_{\theta_0}\Psi(X,\theta_0) = 0$ i $Var_{\theta_0}\Psi(X,\theta_0) = \mathcal{F}(\theta_0)$, primjenom centralnog graničnog teorema slijedi:

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{k=1}^n \Psi(X_k,\theta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0},\mathcal{F}(\theta_0)).$$

Da bi dokazali tvrdnju teorema dovoljno je pokazati da $B_n \xrightarrow{g.s.} \mathcal{F}(\theta_0)$. Naime, po pretpostavci je $\mathcal{F}(\theta_0)$ regularna pa će za dovoljno veliki n postojati B_n^{-1}. Preslikavanje $A \mapsto A^{-1}$, $A \in GL_n(\mathbb{R})$ je neprekidno, pa primjenom teorema 4.1.1 zaključujemo:

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

Neka je $\epsilon > 0$. Koristeći pretpostavku (2), primjenom teorema o dominiranoj konvergenciji lako se pokaže da je $E_{\theta_0}\dot{\Psi}(X,\theta)$ neprekidna u θ u okolini od θ_0 pa postoji $\delta > 0$ tako da vrijedi:

$$d(\theta,\theta_0) < \delta \Rightarrow \left|E_{\theta_0}\dot{\Psi}(X,\theta) - E_{\theta_0}\dot{\Psi}(X,\theta_0)\right| = \left|E_{\theta_0}\dot{\Psi}(X,\theta) + \mathcal{F}(\theta_0)\right| < \epsilon.$$

Sada, koristeći uniformi jaki zakon velikih brojeva zaključujemo da postoji n_0 takav da je:

$$\forall n \geq n_0 \ \sup_{\theta\in S_\delta}\left|\frac{1}{n}\sum_{k=1}^{n}\dot{\Psi}(X_k,\theta) - E_{\theta_0}\dot{\Psi}(X_k,\theta)\right| < \epsilon.$$

Zbog jake konzistentosti postoji n_1 takav da je za $n \geq n_1$, $d(\hat{\theta}_n,\theta_0) < \delta$. Konačno, za svaki $n \geq \max\{n_0,n_1\}$ imamo :

$$\begin{aligned}
\|B_n - \mathcal{F}(\theta_0)\| &\leq \int_0^1 \left|\frac{1}{n}\sum_{k=1}^{n}\dot{\Psi}(X_k,\theta_0 + \lambda(\hat{\theta}_n - \theta_0)) + \mathcal{F}(\theta_0)\right| d\lambda \\
&\leq \int_0^1 \sup_{\theta\in S_\delta}\left[\left|\frac{1}{n}\sum_{k=1}^{n}\dot{\Psi}(X_k,\theta) - E_{\theta_0}\dot{\Psi}(X,\theta)\right| + |E_{\theta_0}\dot{\Psi}(X,\theta) + \mathcal{F}(\theta_0)|\right] d\lambda \\
&< 2\epsilon. \qquad (4.4)
\end{aligned}$$

□

Napomena 4.1.5.
U drugom dijelu prethodnog teorema n_0, n_1 ovise o realizaciji slučajnog uzorka $X_1,X_2,\ldots,X_n$.

Napomena 4.1.6.
Tvrdnja teorema 4.1.4 nam neće uvijek osigurati asimptotsku normalnost procjenitelja maksimalne vjerodostojnosti. Naime, pod uvjetima teorema 4.1.4, procjenitelj maksimalne vjerodostojnosti uopće ne mora zadovoljavati jednadžbu vjerodostojnosti. Nadalje, može zadovoljavati jednadžbu vjerodostojnosti no ne mora biti konzistentan. Isto tako, u slučaju da korijen jednadžbe vjerodostojnosti nije jedinstven, teorem nam ne odgovara na pitanje koji korijen izabrati.

Neka je $x_1,x_2,\ldots,x_n$ realizacija slučajnog uzorka i $(\hat{\theta}_n)_n$ konzistentan niz korijena jednadžbe vjerodostojnosti. Označimo uvjetnu funkciju gustoće slučajne varijable $\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n)$

sa $f_n(\nu \mid x_1, x_2, \ldots x_n)$. Primjenom [9, Tm 11.9.] na funkciju $g(\theta) = \sqrt{n}(\theta - \hat{\theta}_n)$ obzirom na aposteriornu funkciju gustoće slijedi:

$$
\begin{aligned}
f_n(\nu \mid x_1, x_2, \ldots x_n) &= \frac{1}{\sqrt{n}}\, \pi(\hat{\theta}_n + \frac{\nu}{\sqrt{n}} \mid \mathbf{x}_n) \\
&= \frac{1}{\sqrt{n}} \frac{L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\, \pi(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})}{\int L_n(\zeta)\, \pi(\zeta)\, d\zeta} = \left[\begin{smallmatrix} \zeta = \hat{\theta}_n + \frac{\xi}{\sqrt{n}} \\ \sqrt{n}\, d\zeta = d\xi \end{smallmatrix} \right] \\
&= \frac{1}{\sqrt{n}} \frac{L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\, \pi(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})}{\frac{1}{\sqrt{n}} \int L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, d\xi} \\
&= \frac{L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\, \pi(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})}{\int L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, d\xi}.
\end{aligned}
$$

Dakle, uvjetna gustoća od $\nu = \sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n)$ uz danu realizaciju slučajnog uzorka je:

$$
f_n(\nu | x_1, x_2, \ldots, x_n) = \frac{L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})}{\int L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi}. \tag{4.5}
$$

Pogledajmo sljedeći primjer kao motivaciju za asimptotsku konzistentnost aposteriorne distribucije.

Primjer 4.1.7.
U primjeru 1.0.3 *b) smo pokazali da, ako je slučajan uzorak* $X_1, X_2, \ldots, X_n \sim N(\theta, \sigma^2)$ *i* $\boldsymbol{\theta} \sim N(\mu, \tau)$*, da je aposteriorna distribucija također normalno distribuirana:*

$$
\boldsymbol{\theta} \mid \boldsymbol{X}_n \sim N\bigg(\frac{\tau^2 \sigma 2}{n\tau^2 + \sigma^2} \big(\frac{n}{\sigma^2} \overline{X}_n + \frac{\mu}{\tau^2} \big) \, , \, \frac{\sigma^2 \tau^2}{n\tau^2 + \sigma^2} \bigg).
$$

Procjenitelj maksimalne vjerodostojnosi za parametar θ *je* $\hat{\theta}_n = \overline{X}_n$*. Sada je aposteriorna distribucija od* $\sqrt{n}\,(\boldsymbol{\theta} - \overline{X}_n)$*:*

$$
\sqrt{n}\,(\boldsymbol{\theta} - \overline{X}_n) \mid \boldsymbol{X}_n \sim N\left(\frac{\frac{\sigma^2}{\sqrt{n}}}{\sigma^2/n + \tau^2}\, (\mu - \overline{X}_n), \frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2/n} \right).
$$

Pretpostavimo sada da je θ_0 *prava vrijednost parametra* θ*. Po jakom zakonu velikih brojeva je* $\overline{X}_n \xrightarrow{g.s.} \theta_0$*. Označimo sa* f_n *funkciju gustoće aposteriorne distribucije od* $\sqrt{n}\,(\boldsymbol{\theta} - \overline{X}_n)$*.*

Radi jednostavnosti uvodimo oznake za parametre:

$$\alpha_n = \frac{\frac{\sigma^2}{\sqrt{n}}}{\sigma^2/n + \tau^2}(\mu - \overline{X}_n), \quad \beta_n^2 = \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2/n}.$$

Vrijedi $\lim\limits_{n\to\infty} \alpha_n = 0$ *g.s. i* $\lim\limits_{n\to\infty} \beta_n^2 = \sigma^2$ *g.s. te iz toga slijedi:*

$$f_n(v) = \frac{1}{\sqrt{2\pi}\beta_n} e^{-\frac{(v-\alpha_n)^2}{2\beta_n^2}} \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{v^2}{2\sigma^2}} \ g.s.$$

Vidimo da niz $(f_n)_n$ *konvergira ka gustoći* $f \sim N(0, \sigma^2)$*. Teorem 4.1.2 nam govori da vrijedi i više. Točnije, vrijedi:*

$$\int |f_n(v) - \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{v^2}{2\sigma^2}}| dv \to 0 \quad P_{\theta_0}^\infty - g.s.$$

Pokazuje se da je $\mathcal{F}(\theta_0) = \frac{1}{\sigma^2}$ *pa vidimo da imamo sličan rezultat kao u prethodnom teoremu za procjenitelje maksimalne vjerodostojnosti.*

Htjeli bi imati takav rezultat kada apriorna distribucija ili slučajan uzorak ne prate normalnu distribuciju. Pokazat ćemo da takva tvrdnja vrijedi u općenitijem slučaju, ali uz nekoliko pretpostavki. U nastavku poglavlja pretpostavljamo da je parametarski prostor Θ otvoren i konveksan podskup od $\mathbb{R}$.

Teorem 4.1.8. *(Bernstein-von Mises)*
Neka su zadovoljene pretpostavke teorema 4.1.4 te neka je $(\hat{\theta}_n)_n$ *konzistentan niz korijena jednadžbe vjerodostojnosti. Pretpostavimo da još vrijedi:*

(4) apriorna gustoća $\pi(\theta)$ *je neprekidna i strogo pozitivna na* Θ*,*

(5) za svaki θ *i* $\delta > 0$ *postoji* $\epsilon(\delta, \theta) > 0$ *i* N *takav da za svaki* $n \geq N$

$$\sup_{|t-\theta|>\delta} \frac{1}{n}\Big(l_n(t) - l_n(\theta)\Big) \leq -\epsilon(\delta, \theta) \quad P_\theta^\infty - g.s.$$

Tada imamo:

$$\int \Big|f_n(\zeta \mid \boldsymbol{x}_n) - \phi(\zeta)\Big| d\zeta \ \to \ 0 \ P_\theta^\infty - g.s.$$

gdje je ϕ *funkcija gustoće slučajne varijable* $Y \sim N(0\,, \mathcal{F}(\theta)^{-1})$*.*

Dokaz.
Neka je $\theta_0 \in \Theta$ fiksna vrijednost (prava vrijednost parametra). Dokaz provodimo u dva dijela. Prvo dokazujemo da vrijedi:

$$\frac{L_n(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}}) \xrightarrow{g.s} \exp\left\{-\frac{\zeta^2}{2}\mathcal{F}(\theta_0)\right\}\pi(\theta_0), \tag{4.6}$$

a potom da vrijedi:

$$\int \frac{L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, d\xi \xrightarrow{g.s} \int \exp\left\{-\frac{\xi^2}{2}\mathcal{F}(\theta_0)\right\} d\xi \ \pi(\theta_0). \tag{4.7}$$

Korištenjem gornjih tvrdnji rezultat teorema lagano slijedi. Naime, iz relacija (4.6) i (4.7) slijedi

$$f_n(\zeta|x_1, x_2, \ldots, x_n) = \frac{\frac{L_n(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})}{\int \frac{L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\, d\xi} \xrightarrow{g.s} \frac{\exp\left\{-\frac{\zeta^2}{2}\mathcal{F}(\theta_0)\right\}\pi(\theta_0)}{\int \exp\left\{-\frac{\xi^2}{2}\mathcal{F}(\theta_0)\right\} d\xi \ \pi(\theta_0)}$$

$$= \frac{\exp\left\{-\frac{\zeta^2}{2}\mathcal{F}(\theta_0)\right\}}{\sqrt{2\pi}/\sqrt{\mathcal{F}(\theta_0)}} = \phi(\zeta).$$

Dakle, vrijedi $f_n(\zeta|x_1, x_2, \ldots, x_n) \xrightarrow{g.s.} \phi(\zeta)$. Tvrdnja teorema sada proizlazi iz teorema 4.1.2
Dokažimo (4.6). Primjenom Taylorovog teorema srednje vrijednosti imamo:

$$l_n(\theta) = l_n(\hat{\theta}_n) + \dot{l}_n(\hat{\theta}_n)(\theta - \hat{\theta}_n) + \frac{1}{2}(\theta - \hat{\theta}_n)^2\ddot{l}_n(\theta_n^*), \quad \theta_n^* \in (\hat{\theta}_n, \theta).$$

Kako je za dovoljno veliki n $\dot{l}_n(\hat{\theta}_n) = 0$, imamo:

$$\frac{L_n(\theta)}{L_n(\hat{\theta}_n)} = \exp\{l_n(\theta) - l_n(\hat{\theta}_n)\} = \exp\{\frac{1}{2}(\theta - \hat{\theta}_n)^2\ddot{l}_n(\theta_n^*)\}. \tag{4.8}$$

Ako u (4.8) stavimo $\theta = \hat{\theta}_n + \zeta/\sqrt{n}$ tada će (θ_n^*) biti konzistentan niz za θ_0. Naime,

$$|\theta_n^* - \theta_0| \leq |\theta_n^* - \hat{\theta}_n| + |\hat{\theta}_n - \theta_0| < \frac{\zeta}{\sqrt{n}} + |\hat{\theta}_n - \theta_0| \to 0. \tag{4.9}$$

Sada kada imamo (4.9), kao u dokazu druge tvrdnje teorema 4.1.4 (relacija 4.4) se pokazuje konvergencija

$$\frac{1}{n}\,\ddot{l}_n(\theta_n^*) \xrightarrow{g.s.} E_{\theta_0}\dot{\Psi}(X,\theta_0).$$

Pokazali smo da je $E_{\theta_0}\dot{\Psi}(X,\theta_0) = -\mathcal{F}(\theta_0)$. Dakle,

$$\begin{aligned}
&\frac{L_n\left(\hat{\theta}_n + \zeta/\sqrt{n}\right)}{L_n\left(\hat{\theta}_n\right)}\;\pi(\hat{\theta}_n + \zeta/\sqrt{n}) \\
&\qquad = \exp\left\{\frac{\zeta^2}{2\,n}\ddot{l}_n(\theta^*)\right\}\pi(\hat{\theta}_n + \zeta/\sqrt{n}) \\
&\qquad \xrightarrow{g.s.} \exp\left\{-\frac{\zeta^2}{2}\mathcal{F}(\theta_0)\right\}\;\pi(\theta_0).
\end{aligned} \tag{4.10}$$

Preostaje nam dokazati tvrdnju 4.7. Neka je $\delta > 0$ takav da je zatvorena kugla oko θ_0 radijusa δ sadržana u okolini iz uvjeta (2) teorema 4.1.4. Zapišimo $\int \frac{L_n(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\;\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta$ na sljedeći način:

$$\begin{aligned}
\int \frac{L_n(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\;\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta &= \int e^{l_n(\hat{\theta}_n + \zeta/\sqrt{n}) - l_n(\hat{\theta}_n)}\,\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta \\
= \int\limits_{|\zeta|\le\delta\sqrt{n}} e^{l_n(\hat{\theta}_n + \zeta/\sqrt{n}) - l_n(\hat{\theta}_n)}\;\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta &+ \int\limits_{|\zeta|>\delta\sqrt{n}} e^{l_n(\hat{\theta}_n + \zeta/\sqrt{n}) - l_n(\hat{\theta}_n)}\,\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta.
\end{aligned} \tag{4.11}$$

Drugi član u gornjoj sumi zamjenom varijabli $t = \hat{\theta}_n + \frac{\zeta}{\sqrt{n}}$ možemo zapisati kao:

$$\int\limits_{|\zeta|>\delta\sqrt{n}} e^{l_n(\hat{\theta}_n + \zeta/\sqrt{n}) - l_n(\hat{\theta}_n)}\,\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}})\;d\zeta = \sqrt{n}\int\limits_{|t-\hat{\theta}_n|>\delta} e^{l_n(t) - l_n(\hat{\theta}_n)}\,\pi(t)\;dt.$$

Isto tako, kako je $\hat{\theta}_n \xrightarrow{g.s} \theta_0$, za dovoljno veliki n će biti $|\hat{\theta}_n - \theta_0| < \delta/2$. Za svaki $t \in \{t : |t - \hat{\theta}_n| > \delta\}$ je

$$|t-\theta_0| \ge |t-\hat{\theta}_n| - |\hat{\theta}_n - \theta_0| > \delta - \delta/2 = \delta/2.$$

Zbog pretpostavke (5) za dovoljno veliki n je

$$e^{l_n(t) - l_n(\theta_0)} \le e^{-n\epsilon(\delta/2,\theta_0)} \quad P_{\theta_0}^{\infty} - g.s., \tag{4.12}$$

a zbog neprekidnosti funkcije $\theta \mapsto l_n(\theta)$ za $C > 0$ postoji $\delta_C > 0$ takav da vrijedi:

$$|t - \theta_0| < \delta_C \Rightarrow |l_n(t) - l_n(\theta_0)| < C. \tag{4.13}$$

Sada, zbog konzistentnosti niza $(\hat{\theta}_n)_n$, za δ_C i dovoljno veliki n će vrijediti

$$|\hat{\theta}_n - \theta_0| < \delta_C. \tag{4.14}$$

Za dovoljno veliki n iz (4.12) , (4.13) i (4.14) slijedi:

$$\begin{aligned} e^{l_n(t)-l_n(\hat{\theta}_n)} = e^{l_n(t)-l_n(\theta_0)+l_n(\theta_0)-l_n(\hat{\theta}_n)} &\leq e^{l_n(\theta_0)-l_n(\hat{\theta}_n)}\ e^{-n\epsilon(\delta/2,\theta_0)} \\ &\leq e^C e^{-n\epsilon(\delta/2,\theta_0)}. \end{aligned}$$

Dakle,

$$\sqrt{n} \int\limits_{|t-\hat{\theta}_n|>\delta} e^{l_n(t)-l_n(\hat{\theta}_n)}\, \pi(t)\ dt \leq\ \sqrt{n}\ e^C e^{-n\epsilon(\delta/2,\theta_0)} \xrightarrow{n\to\infty} 0. \tag{4.15}$$

Promotrimo prvi sumand u (4.11). Kao u prvom dijelu dokaza imamo

$$l_n(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}}) - l_n(\hat{\theta}_n) = \frac{\zeta^2}{2n}\ddot{l}_n(\theta_n^*(\zeta)), \quad |\hat{\theta}_n - \theta_n^*(\zeta)| < \frac{\zeta}{\sqrt{n}}.$$

Zbog $\zeta \leq \delta\sqrt{n}$ je $|\hat{\theta}_n - \theta_n^*(\zeta)| < \frac{\zeta}{\sqrt{n}} \leq \delta$ pa imamo

$$|\theta_n^*(\zeta) - \theta_0| \leq |\theta_n^*(\zeta) - \hat{\theta}_n| + |\hat{\theta}_n - \theta_0| < \delta + |\hat{\theta}_n - \theta_0| \leq \delta.$$

Posljednja nejednakost će vrijediti za dovoljno veliki n. Sada, kao u prvom dijelu dokaza zaključujemo

$$\frac{1}{n}\ddot{l}_n(\theta_n^*(\zeta)) \xrightarrow{g.s} -\mathcal{F}(\theta_0).$$

pa za $\eta = \frac{1}{2}\mathcal{F}(\theta_0)$ i dovoljno veliki n imamo

$$\frac{1}{n}\ddot{l}_n(\theta_n^*(\zeta)) + \mathcal{F}(\theta_0) < \eta = \frac{1}{2}\mathcal{F}(\theta_0) \Rightarrow \frac{1}{n}\ddot{l}_n(\theta_n^*(\zeta)) < -\frac{1}{2}\mathcal{F}(\theta_0). \tag{4.16}$$

Po pretpostavci je apriorna gustoća neprekidna i $\pi(\theta_0) > 0$ pa će za dovoljno veliki n vrijediti:

$$|\pi(\hat{\theta}_n + \frac{\zeta}{\sqrt{n}}) - \pi(\theta_0)| < \pi(\theta_0). \tag{4.17}$$

Konačno, za dovoljno veliki n iz relacija (4.16) , (4.17) imamo:

$$e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\zeta/\sqrt{n})\mathbb{1}_{\{|\zeta|\le\delta\sqrt{n}\}} = e^{\frac{\zeta^2}{2n}\ddot{l}_n(\theta^*(\zeta))}\ \pi(\hat{\theta}_n+\zeta/\sqrt{n})\mathbb{1}_{\{|\zeta|\le\delta\sqrt{n}\}}$$

$$\le 2\pi(\theta_0)e^{-\frac{\zeta^2}{4}\mathcal{F}(\theta_0)}. \tag{4.18}$$

Pokazali smo da je funkcija pod integralom prvog sumanda u (4.11) dominirana integrabilnom funcijom pa primjenom teorema o dominiranoj konvergenciji imamo

$$\lim_{n\to\infty}\int\limits_{|\zeta|\le\delta\sqrt{n}} e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\ d\zeta = \int \lim_{n\to\infty} e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\mathbb{1}_{\{|\zeta|\le\delta\sqrt{n}\}}\ d\zeta$$

$$= \int \pi(\theta_0)\, e^{-\frac{\zeta^2\mathcal{F}(\theta_0)}{2}}\ d\zeta. \tag{4.19}$$

Objedinjavanjem gornjih rezultata slijedi:

$$\lim_{n\to\infty}\int \frac{L_n(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})}{L_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\ d\zeta =$$

$$= \lim_{n\to\infty}\int\limits_{|\zeta|\le\delta\sqrt{n}} e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\ d\zeta + \lim_{n\to\infty}\int\limits_{|\zeta|>\delta\sqrt{n}} e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\ d\zeta$$

$$= \Big(\text{teorem o dominiranoj konvergenciji i relacija (4.15)}\Big)$$

$$= \int\limits_{|\zeta|\le\delta\sqrt{n}} \lim_{n\to\infty} e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n+\frac{\zeta}{\sqrt{n}})\ d\zeta + 0$$

$$= \int \pi(\theta_0)\, e^{-\frac{-\zeta^2\mathcal{F}(\theta_0)}{2}}\ d\zeta.$$

$\square$

Ilustrirajmo na primjeru primjenu teorema 4.1.8

Primjer 4.1.9.
Neka je $X_1, X_2, \ldots X_n \sim \Gamma(1, 1/\theta)$ slučajan uzorak, pa je funkcija gustoće zadana sa $f(x|\theta) = \theta e^{-\theta x}\mathbb{1}_{(0,\infty)}(x)$. Apriorna distribucija parametra $\boldsymbol{\theta}$ je također Γ sa parametrima $(1, 1)$, to jest $\pi(\theta) = e^{-\theta}\mathbb{1}_{(0,\infty)}(\theta)$. Lako se pokaže da su zadovoljeni uvjeti teorema 4.1.4 te da je Fisherova informacija jednaka $\mathcal{F}(\theta) = \frac{1}{\theta^2}$. Nadalje, procjenitelj maksimalne vjerodostojnosti za parametar θ je $\frac{1}{\bar{X}_n}$. Pokažimo da su zadovoljena dodatna dva uvjeta teorema 4.1.8. Apriorna gustoća je očito neprekidna i strogo pozitivna na $\Theta = (0, \infty)$. Preostaje nam uvjet (5). Neka su θ i $\delta > 0$ proizvoljni. Imamo sljedeće:

$$\frac{1}{n}(l_n(t) - l_n(\theta)) = \frac{1}{n}\sum_{k=1}^{n}(\log f(x_k|t) - \log f(x_k|\theta))$$

$$= \frac{1}{n}\sum_{k=1}^{n}\log\frac{te^{-tx_k}}{\theta e^{-\theta x_k}} \to E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}})\ \ P_\theta^\infty - g.s. \tag{4.20}$$

Izračunajmo $E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}})$:

$$E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}}) = \int_0^\infty \log\frac{te^{-tx}}{\theta e^{-\theta x}}\theta e^{-\theta x}dx$$

$$= \int_0^\infty(\log t - xt - \log\theta + x\theta)\theta e^{-\theta x}dx$$

$$= \log t - \log\theta - \frac{t}{\theta} + 1. \tag{4.21}$$

Ako promotrimo $E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}})$ kao funkciju po t, lako se pokaže da poprima jedinstveni globalni maksimum u θ i on je jednak 0. Time smo pokazali da je $\sup\limits_{|t-\theta|>\delta} E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}}) < 0$. Za $-\epsilon(\delta,\theta) := \sup\limits_{|t-\theta|>\delta} E_\theta(\log\frac{te^{-tX}}{\theta e^{-\theta X}})$ će zbog relacije (4.20) za dovoljno veliki n, odnosno za sve n osim konačno mnogo njih vrijediti

$$\sup_{|t-\theta|>\delta}\frac{1}{n}\Big(l_n(t) - l_n(\theta)\Big) \leq \frac{-\epsilon(\delta,\theta)}{2} \quad P_\theta^\infty - g.s.$$

Dakle, zadovoljeni su uvjeti teorema 4.1.8. Ako sa $f(\nu|\mathbf{x}_n)$ označimo aposteriornu gustoću od $\sqrt{n}(\boldsymbol{\theta} - \frac{1}{\bar{x}_n})$, primjenom teorema 4.1.8 slijedi:

$$\int |f(\nu|\mathbf{x}_n) - \phi(\nu)|d\nu \xrightarrow{g.s.} 0.$$

4.2 Asimptotska normalnost Bayesovog procjenitelja

Na početku poglavlja smo iskazali i dokazali rezultat vezan za asimptotsku normalnost procjenitelja maksimalne vjerodostojnosti. U jednom od prethodnih poglavlja smo definirali Bayesove procjenitelje za apsolutnu i kvadratnu funkciju gubitka. Sada nas zanima vrijedi li isti asimptotski rezultat i za takve procjenitelje te u kakvom su odnosu sa procjeniteljem maksimalne vjerodostojnosti. Pokazuje se da vrijedi te da su zapravo asimptotski ekvivalentni sa procjeniteljem maksimalne vjerodostojnosti. Ti rezultati su zapravo posljedica teorema 4.1.8

Teorem 4.2.1.
Pretpostavimo da uz uvjete teorema 4.1.8 vrijedi i $\int \theta^2\, \pi(\theta)\, d\theta < \infty$*. Neka je* θ_M *medijan aposteriorne distribucije od* $\boldsymbol{\theta}$*. Tada za svaki* $\theta_0 \in \Theta$ *imamo:*

(i) $\sqrt{n}\,(\theta_M - \hat{\theta}_n) \to 0\ \ P^{\infty}_{\theta_0} - g.s.$

(ii) $\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \hat{\theta}_n) \to 0\ \ P^{\infty}_{\theta_0} - g.s.$

Prije dokaza teorema iskažimo i dokažimo neke pomoćne tvrdnje.

Teorem 4.2.2. *(Polya)*
Neka niz $(X_n,\ n \in \mathbb{N})$ *realnih slučajnih varijabli konvergira po distribuciji prema slučajnoj varijabli X. Označimo sa* F_n *i F pripadne funkcije distribucije od* X_n, X*. Pretpostavimo da je F neprekidna funkcija. Tada vrijedi:*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0.$$

Dokaz prethodnog teorema može se naći u [2].

Lema 4.2.3.
Neka niz $(X_n,\ n \in \mathbb{N})$ *realnih slučajnih varijabli konvergira po distribuciji prema slučajnoj varijabli X. Označimo sa* F_n *i F pripadne funkcije distribucije od* X_n, X*. Pretpostavimo da je F neprekidna i strogo rastuća funkcija distribucije. Tada za svaki* $0 < \alpha < 1$ *vrijedi:*

$$|F_n^{-1}(\alpha) - F^{-1}(\alpha)| \to 0\,, \tag{4.22}$$

gdje je $F_n^{-1}(\alpha) = \inf\{x \in \mathbb{R}\ :\ F_n(x) \geq \alpha\}$.

Dokaz.
Definirajmo za $n \in \mathbb{N}$ skup $S_n = \{x \in \mathbb{R}\ :\ F_n(x) \geq \alpha\}$. Budući da je $F_n^{-1}(\alpha) = \inf S_n$, za svaki n postoji $z_n \in S_n$ takav da je

$$z_n - \frac{1}{n} < F_n^{-1}(\alpha) \leq z_n. \tag{4.23}$$

Iz toga slijedi da $z_n - \frac{1}{n}$ nije element skupa S_n pa zaključujemo:

$$F_n(z_n - \frac{1}{n}) < \alpha \leq F_n(z_n). \tag{4.24}$$

Da bismo dokazali (4.22), dovoljno je pokazati da je $\lim_{n\to\infty} z_n = F^{-1}(\alpha)$. Ako to pokažemo, tvrdnja leme će slijediti iz (4.23) primjenom teorema o sendviču. Zbog neprekidnost i strogog rasta funkcije F postoji jedinstveni $x \in \mathbb{R}$ takav da je $F(x) = \alpha$. Pokažimo da je $\lim_{n\to\infty} F_n(z_n) = F(x)$. Ako za sve osim najviše konačno mnogo n vrijedi $x \geq z_n$, tada postoji n_0 takav da za svaki $n \geq n_0$ (uz primjenu relacije (4.24)) vrijedi

$$F(x) \leq F_n(z_n) \leq F_n(x).$$

Iz pretpostavke konvergencije po distribuciji i korištenjem teorema o sendviču slijedi da je $\lim_{n\to\infty} F_n(z_n) = F(x)$. Ako je za beskonačno mnogo n, $z_n > x$ tada imamo podniz $(z_{n_k})_k$ takav da za svaki k vrijedi $z_{n_k} > x$. Za dovoljno veliki k slijedi $z_{n_k} - \frac{1}{n_k} > x$ pa iz (4.24) slijedi

$$F_{n_k}(x) \leq F_{n_k}(z_{n_k} - \frac{1}{n_k}) < F(x).$$

Zaključujemo da je $\lim_{k\to\infty} F_{n_k}(z_{n_k} - \frac{1}{n_k}) = F(x)$. Pokažimo sada da je $\lim_{k\to\infty} F_{n_k}(z_{n_k}) = F(x)$. Vrijedi sljedeće:

$$\begin{aligned} |F_{n_k}(z_{n_k}) - F(x)| \leq & |F_{n_k}(z_{n_k}) - F(z_{n_k})| + |F(z_{n_k}) - F(z_{n_k} - 1/n_k)| \\ & + |F(z_{n_k} - 1/n_k) - F_{n_k}(z_{n_k} - 1/n_k)| + |F_{n_k}(z_{n_k} - 1/n_k) - F(x)|. \end{aligned}$$

Prvi i treći sumand u nejednakosti teže ka nuli zbog teorema 4.2.2, a zadnji sumand po prethodno dokazanom. F je neprekidna funkcija distribucije pa je stoga uniformno neprekidna. Zbog toga drugi sumand u nejednakosti teži k nuli. Prema tome, imamo $\lim_{k\to\infty} F_{n_k}(z_{n_k}) = F(x)$. Time smo dokazali da za cijeli niz vrijedi $\lim_{n\to\infty} F_n(z_n) = F(x)$. Primjenom te činjenice i teorema 4.2.2 slijedi $\lim_{n\to\infty} F(z_n) = F(x)$. Naime, vrijedi sljedeće:

$$|F(z_n) - F(x)| \leq |F(z_n) - F_n(z_n)| + |F_n(z_n) - F(x)|.$$

Funkcija F je neprekidna i strogo rastuća pa postoji inverzna funkcija od F, $F^{-1} : (0,1) \to \mathbb{R}$, koja je neprekidna. Iz neprekidnosti funkcije F^{-1} slijedi

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} F^{-1}(F(z_n)) = F^{-1}(F(x)) = x = F^{-1}(\alpha).$$

Time je dokazana tvrdnja leme. $\square$

U nastavku slijedi dokaz teorema 4.2.1

Dokaz.

(i) Označimo sa F_n funkciju distribucije od $\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n)$ uvjetno na $\mathbf{X}_n = \mathbf{x}_n$. Iz teorema 4.1.8 i 4.2.2 slijedi:

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \to 0 \quad P_{\theta_0}^{\infty} - g.s.$$

gdje je Φ funkcija distribucije slučajne varijable $Y \sim N(0, \mathcal{F}(\theta_0)^{-1})$. Koristeći Lemu 4.2.3 za $\alpha = \frac{1}{2}$ dobivamo:

$$|F_n^{-1}(\frac{1}{2}) - \Phi^{-1}(\frac{1}{2})| \to 0.$$

Kako je $\Phi^{-1}(\frac{1}{2}) = 0$, slijedi

$$|F_n^{-1}(\frac{1}{2})| \to 0 \quad P_{\theta_0}^{\infty} - g.s. \tag{4.25}$$

Općenito za slučajno varijablu X je $F_X^{-1}(\frac{1}{2})$ upravo medijan od X. Pokažimo da je medijan aposteriorne distribucije od $\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n)$ upravo $\sqrt{n}(\theta_M - \hat{\theta}_n)$.

$$P(\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n) \leq \sqrt{n}(\theta_M - \hat{\theta}_n) \mid \mathbf{X}_n) = P(\boldsymbol{\theta} \leq \theta_M \mid \mathbf{X}_n) \geq \frac{1}{2}$$

$$P(\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n) \geq \sqrt{n}(\theta_M - \hat{\theta}_n) \mid \mathbf{X}_n) = P(\boldsymbol{\theta} \geq \theta_M \mid \mathbf{X}_n) \geq \frac{1}{2}.$$

pa iz (4.25) imamo

$$\sqrt{n}(\theta_M - \hat{\theta}_n) \to 0 \quad P_{\theta_0}^{\infty} - g.s.$$

(ii) Da bi dokazali (ii) dovoljno je pokazati da vrijedi

$$\int t\, f_n(t \mid x_1, x_2, \ldots, x_n)\; dt \xrightarrow{g.s} 0. \tag{4.26}$$

Naime, vrijedi sljedeće:

$$E(\theta \,|\, \mathbf{x}_n) = \int \theta\, \pi(\theta \mid \mathbf{x}_n)\, d\theta \;\; = \begin{bmatrix} \theta = \hat{\theta}_n + t/\sqrt{n} \\ \sqrt{n}\, d\theta = dt \end{bmatrix}$$

$$= \frac{1}{\sqrt{n}} \int (\hat{\theta}_n + t/\sqrt{n})\; \pi(\hat{\theta}_n + t/\sqrt{n} \mid \mathbf{x}_n)\, dt. \tag{4.27}$$

Raspišimo $\pi(\hat{\theta}_n + t/\sqrt{n} \mid \mathbf{x}_n)$:

$$\pi(\hat{\theta}_n + t/\sqrt{n} \mid \mathbf{x}_n) = \frac{L_n(\hat{\theta}_n + \frac{t}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{t}{n})}{\int L_n(\theta)\pi(\theta)\, d\theta} = \left[\begin{smallmatrix} \theta=\hat{\theta}_n+\nu/\sqrt{n} \\ \sqrt{n}\, d\theta = d\nu \end{smallmatrix} \right]$$

$$= \frac{L_n(\hat{\theta}_n + \frac{t}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{t}{n})}{\frac{1}{\sqrt{n}} \int L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{\nu}{n})\, d\nu}. \tag{4.28}$$

Dakle:

$$E(\theta \mid \mathbf{x}_n) = \frac{1}{\sqrt{n}} \int (\hat{\theta}_n + t/\sqrt{n}) \frac{L_n(\hat{\theta}_n + \frac{t}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{t}{n})}{\frac{1}{\sqrt{n}} \int L_n(\hat{\theta}_n + \frac{\nu}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{\nu}{n})\, d\nu}\, dt$$

$$= \int (\hat{\theta}_n + t/\sqrt{n})\, f_n(t \mid x_1, x_2, \ldots, x_n)\, dt.$$

Iz toga, pod pretpostavkom da vrijedi (4.26), slijedi:

$$\sqrt{n}(E(\theta|\mathbf{x}_n) - \hat{\theta}_n) = \int t\, f_n(t \mid x_1, x_2, \ldots, x_n)\, dt \xrightarrow{g.s.} 0. \tag{4.29}$$

Dokažimo sada (4.26). Neka su $M \in \mathbb{N}$, $\delta, \epsilon > 0$ proizvoljni. Rastavimo $\int t\, f_n(t \mid \mathbf{x}_n)\, dt$ na sljedeći način:

$$\int t\, f_n(t \mid \mathbf{x}_n)\, dt = \int\limits_{|t|\le M} t\, f_n(t \mid \mathbf{x}_n)\, dt + \int\limits_{M<|t|\le\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\, dt + \int\limits_{|t|>\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\, dt. \tag{4.30}$$

Za prvi sumand u (4.30) vrijedi sljedeće:

$$\left| \int_{|t|\le M} t\, f_n(t \mid \mathbf{x}_n)\, dt \right| \le \left| \int_{|t|\le M} t\,(f_n(t \mid \mathbf{x}_n) - \phi(t))\, dt \right| + \left| \int_{|t|\le M} t\,\phi(t)\, dt \right|$$

$$= \Big(t \mapsto t\phi(t)) \text{ je neparna funkcija}\Big)$$

$$\le \int_{|t|\le M} |t|\, |f_n(t \mid \mathbf{x}_n) - \phi(t)|\, dt + 0$$

$$\le M \int_{-\infty}^{\infty} |f_n(t \mid \mathbf{x}_n) - \phi(t)|\, dt. \tag{4.31}$$

Po teoremu 4.1.8 izraz (4.31) teži ka nuli $P_{\theta_0}^{\infty} - g.s.$ Prema tome, postoji $n_0 \in \mathbb{N}$ takav da za je za svaki $n \geq n_0$:

$$\left| \int_{|t| \leq M} t\, f_n(t \mid \mathbf{x}_n)\; dt \right| < \epsilon. \tag{4.32}$$

Pogledajmo treći sumand u (4.30):

$$\int_{|t|>\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\; dt = \int_{|t|>\delta\sqrt{n}} t\, \frac{L_n(\hat{\theta}_n + \frac{t}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{t}{\sqrt{n}})}{\int L_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi}\; dt = \left[\begin{smallmatrix} t=\sqrt{n}(s-\hat{\theta}_n) \\ dt=\sqrt{n}\,ds \end{smallmatrix} \right]$$

$$= \frac{1}{\int e^{l_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}}) - l_n(\hat{\theta}_n)} \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi} \int\limits_{|s-\hat{\theta}_n|>\delta} n(s-\hat{\theta}_n) e^{l_n(s) - l_n(\hat{\theta}_n)} \pi(s)\, ds$$

$$\leq \frac{n\, e^{-n\,\epsilon(\delta/2, \theta_0)}}{\int e^{l_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}}) - l_n(\hat{\theta}_n)} \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi} \left(\int\limits_{|s-\hat{\theta}_n|>\delta} s\, \pi(s)\, ds - \hat{\theta}_n \int\limits_{|s-\hat{\theta}_n|>\delta} \pi(s)\, ds \right).$$

Nejednakost u trećem retku vrijedi jer na skupu $\{s : |s - \hat{\theta}_n| > \delta\}$ za dovoljno veliki n vrijedi $e^{l_n(s) - l_n(\hat{\theta}_n)} \leq e^{-n\,\epsilon(\delta/2, \theta_0)}$ $P_{\theta_0}^{\infty} - g.s.$

Po pretpostavci je $\left| \int s\, \pi(s)\, ds \right| < \infty$. Zbog jake konzistentnosti niza $(\hat{\theta}_n)_n$ postoji $C_1 > 0$ takav da je $|\hat{\theta}_n| \leq C_1$ za svaki $n \in \mathbb{N}$. Iz teorema 4.1.8 imamo

$$\int e^{l_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}}) - l_n(\hat{\theta}_n)} \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi \to \frac{\sqrt{\mathcal{F}(\theta_0)}}{\sqrt{2\pi}}$$

pa postoji $C_2 > 0$ takav da je za dovoljno veliki n

$$\frac{1}{\int e^{l_n(\hat{\theta}_n + \frac{\xi}{\sqrt{n}}) - l_n(\hat{\theta}_n)} \pi(\hat{\theta}_n + \frac{\xi}{\sqrt{n}})\; d\xi} < C_2. \tag{4.33}$$

Konačno, za dovoljno veliki n vrijedi

$$\int_{|t|>\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\; dt \leq \frac{n\, e^{-n\epsilon(\delta/2, \theta_0)}}{C_2} \left(\int s\, \pi(s)\, ds + C_1 \right) \to 0.$$

Dakle, postoji $n_1 \in \mathbb{N}$ takav da je za svaki $n \geq n_1$

$$\left| \int_{|t|>\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\ dt \right| < \epsilon. \tag{4.34}$$

Sada su za svaki $n \geq \max\{n_0, n_1\}$ izrazi (4.32) i (4.34) strogo manji od ϵ. Preostaje nam pokazati da je $\int_{M<|t|\leq\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x_n})\ dt$ proizvoljno malen. Kao u drugom dijelu teorema 4.1.8, jer je $|t| < \delta\sqrt{n}$, koristeći relaciju (4.18) zaključujemo da je za dovoljno veliki n

$$e^{l_n(\hat{\theta}_n+\zeta/\sqrt{n})-l_n(\hat{\theta}_n)}\ \pi(\hat{\theta}_n + \zeta/\sqrt{n})\mathbb{1}_{\{M<|\zeta|\leq\delta\sqrt{n}\}} \leq 2\pi(\theta_0)e^{-\frac{\zeta^2}{4}\mathcal{F}(\theta_0)}\ \mathbb{1}_{\{M<|\zeta|\}}(\zeta).$$

Uz relaciju (4.33) imamo

$$\int_{M<|t|\leq\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\ dt \leq \int\ 2\, C_2\, \pi(\theta_0)\, t\, e^{-\frac{t^2}{4}\mathcal{F}(\theta_0)}\ \mathbb{1}_{\{M<|t|\}}(t)\, dt. \tag{4.35}$$

Funkcija unutar integrala s desne strane nejednakosti je integrabilna i $(\{M < |\xi|\})_M$ je padajući niz događaja pa postoji M takav da je desna strana u (4.35) strogo manja od ϵ. Konačno, za $n \geq \max\{n_0, n_1\}$ vrijedi

$$\begin{aligned}\int t\, f_n(t \mid \mathbf{x}_n)\ dt &= \int_{|t|\leq M} t\, f_n(t \mid \mathbf{x}_n)\ dt + \int_{M<|t|\leq\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\ dt + \int_{|t|>\delta\sqrt{n}} t\, f_n(t \mid \mathbf{x}_n)\ dt \\ &< 3\epsilon.\end{aligned}$$

□

Teoremom 4.2.1 smo zapravo pokazali da uz uvjete teorema 4.1.4 i 4.1.8 jaka konzistentnost procjenitelja maksimalne vjerodostojnosti ekvivalentna jakoj konzistentnosti Bayesovih procjenitelja obzirom na kvadratnu i apsolutnu funkciju gubitka. Dodatno, taj rezultat tada povlači i asimptotsku normalnost Bayesovih procjenitelja. O tome nam govori sljedeći teorem.

Teorem 4.2.4.
Neka su zadovoljene pretpostavke kao u teoremu 4.2.1. Tada je:

(i) $\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \theta_0) \xrightarrow{\mathcal{D}} N(0, \mathcal{F}(\theta_0)^{-1})$,

(ii) $\sqrt{n}\,(\theta_M - \theta_0) \xrightarrow{\mathcal{D}} N(0, \mathcal{F}(\theta_0)^{-1})$.

Dokaz.
Obje tvrdnje slijede iz prethodnog teorema. Naime,

$$\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \theta_0) = \sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \hat{\theta}_n) + \sqrt{n}\,(\hat{\theta}_n - \theta_0).$$

Po teoremu 4.2.1 je $\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \hat{\theta}_n) \xrightarrow{g.s.} 0$, a po teoremu 4.1.4 $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, \mathcal{F}(\theta_0)^{-1})$. Primjenom teorema 4.1.1 na nizove $(\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \hat{\theta}_n))_n$ i $(\sqrt{n}\,(\hat{\theta}_n - \theta_0))_n$ slijedi tvrdnja. Analogno se pokazuje tvrdnja (*ii*). □

Primjer 4.2.5.
Neka su $X_1, X_2, \dots X_n$ nezavisne i Bernoullijeve slučajne varijable sa parametrom θ te neka je apriorna distribucija $\Pi \sim Beta(\alpha, \beta)$. U primjeru 2.0.5 *smo pokazali da je Bayesov procjenitelj obzirom na kvadratnu funkciju gubitka jednak*

$$E(\boldsymbol{\theta} \mid X_n) = \frac{\alpha}{\alpha + \beta + n} + \frac{\sum_{i=1}^{n} X_i}{\alpha + \beta + n}.$$

Pokazuje se da je Fisherova informacija jednaka $\mathcal{F}(\theta_0) = \frac{1}{\theta(1-\theta)} > 0$. Opet, uvjeti teorema 4.1.4 su zadovoljeni i apriorna gustoća je strogo pozitivna na $\Theta = (0, 1)$. Pokažimo da je zadovoljen uvjet (5) *teorema 4.1.8:*

$$\frac{1}{n}(l_n(t) - l_n(\theta)) \to E_\theta \log \frac{f(X|t)}{f(X|\theta)} \quad P_\theta^\infty - g.s.$$

$$\log \frac{f(X|t)}{f(X|\theta)} = X \log t + (1 - X) \log(1 - t) - X \log \theta - (1 - X) \log(1 - \theta).$$

Sada imamo:

$$E_\theta \log \frac{f(X|t)}{f(X|\theta)} = \theta(\log t - \log \theta) + (1 - \theta)(\log(1 - t) - \log(1 - \theta)).$$

Lako se pokazuje da gornji izraz kao funkcija po t poprima jedinstveni globalni maksimum u $t = \theta$ te je jednak 0. Sada se kao u primjeru 4.1.9 *pokazuje da postoji $\epsilon(\delta, \theta) > 0$ takav da vrijedi:*

$$\sup_{|t-\theta|>\delta} \frac{1}{n}\Big(l_n(t) - l_n(\theta)\Big) \le \frac{-\epsilon(\delta, \theta)}{2} \quad P_\theta^\infty - g.s.$$

Sada primjenom teorema 4.2.4 *zaključujemo da je $E(\boldsymbol{\theta}|\mathbf{X}_n)$ asimptotski normalan, odnosno*

$$\sqrt{n}\,(E(\boldsymbol{\theta} \mid \mathbf{X_n}) - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1 - \theta)).$$

Svi asimptotski rezultati do sada iskazani su vrijedili u regularnom statističkom modelu, to jest modeleu koji zadovoljava uvjete teorema 4.1.4. Na sljedećem primjeru ćemo pokazati da je varijanta teorema 4.1.8 moguća i za neregularne modele, ali uz neke modifikacije.

Primjer 4.2.6.
Neka je $X_1, X_2, \ldots X_n$ slučajan uzorak iz uniformne distribucije na intervalu $(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Pretpostavimo da je apriorna gustoća parametra $\boldsymbol{\theta}$, $\pi(\theta)$, neprekidna, omeđena i pozitivna na Θ. Neka je θ_0 stvarna vrijednost parametra i $\hat{\theta}_n := X_{(n)}$ procjenitelj maksimalne vjerodostojnosti za θ. Pokažimo da aposteriorna gustoća od $\nu = n(\boldsymbol{\theta} - \hat{\theta}_n)$ konvergira ka gustoći eksponencijalne distribucije sa očekivanjem θ_0.
Analognim izvodom kao za $\sqrt{n}(\boldsymbol{\theta} - \hat{\theta}_n)$ se pokazuje da je gustoća od $n(\boldsymbol{\theta} - \hat{\theta}_n)$:

$$
\begin{aligned}
f_n(\nu|x_1, x_2, \ldots, x_n) &= \frac{L_n(\hat{\theta}_n + \frac{\nu}{n})\pi(\hat{\theta}_n + \frac{\nu}{n})}{\int_\Theta L_n(\hat{\theta}_n + \frac{\xi}{n})\pi(\hat{\theta}_n + \frac{\xi}{n})\ d\xi} \\
&= \frac{(\hat{\theta}_n + \frac{\nu}{n})^{-n}\ \pi(\hat{\theta}_n + \frac{\nu}{n})}{\int_0^\infty (\hat{\theta}_n + \frac{\xi}{n})^{-n}\ \pi(\hat{\theta}_n + \frac{\xi}{n})\, d\xi} \\
&= \frac{\hat{\theta}_n^{-n}(1 + \frac{\nu/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\nu}{n})}{\hat{\theta}_n^{-n}\int_0^\infty (1 + \frac{\xi/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\xi}{n})\, d\xi} \\
&= \frac{(1 + \frac{\nu/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\nu}{n})}{\int_0^\infty (1 + \frac{\xi/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\xi}{n})\, d\xi}. \qquad (4.36)
\end{aligned}
$$

U primjeru 3.0.16 *smo pokazali da je $\hat{\theta}_n$ jako konzistentan procjenitelj za θ_0. Pokažimo sada da vrijedi*

$$f_n(\nu|x_1, x_2, \ldots, x_n) \xrightarrow{g.s.} \frac{1}{\theta_0} e^{-\frac{\nu}{\theta_0}}\ \mathbb{1}_{(0,\infty)}(\nu).$$

Zbog jake konzistentnosti niza $(\hat{\theta}_n)_n$ i neprekidnosti funkcije $\pi(\theta)$ u θ_0 vrijedi

$$\pi(\hat{\theta}_n + \frac{\nu}{n}) \to \pi(\theta_0). \qquad (4.37)$$

Isto tako, vrijedi

$$(1 + \frac{\nu/\hat{\theta}_n}{n})^{-n} \to e^{-\frac{\nu}{\theta_0}}. \qquad (4.38)$$

Pokažimo da je funkcija $\nu \mapsto (1 + \frac{\nu/\hat{\theta}_n}{n})^{-n}\pi(\hat{\theta}_n + \frac{\nu}{n})$ *dominirana integrabilnom funkcijom. Zbog omeđenosti funkcije* $\pi(\theta)$ *postoje konstante* $M, m > 0$ *tako da je* $m \leq \pi(\theta) \leq M$ *za svaki* $\theta \in \Theta$. *Iz toga imamo*

$$1 \leq \frac{\pi(\theta)}{m} \Rightarrow M \leq M\frac{\pi(\theta)}{m}.$$

Konačno,

$$(1 + \frac{\nu/\hat{\theta}_n}{n})^{-n}\pi(\hat{\theta}_n + \frac{\nu}{n}) \leq 1 \cdot M \leq M\frac{\pi(\nu)}{m}.$$

Sada, primjenom teorema o dominiranoj konvergenciji na funkciju $\nu \mapsto (1+\frac{\nu/\hat{\theta}_n}{n})^{-n}\pi(\hat{\theta}_n+\frac{\nu}{n})$ *i iz relacija* (4.37) *i* (4.38) *slijedi*

$$\begin{aligned}
\lim_{n\to\infty} f_n(\nu|x_1, x_2, \ldots, x_n) &= \frac{\lim\limits_{n\to\infty}(1 + \frac{\nu/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\nu}{n})}{\lim\limits_{n\to\infty}\int_0^\infty (1 + \frac{\xi/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\xi}{n})\, d\xi} \\
&= \frac{e^{-\frac{\nu}{\theta_0}}\, \pi(\theta_0)}{\int_0^\infty \lim\limits_{n\to\infty}(1 + \frac{\xi/\hat{\theta}_n}{n})^{-n}\, \pi(\hat{\theta}_n + \frac{\xi}{n})\, d\xi} \\
&= \frac{e^{-\frac{\nu}{\theta_0}}\, \pi(\theta_0)}{\pi(\theta_0)\int_0^\infty e^{-\frac{\nu}{\theta_0}}\, d\nu} = \frac{1}{\theta_0}e^{-\frac{\nu}{\theta_0}}.
\end{aligned}$$

Bibliografija

[1] P. Baldi, L. Mazliak i P. Priouret, *Martingales and Markov Chains: Solved Exercises and Elements of Theory*, CRC Press, 2002.

[2] P.J. Bickel i K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics, Volume I, Second Edition*, CRC Press, 2015.

[3] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer New York, 2008.

[4] T.S. Ferguson, *A Course in Large Sample Theory*, Taylor & Francis, 1996.

[5] J.K. Ghosh i R.V. Ramamoorthi, *Bayesian Nonparametrics*, Springer, 2014.

[6] E.L. Lehmann i G. Casella, *Theory of Point Estimation*, Springer New York, 2006.

[7] J.W. Miller, *A detailed treatment of Doob's theorem*, (2018), `https://arxiv.org/pdf/1801.03122.pdf`, , pristupljeno: 27.08.2018.

[8] K.R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1972.

[9] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, 1992.

[10] M.J. Schervish, *Theory of Statistics*, Springer New York, 2012.

Sažetak

U ovom radu smo se bavili asimptotskom teorijom u bayesovskoj statistici. Uveli smo osnovne pojmove bayesovske statistike te dali metodu konstruiranja aposteriorne distribucije. Opisali smo Bayesove procjenitelje obzirom na kvadratnu i apsolutnu funkciju gubitka. Zatim smo uveli pojam konzistentnosti aposteriorne distribucije koji nam na neki način daje opravdanje bayesovskog pristupa sa gledišta klasične statistike. U sklopu konzistentnosti smo promatrali kako se asimptoski ponašaju dvije konzistentne aposteriorne distribucije koje su proizašle iz različitih apriornih distribucija. Pokazali smo da su, uz neke pretpostavke, asimptoski ekvivalentne. Zatim smo iskazali i dokazali Doobov teorem koji nam govori da je aposteriorna distribucija konzistentna na cijelom nosaču apriorne distribucije. Također smo opisali jednu metodu procjene proizašle iz klasične statistike: metodu maksimalne vjerodostojnosti. Pokazali smo da uz određene uvjete imamo istovremeno jaku konzistentnost procjenitelja maksimalne vjerodostojnosti i konzistentnost aposteriorne distribucije, ali smo i naveli primjer kada to ne vrijedi.

Nakon toga, bavili smo se asimptotskim svojstvima aposteriorne distribucije i Bayesovih procjenitelja. Teoremom Bernstein-von Mises opisali smo asimptotsko ponašanje aposteriorne distribucije u regularnim statističkim modelima, rezultatom koji je zapravo analogon teorema o asimptotskoj normalnosti procjenitelja maksimalne vjerodostojnosti. Posljedično tome, dokazali smo asimptotsku ekvivalenciju Bayesovih procjenitelja sa procjeniteljem maksimalne vjerodostojnosti, odnosno dokazali smo njihovu asimptotsku normalnost. Rezultati i razmatranja su potkrijepljeni primjerima. Na kraju, primjerom smo ilustrirali asimptotsku normalnost aposteriorne distribucije i u jednom neregularnom modelu.

Summary

In this thesis we dealt with large sample theory in Bayesian statistics. We introduced some basic concepts in Bayesian statistics and described how the posterior distribution is constructed. Then we described the Bayes estimator under the square error and absolute loss functions. Furthermore, the posterior consistency was introduced, a sort of frequentist validation of the Bayesian approach. We observed the asymptotic behaviour of two consistent posteriors that arose from different priors and showed that, under some assumptions, they are asymptotically equivalent. Next, we proved Doobs theorem that states that the posterior is consistent on the support of the prior. We also described one frequentist estimation method: maximum likelihood estimation. Then we showed that under certain conditions the strong consistency of the maximum likelihood estimator and the consistency of the posterior distribution occurs.

However, we showed in an example that the previous statement is not valid in general. Afterwards, we dealt with the asymptotic properties of the posterior distribution and Bayes estimators. We stated the Bernstein-von Mises theorem about the asymptotic normality of the posterior when the regularity conditions hold. It is an analogous result to the asymptotic normality of the maximum likelihood estimator. Moreover, the asymptotic equivalence of the Bayes estimators and the maximum likelihood estimator was proven. Consequently, the asymptotic normality of the Bayes estimator was obtained. We provided examples that illustrate the main results described in the thesis. Lastly, we gave an example of a result in asymptotic normality of the posterior when the distribution is nonregular.

Životopis

Rođen sam 11. srpnja 1994. godine u Rijeci. Osnovnu školu sam završio u Karlovcu. U istome gradu 2013. godine završavam gimnaziju, matematički smjer. Potom upisujem preddiplomski studij Matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu. Nakon završetka preddiplomskog studija, 2016.godine, upisujem diplomski studij Matematička statistika na istome fakultetu.