Univerzitet u Beogradu

Matematički fakultet

Neki primeri primene funkcije verodostojnosti u matematičkoj statistici

Master rad

Student:

Ana Stojadinović

Mentor:

Vesna Jevremović

Beograd,2012.

1. Uvod

Pri proučavanju obeležja na osnovu uzorka javljaju se dva osnovna zadatka. Prvi je naći raspodelu obeležja na populaciji, a drugi, ako je tip raspodele poznat, odrediti parametre raspodele. U teoriji verovatnoća su se, pored ostalog, proučavali različiti modeli za raspodele slučajnih obeležja i uslovi pod kojima ti modeli važe. Tako smo proučavali model binomne, Puasonove[1], normalne i drugih raspodela, koje smo reprezentovali odgovarajućim funkcijama raspodele. Kod svih njih se, zbog određenih početnih pretpostavki, pojavljuju neki empirijski parametri, od kojih zavisi i sama funkcija raspodele.

Nepoznate parametre u raspodeli posmatranog obeležja ocenjujemo na osnovu uzorka, računajući realizovane vrednosti pogodno odabranih statistika.

Pošto je realizovana vrednost statistike neki realan broj, tj. neka tačka na realnoj pravoj, to ovakve ocene parametara nazivamo tačkaste ocene.

Neka je $(X_1, \dots, X_n)$ prost slučajan uzorak obima n za posmatrano obeležje X. Statistika $Y = f(X_1, \dots, X_n)$ je slučajna promenljiva koja implicitno zavisi od parametra u raspodeli obeležja X. Ako se statistikom Y ocenjuje parametar θ, tada se, često, statistika Y naziva ocena parametra θ i označava sa $\hat{\theta}$.

Za ocenjivanje raznih parametara u raspodeli obeležja biramo različite statistike. Moguće je takođe koristiti različite statistike za ocenu jednog parametra. Statistike koje koristimo moraju da imaju određene osobine kojima se opravdava njihova primena u ocenjivanju parametara.

U ovom radu biće opisana jedna od metoda za dobijanje tačkastih ocena - metoda maksimalne verodostojnosti, potom ćemo govoriti o Bajesovim ocenama, bitnom delu teorije odlučivanja, zatim ćemo se osvrnuti na numeričke metode za dobijanje ovih ocena i na kraju, u praktičnom delu, prikazaćemo primenu navedene teorije u statističkom softveru R.

[1] Siméon Denis Poisson (1781 – 1840), francuski matematičar i fizičar.

2. Definicija funkcije verodostojnosti

Neka je $\boldsymbol{X}$ =($X_1, \dots, X_n$) prost slučajan uzorak obima n za posmatrano obeležje X u čijoj raspodeli figuriše nepoznati parametar θ. Podsetimo se da je reč o n-dimenzionalnoj slučajnoj promenljivoj ($X_1, \dots, X_n$) pri čemu su slučajne promenljive $X_1, \dots, X_n$ nezavisne i sve imaju istu raspodelu kao posmatrano obeležje X. Neka je gustina raspodele obeležja X, ako je ono neprekidnog tipa:

$$g(x;\theta), x \in R, \theta \in \Theta,$$

odnosno neka je zakon raspodele obeležja X, ako je ono diskretnog tipa:

$$p(x;\theta), x \in R, \theta \in \Theta.$$

Ovde Θ označava skup dopustivih (mogućih) vrednosti parametra, tzv. parametarski prostor.

Funkcija verodostojnosti posmatrana kao slučajna veličina je, u neprekidnom slučaju

$$L(\boldsymbol{X};\theta) = \prod_{j=1}^{n} g(X_j;\theta),$$

a u diskretnom slučaju

$$L(\boldsymbol{X};\theta) = \prod_{j=1}^{n} p(X_j;\theta).$$

Radi jednostavnosti, u daljem tekstu pretpostavimo da je prost slučajan uzorak $\boldsymbol{X}$ =($X_1, \dots, X_n$) dat svojom zajedničkom gustinom raspodele[2], u slučaju obeležja neprekidnog tipa, odnosno zajedničkim zakonom raspodele[3], u slučaju obeležja diskretnog tipa, u oba slučaja koristimo oznaku $f(\boldsymbol{x};\theta), \boldsymbol{x} \in R, \theta \in \Theta$. Neka je data realizovana vrednost uzorka $\boldsymbol{x}$, tada definišemo funkciju verodostojnosti kao

$$L(\theta) = f(\boldsymbol{x};\theta);$$

za svaki mogući realizovan uzorak $\boldsymbol{x}$ =($x_1, \dots, x_n$). Funkcija verodostojnosti je sada realna funkcija definisana na parametarskom prostoru Θ.

[2] Pretpostavimo da su $X_1, \dots, X_n$ slučajne promenljive neprekidnog tipa definisane na istom prostoru verovatnoća i neka je $P[X_1 \le x_1, \dots, X_n \le x_n] = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n$ za sve $x_1, \dots, x_n$. Tada je $f(x_1, \dots, x_n)$ zajednička funkcija gustine za $(X_1, \dots, X_n)$.

[3] Pretpostavimo da su $X_1, \dots, X_n$ slučajne promenljive diskretnog tipa definisane na istom prostoru verovatnoća. Tada je sa $f(x_1, \dots, x_n) = P[X_1 = x_1, \dots, X_n = x_n]$ definisan zajednički zakon raspodele za $(X_1, \dots, X_n)$.

3. Dovoljne statistike

U ovom odeljku ćemo se baviti jednim od kriterijuma za izbor statistike. Za redukciju uzoračkih podataka, koja se sastoji u odbacivanju onog dela podataka koji ne sadrži informaciju o nepoznatoj raspodeli, prema tome, nekorisna je za bilo koji problem odlučivanja o toj raspodeli, R.A.Fišer[4] je uveo pojam *dovoljne statistike.* Fišer je razmatrao ocenjivanje disperzije σ^2 za normalnu raspodelu zasnovanu na nezavisnim i jednako raspodeljenim slučajnim veličinama $X_1, \ldots, X_n$. On je razmatrao ocenjivanje parametra σ^2 na osnovu dve statistike

$$T_1 = \sum_{i=1}^{n} |X_i - \bar{X}| \qquad i \qquad T_2 = \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Fišer je pokazao da raspodela za T_1 pri uslovu da je statistika T_2 fiksna, ne zavisi od parametra σ, dok raspodela za T_2 pri uslovu da je statistika T_1 fiksna, zavisi od parametra σ. Zaključio je da statistika T_2 sadrži sve informacije o parametru σ^2 i da bilo koja ocena za σ^2 treba da bude proporcionalna T_2, tj. bilo koja ocena za σ^2 proporcionalna statistici T_1, može biti poboljšana korišćenjem informacije o T_2, dok T_2 ne može biti poboljšana korišćenjem T_1.

Sada ćemo generalizovati ovo. Pretpostavimo da $\boldsymbol{X} = (X_1, \ldots, X_n) \sim F_\theta$ za neko $\theta \in \Theta$ i neka je $T = T(\boldsymbol{X})$ statistika. Za svako t u intervalu od T, definišimo skupove

$$A_t = \{\boldsymbol{x} : T(\boldsymbol{x}) = t\}.$$

Sada posmatrajmo raspodelu za $\boldsymbol{X}$ na skupu A_t, tj. uslovnu raspodelu za $\boldsymbol{X}$ pri $T = t$. Ako ova uslovna raspodela ne zavisi od θ, onda $\boldsymbol{X}$ ne sadrži informaciju o θ na skupu A_t, tj. $\boldsymbol{X}$ je drugorazredna statistika[5] na A_t. Ako je ovo tačno za svako t u intervalu statistike T, sledi da T sadrži iste informacije o θ kao i $\boldsymbol{X}$. Dakle, statistika T je dovoljna statistika za θ. Statistika se smatra dovoljnom ako sadrži sve informacije o parametru.

Definicija 1. Statistika T je dovoljna statistika za parametar θ ako uslovna raspodela uzorka $\boldsymbol{X} = (X_1, \ldots, X_n)$, pri fiksiranoj vrednosti statistike T, ne zavisi od θ.

Dovoljne statistike nisu jedinstvene. Iz definicije dovoljnosti sledi da ako je g jedan-jedan funkcija, onda je i $g(T)$ dovoljna statistika. Ovo naglašava činjenicu da dovoljne statistike same po sebi nisu bitne, već je bitna podela uzoračkog prostora indukovana datom statistikom.

Teorema 1. Ako je T dovoljna statistika za parametar θ, a v bijektivna funkcija, tada je i $v(T)$ dovoljna statistika za θ.

Ne dopušta svaka familija uzoračkih raspodela redukciju bez gubitka informacija o parametru θ, te se kao važan problem postavlja utvrđivanje egzistencije

[4] Sir Ronald Aylmer Fisher (1890 – 1962), engleski statističar i genetičar.
[5] Statistika čija raspodela ne zavisi od parametra naziva se drugorazredna statistika.

dovoljne statistike za θ. Rešavanje tog problema korišćenjem samo prethodne definicije povezano je sa teškoćama zbog određivanja uslovnih raspodela. Jednostavan kriterijum za utvrđivanje egzistencije dovoljne statistike, koji ujedno daje i način za njeno određivanje, jeste sledeća *teorema o faktorizaciji Fišer-Nojmana*[6]*:*

Teorema 2. Statistika T je dovoljna statistika za parametar θ u raspodeli obeležja X ako i samo ako se funkcija verodostojnosti može predstaviti u obliku

$$L(\theta) = f(\boldsymbol{x};\theta) = g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})$$

gde je g funkcija koja zavisi od $\boldsymbol{x} = (x_1, \dots, x_n)$ samo preko T i zavisi od θ, dok funkcija h ne zavisi od θ.

Primer 1. Neka je $\boldsymbol{X} = (X_1, \dots, X_n)$ uzorak iz Puasonove raspodele sa parametrom λ. Tada je funkcija verodostojnosti oblika:

$$L(\boldsymbol{X};\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{j=1}^{n} x_j!}$$

Koristeći teoremu o faktorizaciji zaključujemo da je funkcija $h(\boldsymbol{x}) = \frac{1}{\prod_{j=1}^{n} x_j!}$, a funkcija $g(T(\boldsymbol{x});\lambda) = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}$, pa odatle proizilazi da je statistika $T = \sum_{i=1}^{n} x_i$ dovoljna za parametar λ. ◊

Primer 2. Neka je $\boldsymbol{X} = (X_1, \dots, X_n)$ uzorak iz normalne raspodele $N(\theta, 1)$. Tada funkcija verodostojnosti ima oblik:

$$L(\boldsymbol{X};\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2}} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2 - \frac{n\theta^2}{2}} e^{\theta \sum_{i=1}^{n} x_i} =$$

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

Koristeći teoremu o faktorizaciji zaključujemo da je funkcija $h(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}$, a funkcija $g(T(\boldsymbol{x});\theta) = e^{\theta \sum_{i=1}^{n} x_i - \frac{n\theta^2}{2}}$, pa odatle proizilazi da je statistika $T = \sum_{i=1}^{n} x_i$ dovoljna za parametar θ. ◊

Možemo primetiti da se za dovoljnu statistiku našeg parametra u primeru 2., pored uzoračkog totala, sasvim legitimno može uzeti i uzoračka sredina $\bar{X}_n = \sum_{i=1}^{n} x_i / n$, što opravdava zaključak da dovoljna statistika nije jedinstvena i ilustruje zaključak Teoreme 1.

Iz definicije dovoljnosti sledi da je uzorak $\boldsymbol{X}$ uvek dovoljna statistika. Tako dovoljnost ne bi bio posebno koristan koncept ukoliko ne bismo mogli da nađemo dovoljnu statistiku koja zaista redukuje podatke, pa tako pravi problem zapravo leži u

[6] Jerzy Neyman (1894 – 1981), američki matematičar poljskog porekla.

određivanju da li izabrana dovoljna statistika predstavlja najbolju moguću redukciju podataka.

Postoje dva shvatanja o tome šta se podrazumeva pod „najbolje moguće“ smanjenje podataka. Prvi koncept je minimalna dovoljna statistika. Dovoljna statistika T je minimalna dovoljna statistika ako za svaku drugu dovoljnu statistiku S postoji funkcija g takva da je $T = g(S)$. Tako je minimalna dovoljna statistika zapravo dovoljna statistika koja predstavlja maksimalnu redukciju podataka, tako da ti podaci sadrže onoliko informacija o nepoznatom parametru koliko i sam uzorak.

Teorema 3. Neka su $\boldsymbol{x}$ i $\boldsymbol{y}$ uzorci istog obima. T je minimalna dovoljna statistika ako

$$\frac{L(\boldsymbol{x};\theta)}{L(\boldsymbol{y};\theta)} \text{ nije funkcija od } \theta <=> T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

Drugi koncept je kompletnost. Ako $\boldsymbol{X} = (X_1, \dots, X_n) \sim F_\theta$ onda je statistika $T = T(\boldsymbol{X})$ kompletna ako $E_\theta(g(T)) = 0$ za svako $\theta \in \Theta$. Podrazumeva se da $P_\theta(g(T) = 0) = 1$ za svako $\theta \in \Theta$. Konkretno, ako je T dovoljna statistika onda je $g(T)$ drugorazredna statistika za θ jedino ako je $g(T)$ konstanta; dakle kompletna statistika ne sadrži drugorazredne informacije.

Koliko je u praksi bitan pojam dovoljnosti? Sva dosadašnja razmatranja ukazuju na to da bi bilo koja statistička procedura trebala da zavisi jedino od minimalne dovoljne statistike. Zapravo, optimalne statističke procedure (tačkaste ocene, testovi hipoteza, itd.) skoro uvek zavise od minimalne dovoljne statistike. Dalje u radu biće pokazano da su ocene metodom maksimalne verodostojnosti funkcije dovoljne statistike. Ipak, statistički modeli služe samo kao aproksimacija stvarnih podataka, pa procedure koje se nameću kao optimalne u praksi se mogu pokazati nezadovoljavajućim. Zato treba jako fleksibilno razvijati statističke modele.

Uvedimo sada pojam *eksponencijalne familije raspodela*, kojom se obuhvataju mnoge poznate raspodele.

Definicija 2. Za familiju raspodela se kaže da je eksponencijalnog tipa ako je njena gustina oblika:

$$f(\boldsymbol{x};\theta) = exp\{A(\theta)K(\boldsymbol{x}) + S(\boldsymbol{x}) + q(\theta)\}.$$

Neka je $\boldsymbol{X} = (X_1, \dots, X_n)$ uzorak iz populacije čije obeležje ima gustinu koja pripada slučaju eksponencijalne familije. Tada je uzoračka gustina:

$$\prod_{i=1}^{n} f(x_i;\theta) = exp\left\{A(\theta)\sum_{i=1}^{n} K(x_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta)\right\}$$

što se može predstaviti u obliku $C(\theta)e^{A(\theta)\sum_{i=1}^{n}K(x_i)}h(\boldsymbol{x})$, gde je $h(\boldsymbol{x}) = e^{\sum_{i=1}^{n}S(x_i)}$, a $g = C(\theta)e^{A(\theta)\sum_{i=1}^{n}K(x_i)}$, pa je na osnovu teoreme o faktorizaciji $T = \sum_{i=1}^{n}K(x_i)$ dovoljna statistika za parametar θ.

Primer 3. Neka je $\boldsymbol{X}=(X_1,\dots,X_n)$ uzorak iz normalne raspodele $N(m,\sigma^2)$. Normalna raspodela je iz familije eksponencijalnih raspodela. Ukoliko je poznat parametar σ^2 onda je raspodela iz eksponencijalne familije raspodela po parametru m i obrnuto.Dakle, uzoračka gustina ima oblik:

$$\prod_{i=1}^{n} f(x_i;\theta) =$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{m}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{nm^2}{2\sigma^2}}$$

pa zaključujemo da je par statistika $(\sum_{i=1}^{n}x_i, \sum_{i=1}^{n}x_i^2)$ dovoljna statistika za $\theta = (m,\sigma^2)$. Ovde treba napomenuti da je za višedimenzionalni parametar netačno zaključivanje da ako je par (u,v) dovoljna statistika za (θ_1,θ_2) sledi da je u dovoljna statistika za θ_1, a v dovoljna statistika za θ_2. ◊

4. Funkcija verodostojnosti i ocenjivanje parametara

U odeljku 2. definisali smo funkciju verodostojnosti, a sada ćemo videti kako se pomoću nje definiše *ocena po metodi maksimalne verodostijnosti.*

Definicija 3. Pretpostavimo da za realizovani uzorak $\boldsymbol{x}=(x_1,\dots,x_n)$, funkcija verodostojnosti $L(\theta)$ postiže supremum u $\theta = S(\boldsymbol{x})$:

$$\sup_{\theta\in\Theta} L(\theta) = L(S(\boldsymbol{x}))$$

Tada statistiku $\hat{\theta} = S(\boldsymbol{x})$ zovemo ocena po metodi maksimalne verodostojnosti za parametar θ (u daljem tekstu za ovu ocenu koristićemo skraćenicu MLE[7]).

U praksi obično biramo model raspodele za obeležje X tako da postoji $\sup_{\theta\in\Theta} L(\theta)$.

Ako je $T = T(\boldsymbol{X})$ dovoljna statistika za θ, onda iz Teoreme o faktorizaciji sledi da

$$L(\theta) \propto g(T(\boldsymbol{x});\theta).^{8}$$

[7] Skraćenica od maximum likelihood estimator.
[8] Ako je funkcija A proporcionalna funkciji B koristimo zapis A∝B.

Iz ovoga dalje sledi da ako je MLE $\hat{\theta}$ jedinstvena, onda je $\hat{\theta}$ funkcija dovoljne statistike T.

Još jedno atraktivno svojstvo MLE je invarijantnost. Na primer, ako $\phi = g(\theta)$, gde je g monotona funkcija (ili opštije, jedan-jedan funkcija) i $\hat{\theta}$ je MLE za θ, onda je $g(\hat{\theta})$ MLE za ϕ. Uobičajeno je proširiti ovo svojstvo invarijantnosti na proizvoljne funkcije, dakle ako $\phi = g(\theta)$ onda obično kažemo da $\hat{\phi} = g(\theta)$ je MLE za ϕ.

Suštinski postoje dve različite metode za nalaženje MLE:

- Direktno proučavamo $L(\theta)$ da bismo ustanovili za koju vrednost θ funkcija verodostojnosti $L(\theta)$ postiže supremum. U nekim slučajevima to će ujedno biti maksimalna vrednost funkcije verodostojnosti.Ova metoda je posebno korisna kada opseg podataka zavisi od parametra.
- Funkcije verodostojnosti - ako opseg podataka ne zavisi od parametra, parametarski prostor Θ je otvoren skup, i funkcija verodostojnosti je diferencijabilna po $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ na prostoru Θ, pa tada MLE $\hat{\boldsymbol{\theta}}$ zadovoljava sledeće jednačine
$$\frac{\partial}{\partial \theta_k} \ln L(\hat{\boldsymbol{\theta}}) = 0 \quad za\; k = 1, \dots, p$$
Ove jednačine se zovu jednačine verodostojnosti, a $lnL(\theta)$ se zove log-funkcija verodostojnosti.

Za ogromnu većinu statističkih modela možemo koristiti jednačine verodostojnosti kako bismo odredili MLE. Ako $\hat{\theta}$ maksimizira $L(\theta)$, takođe maksimizira i $lnL(\theta)$. $L(\theta)$ je izražena kao proizvod, pa $lnL(\theta)$ postaje suma, što je lakše za diferenciranje. Jednačine verodostojnosti mogu imati više rešenja, pa je važno proveriti da li izabrano rešenje zaista maksimizira funkciju verodostojnosti, npr. ako je funkcija verodostojnosti dvaput diferencijabilna, tada ispitujemo znak njenog drugog izvoda u tačkama u kojima je prvi izvod jednak nuli ili u slučaju višedimenzionalnog parametra, ako funkcija verodostojnosti ima neprekidne parcijalne izvode drugog reda, onda formiramo totalni diferencijal drugog reda te funkcije i utvrđujemo njegov znak u tačkama koje smo dobili kao moguće ocene. Ako parametarski prostor Θ nije otvoren skup treba proveriti da maksimum ne bude na granicama parametarskog prostora.

Primer 4. Pretpostavimo da su slučajne promenljive $X_1, \dots, X_n$ nezavisne i sve imaju uniformnu raspodelu na intervalu $[0, \theta]$ za neko θ>0. Tada je funkcija verodostojnosti
$$L(\theta) = \theta^{-n} I(0 < x_1, \dots, x_n < \theta) = \theta^{-n} I(\theta \geq \max(x_1, \dots, x_n)).$$
Tako ako $\theta < \max(x_1, \dots, x_n)$, $L(\theta) = 0$ dok je $L(\theta)$ opadajuća funkcija od θ za $\theta \geq \max(x_1, \dots, x_n)$. Zato $L(\theta)$ dostiže svoj maksimum u $\theta = \max(x_1, \dots, x_n)$ pa

je

$$\widehat{\theta} = X_{(n)} = \max(X_1, \dots, X_n)$$

MLE za θ. ◊

Primer 5. Neka je $\boldsymbol{X} = (X_1, \dots, X_n)$ uzorak iz Puasonove raspodele sa parametrom $\lambda > 0$. Tada je funkcija verodostojnosti oblika:

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{j=1}^{n} x_j!}$$

a log-funkcija verodostojnosti

$$lnL(\lambda) = -n\lambda + ln\lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!).$$

Uz pretpostavku da je $\sum_{i=1}^{n} x_i > 0$ i diferencirajuci po λ, nalazimo da je

$$\frac{d}{d\lambda} lnL(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i$$

a izjednačavajući prvi izvod sa 0 zaključujemo da je $\hat{\lambda} = \bar{x}$. Da bismo potvrdili da je ovo zaista maksimum naći ćemo drugi izvod i posmatrati njegov znak

$$\frac{d^2}{d\lambda^2} lnL(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^{n} x_i$$

Zaključujemo da je drugi izvod uvek negativan. Tako $\bar{x}$ maksimizira funkciju verodostojnosti za dati izorak $x_1, \dots, x_n$ i MLE za λ je $\hat{\lambda} = \bar{X}$ kada je $\sum_{i=1}^{n} x_i > 0$.

Interesantno je razmotriti slučaj kada je $\sum_{i=1}^{n} x_i = 0$. Onda, strogo govoreći, MLE ne postoji, jer funkcija verodostojnosti tada ima oblik $L(\lambda) = e^{-n\lambda}$ i važi $supL = 1$ kada $\lambda \to 0$, ali se ne dostiže zato što $\lambda \neq 0$. Ali verovatnoća da dođe do ovog slučaja je jako mala, jer bi sve realizovane vrednosti uzorka morale biti jednake nuli. ◊

Ocena metodom maksimalne verodostojnosti je primenljiva u ogromnom broju problema i ima veoma atraktivna svojstva optimizacije, zato se smatra „zlatnim standardom“ procedura za ocenjivanje.

5. Princip verodostojnosti

Funkcija verodostojnosti ima mnogo veće značenje u statističkom zaključivanju, o čemu će biti reči u ovom odeljku.

Pretpostavimo da možemo da biramo između dva eksperimenta za ocenjivanje parametra θ. Iz prvog eksperimenta dobijamo podatke $\boldsymbol{x}$, dok iz drugog eksperimenta dobijamo podatke $\boldsymbol{x}^*$. Princip verodostojnosti obezbeđuje jednostavan kriterijum kada treba izvesti identične zaključke za θ iz $\boldsymbol{x}$ i $\boldsymbol{x}^*$. Taj kriterijum se navodi na sledeći način:

Princip verodostojnosti. Neka je $L(\theta)$ funkcija verodostojnosti za parametar θ bazirana na uzorku $\boldsymbol{X} = \boldsymbol{x}$, a $L^*(\theta)$ funkcija verodostojnosti za θ bazirana na uzorku $\boldsymbol{X}^* = \boldsymbol{x}^*$. Ako je $L(\theta) = kL^*(\theta)$ (za neko k koje ne zavisi od θ) onda se za θ može izvući isti zaključak iz oba uzorka.

Kod tačkastih ocena, princip verodostojnosti podrazumeva da ako su $\boldsymbol{x}$ i $\boldsymbol{x}^*$ dva uzorka sa funkcijama verodostojnosti $L(\theta)$ i $L^*(\theta)$ i važi $L(\theta) = kL^*(\theta)$ (za svako θ), tada dve tačkaste ocene $T(\boldsymbol{x})$ i $T^*(\boldsymbol{x})$ treba da budu jednake. Očito je da ocena metodom maksimalne verodostojnosti zadovoljava ovaj uslov.

Treba napomenuti da se princip verodostojnosti odnosi samo na informaciju o θ koja je sadržana u uzorku, mada mi takođe možemo imati dodatnu informaciju o θ i van uzorka. Na primer, možemo izraziti naša uverenja o pravoj vrednosti parametra θ (pre posmatranja uzorka) tako što ćemo odrediti raspodelu verovatnoća nad parametarskim prostorom Θ i ova raspodela se naziva apriorna raspodela. Ovakav pristup (naziva se bajesovski[9] pristup) je često oslovljavan kao „korektan“ pristup implementacije principa verodostojnosti.

Postavlja se pitanje da li princip verodostojnosti ima smisla. Ovo pitanje pokreće mnoge kontroverze u teorijskoj statistici i ne postoji jasan odgovor. Princip verodostojnosti se može dokazati podrazumevajući druga dva, manje kontroverzna, principa, princip dovoljnosti i princip uslovljavanja. Princip dovoljnosti tvrdi da dovoljna statistika sadrži onoliko informacija o vrednosti parametra koliko i sami podaci, dok princip uslovljavanja u suštini tvrdi da nema gubitaka podataka prilikom uslovljavanja drugorazredne statistike.

Princip verodostojnosti nije univerzalno prihvaćen, štaviše, u statističkoj praksi je često kršen. Naravno, princip verodostojnosti pretpostavlja da je pravi parametarski model za podatke poznat (statistički modeli su skoro neizbežni za aproksimaciju realne slike). Ovo, naravno, ne dovodi odmah u pitanje validnost principa verodostojnosti, već samo naglašava potrebu da se svaki princip i teorija moraju pomno ispitati u praksi.

---

[9] Thomas Bayes (1702–1761), engleski matematičar.

6. Asimptotska svojstva MLE

Postavlja se pitanje pod kojim uslovima je ocena metodom maksimalne verodostojnosti postojana[10] i sa asimptotski normalnom raspodelom. Pokazaće se da je pod prilično blagim uslovima regularnosti moguće dokazati postojanost i asimptotsku normalnost za MLE realnog parametra kada je dat niz nezavisnih i jednako raspodeljenih slučajnih veličina. U mnogim slučajevima je moguće pronaći asimptotsku raspodelu niza MLE koristeći standardne tehnike.

Primer 6. Neka je $X_1, \dots, X_n$ niz nezavisnih slučajnih veličina iz geometrijske raspodele sa funkcijom gustine

$$f(x;\theta) = \theta(1-\theta)^x \quad za\ x = 0,1,2 \dots$$

Ocena metodom maksimalne verodostojnosti na osnovu uzorka $X_1, \dots, X_n$ je

$$\hat{\theta}_n = \frac{1}{\bar{X}_n + 1}$$

Koristeći Centralnu graničnu teoremu[11] dobijamo da

$$\sqrt{n}(\bar{X}_n - (\theta^{-1} - 1)) \underset{d}{\rightarrow} N(0, \theta^{-2}(1-\theta)).$$

Tako dobijamo, primenjujući delta metod[12] sa izabranim funkcijama $g(x) = 1/(1+x)$ i $g'(x) = -1/(1+x)^2$:

$$\sqrt{n}\big(\hat{\theta}_n - \theta\big) = \sqrt{n}(g(\bar{X}_n) - g(\theta^{-1} - 1)) \underset{d}{\rightarrow} N(0, \theta^2(1-\theta)). \qquad \lozenge$$

Primer 7. Neka je $X_1, \dots, X_n$ niz nezavisnih slučajnih veličina iz uniformne raspodele na intervalu $[0, \theta]$. Ako je data funkcija raspodele parametra θ kao

$$P\big(\hat{\theta}_n \leq x\big) = (x/\theta)^n \ za\ 0 \leq x \leq \theta$$

MLE parametra θ je

$$\hat{\theta}_n = X_{(n)} = \max(X_1, \dots, X_n)$$

Tako za ε> 0 imamo

$$P\big(\big|\hat{\theta}_n - \theta\big| > \varepsilon\big) = P\big(\hat{\theta}_n < \theta - \varepsilon\big) = (\theta - \varepsilon/\theta)^n \rightarrow 0, \qquad n \rightarrow \infty$$

pošto $\theta - \varepsilon/\theta < 1$.

---

[10] Za niz ocena $\{\hat{\theta}_n\}$ kažemo da je postojan za parametar θ ako taj niz konvergira u verovatnoći ka θ, tj.ako važi $\lim_{n\to\infty} P\big[\big|\hat{\theta}_n - \theta\big| > \varepsilon\big] = 0$ za svako $\varepsilon > 0$ i svako θ.

[11] Pretpostavimo da je $X_1, \dots, X_n$ niz nezavisnih slučajnih veličina sa istom raspodelom, čije je matematičko očekivanje μ i konačna disperzija $\sigma^2 > 0$. Ako je $S_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$,tada $S_n \underset{d}{\rightarrow} Z \sim N(0,1)$ kada $n \rightarrow \infty$.

[12] Delta metod koristimo za dobijanje približne raspodele funkcije od ocene sa asimptotski normalnom raspodelom. Pretpostavimo da $\sqrt{n}(X_n - \theta) \underset{d}{\rightarrow} N(0, \sigma^2)$, gde su θ i σ^2 konačne konstante a X_n su slučajne veličine. Ako je $g(x)$ funkcija sa izvodom $g'(\theta)$ u $x = \theta$, onda $\sqrt{n}(g(X_n) - g(\theta)) \underset{d}{\rightarrow} N(0, \sigma^2[g'(\theta)]^2)$.

Takođe imamo i

$$P\left[n(\theta-\hat{\theta}_n)\le x\right]=P\left[\hat{\theta}_n\ge\theta-\frac{x}{n}\right]=1-(1-x/\theta n)^n\to 1-e^{-\frac{x}{\theta}}, x\ge 0$$

Pa tako imamo da $n(\theta-\hat{\theta}_n)$ konvergira u raspodeli (distribuciji)[13] ka slučajnoj veličini iz eksponencijalne raspodele. ◊

Pretpostavimo da su $X_1,\dots,X_n$ nezavisne i jednako raspodeljene slučajne veličine, svaka sa sa gustinom (ili zakonom) raspodele $f(x;\theta)$, gde je θ realan parametar.

Definišimo $l(x;\theta)=\ln f(x;\theta)$ i neka su $l'(x;\theta), l''(x;\theta), l'''(x;\theta)$ prva tri izvoda ove funkcije po θ. Uvodimo sledeće pretpostavke o $f(x;\theta)$:

(A1) Parametarski prostor Θ je otvoren podskup realne prave.
(A2) Skup $A=\{x: f(x;\theta)>0\}$ ne zavisi od θ.
(A3) $f(x;\theta)$ je tri puta neprekidno diferencijabilna po θ za svako $x\in A$.
(A4) Matematičko očekivanje $E_\theta[l'(X_i;\theta)]=0$ za svako θ i disperzija $D_\theta[l'(X_i;\theta)]=I(\theta)$ gde je $0<I(\theta)<\infty$ za svako θ.
(A5) $E_\theta[l''(X_i;\theta)]=-J(\theta)$ gde je $0<J(\theta)<\infty$ za svako θ.
(A6) Za svako θ i $\delta>0, |l'''(x;t)|\le M(x)\ za\ |\theta-t|\le\delta$ gde je $E_\theta[M(X_i)]<\infty$.

Pretpostavimo da je $f(x;\theta)$ funkcija gustine. Ako važi uslov (A2) onda imamo

$$\int_A f(x;\theta)dx=1 \quad \text{za svako } \theta\in\Theta$$

pa je

$$\frac{d}{d\theta}\int_A f(x;\theta)dx=0$$

Ako operator diferenciranja i znak integrala mogu da zamene mesta onda imamo

$$0=\int_A \frac{\partial}{\partial\theta}f(x;\theta)dx=\int_A l'(x;\theta)\,f(x;\theta)dx=E_\theta[l'(X_i;\theta)]$$

pa je pretpostavka da je $E_\theta[l'(X_i;\theta)]=0$, u mnogim slučajevima, prirodna posledica uslova (A2). Još češće, ako je moguće diferencirati $f(x;\theta)$ dva puta pod znakom integrala, imamo:

[13] Niz slučajnih veličina (X_n) datih nizom svojih funkcija raspodele (F_n) konvergira u raspodeli ka slučajnoj veličini X, sa funkcijom raspodele $F(x)$, ako važi $\lim_{n\to\infty}F_n(x)=F(x)$. Za konvergenciju u raspodeli koristimo oznaku $X_n\underset{d}{\to}X$.

$$
\begin{aligned}
0 &= \int_A \frac{\partial}{\partial\theta}\left(l'(x;\theta) f(x;\theta)\right) dx \\
&= \int_A l''(x;\theta) f(x;\theta) dx \\
&+ \int_A \left(l'(x;\theta)\right)^2 f(x;\theta) dx = -J(\theta) + I(\theta)
\end{aligned}
$$

pa je onda $J(\theta) = I(\theta)$. $I(\theta)$ se naziva *informaciona funkcija Fišera.*

Primer 8. Neka je $X_1, \dots, X_n$ niz nezavisnih slučajnih veličina iz logističke raspodele čija je funkcija gustine

$$f(x;\theta) = \frac{\exp(x-\theta)}{[1+\exp(x-\theta)]^2}$$

Prvi i drugi izvod po θ funkcije $l(x;\theta) = \ln f(x;\theta)$ su

$$l'(x;\theta) = \frac{\exp(x-\theta) - 1}{1+\exp(x-\theta)}$$

i

$$l''(x;\theta) = -2\frac{\exp(x-\theta)}{[1+\exp(x-\theta)]^2}$$

Dalje imamo da je

$$E_\theta[l'(X_i;\theta)] = \int_{-\infty}^{\infty} \frac{\exp\left(2(x-\theta)\right) - \exp(x-\theta)}{[1+\exp(x-\theta)]^3} dx = 0$$

$$D_\theta[l'(X_i;\theta)] = I(\theta) = \int_{-\infty}^{\infty} (\exp(x-\theta) - 1)^2 \frac{\exp(x-\theta)}{[1+\exp(x-\theta)]^4} dx = \frac{1}{3}$$

i

$$-E_\theta[l''(X_i;\theta)] = J(\theta) = 2\int_{-\infty}^{\infty} \frac{\exp(2(x-\theta))}{[1+\exp(x-\theta)]^4} dx = \frac{1}{3}$$

pa dobijamo da je $I(\theta) = J(\theta)$. ◊

Pod pretpostavkom da važe uslovi od (A1) do (A3), ako $\hat{\theta}_n$ maksimizira verodostojnost, imamo

$$\sum_{i=1}^{n} l'\left(X_i;\hat{\theta}_n\right) = 0$$

a razvijajući levu stranu ove jednakosti Tejlorovim razvojem dobijamo

$$\begin{aligned} 0 &= \sum_{i=1}^{n} l'\left(X_i;\hat{\theta}_n\right) \\ &= \sum_{i=1}^{n} l'(X_i;\theta) + (\hat{\theta}_n-\theta)\sum_{i=1}^{n} l''(X_i;\theta) \\ &+ \frac{1}{2}(\hat{\theta}_n-\theta)^2\sum_{i=1}^{n} l'''(X_i;\theta_n^*) \end{aligned}$$

gde je θ_n^* vrednost između θ i $\hat{\theta}_n$. Deljenjem obe strane jednakosti sa $\sqrt{n}$ dobijamo

$$\begin{aligned} 0 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l'(X_i;\theta) + \sqrt{n}(\hat{\theta}_n-\theta)\frac{1}{n}\sum_{i=1}^{n} l''(X_i;\theta) \\ + \frac{1}{2}\sqrt{n}(\hat{\theta}_n-\theta)^2\frac{1}{n}\sum_{i=1}^{n} l'''(X_i;\theta_n^*) \end{aligned}$$

a sređivanjem i rešavanjem gornje jednakosti dobijamo

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right) = \frac{-n^{-1/2}\sum_{i=1}^{n} l'(X_i;\theta)}{n^{-1}\sum_{i=1}^{n} l''(X_i;\theta) + (2n)^{-1}(\hat{\theta}_n-\theta)\sum_{i=1}^{n} l'''(X_i;\theta_n^*)}$$

A sada iz Centralne granične teoreme i uslova (A4) sledi

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l'(X_i;\theta) \underset{d}{\rightarrow} Z\sim N(0,I(\theta))$$

a iz slabog zakona velikih brojeva[14] i uslova (A5) imamo

$$\frac{1}{n}\sum_{i=1}^{n} l''(X_i;\theta) \underset{p}{\rightarrow} -J(\theta).$$

[14] Neka je $X_1,\ldots,X_n$ niz nezavisnih slučajnih veličina sa istom raspodelom i konačnim matematičkim očekivanjem jednakim µ. Tada $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{p}{\rightarrow} \mu, kada\ n \rightarrow \infty$.

Tako iz teoreme Sluckog[15] imamo da

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \underset{d}{\rightarrow} \frac{Z}{J(\theta)} \sim N(0, I(\theta)/J^2(\theta))$$

pod uslovom

$$\left(\hat{\theta}_n - \theta\right) \frac{1}{n} \sum_{i=1}^{n} \psi'''(X_i; \theta_n^*) \underset{p}{\rightarrow} 0.$$

Kasnije će biti pokazano da ovo poslednje važi pod uslovom da važi (A6) i $\hat{\theta}_n \underset{p}{\rightarrow} \theta$.

Dokazivanje postojanosti ocene metodom maksimalne verodostojnosti $\hat{\theta}_n$ je nešto suptilnije. Razmotrimo (slučajnu) funkciju

$$\phi_n(t) = \frac{1}{n} \sum_{i=1}^{n} [\ln f(X_i; t) - \ln f(X_i; \theta)]$$

koja dostiže maksimum u tački $t = \hat{\theta}_n$. Prema slabom zakonu velikih brojeva, za svako fiksirano $t \in \Theta$, imamo

$$\phi_n(t) \underset{p}{\rightarrow} \phi(t) = E_\theta[\ln(f(X_i; t)/f(X_i; \theta))].$$

Sada primetimo da $-\phi(t)$ dostiže maksimum kada $t = \theta$, pa i $\phi(t)$ dostiže maksimum u $t = \theta$ ($\phi(\theta) = 0$). Štaviše, osim ako $f(x; t) = f(x; \theta)$ za svako $x \in A$ onda $\phi(t) < 0$, pošto pretpostavljamo (implicitno) da znamo o kom se statističkom modelu radi, sledi da $\phi(t)$ dostiže jedinstveni maksimum u $t = \theta$.

Da li činjenica da $\phi_n(t) \underset{p}{\rightarrow} \phi(t)$ za svako t (gde $\phi(t)$ dostiže maksimum u $t = \theta$) podrazumeva da $\hat{\theta}_n \underset{p}{\rightarrow} \theta$? Nažalost, odgovor na ovo pitanje je, uglavnom, ne (osim ako ne uvedemo dodatne pretpostavke o ϕ_n). Da bismo pojednostavili stvari, razmotrićemo neke primere u kojima $\{\phi_n(t)\}$ i $\phi(t)$ nisu slučajne funkcije.

Primer 9. Neka je $\{\phi_n(t)\}$ funkcija definisana sa

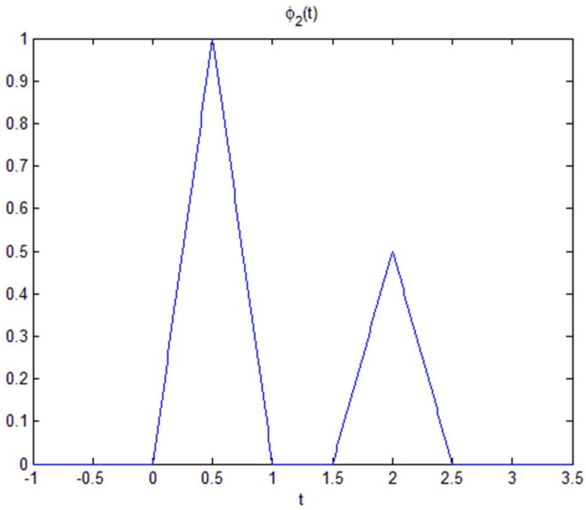
$$\phi_n(t) = \begin{cases} 1 - n\left|t - \frac{1}{n}\right|, & 0 \le t \le 2/n \\ \frac{1}{2} - |t - 2|, & 3/2 \le t \le 5/2 \\ 0, & inače \end{cases}$$

Posmatrajmo grafike $\phi_n(t)$ za određene vrednosti n

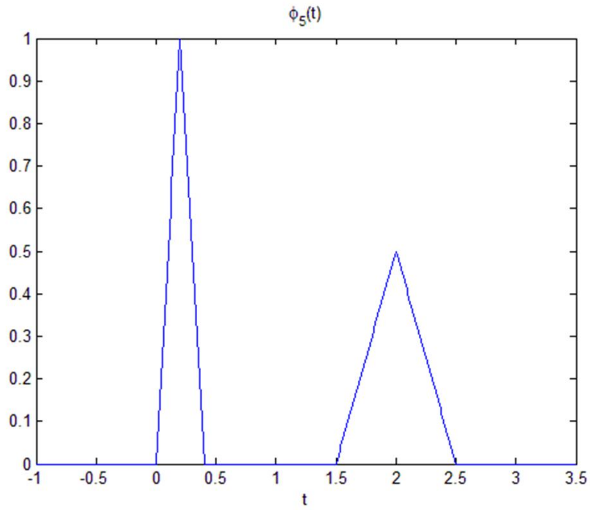
[15] Evgeny Slutsky (1880-1948), ruski matematičar. Pretpostavimo da $X_n \underset{d}{\rightarrow} X$ i $Y_n \underset{p}{\rightarrow} \theta$ (konstanta). Tada

(a) $X_n + Y_n \underset{d}{\rightarrow} X + \theta$

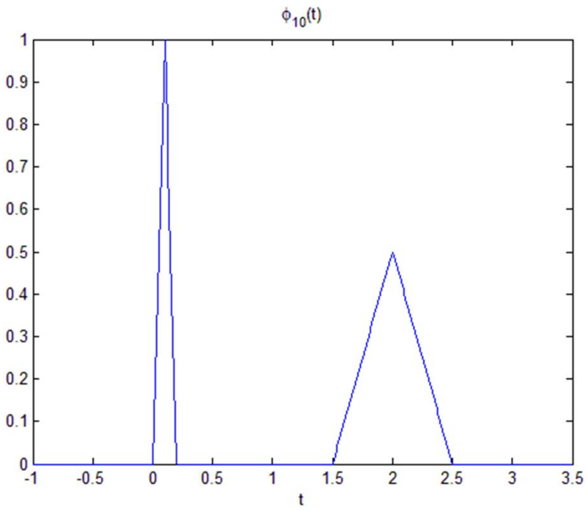
(b) $X_n Y_n \underset{d}{\rightarrow} \theta X$



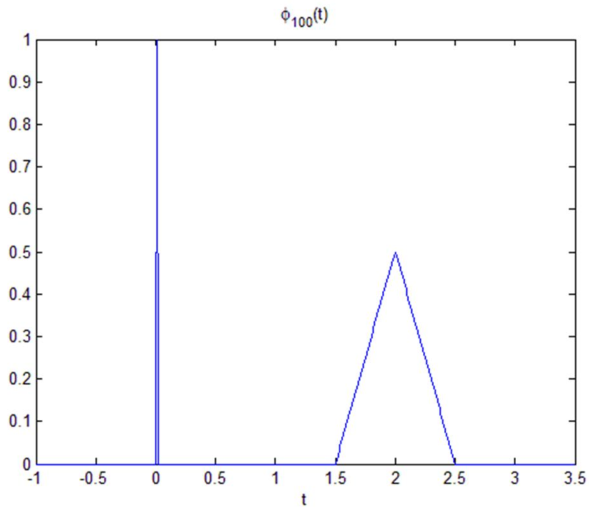
Izgled grafika za $n = 2$



Izgled grafika za $n = 5$



Izgled grafika za $n = 10$



Izgled grafika za $n = 100$

Možemo primetiti da $\phi_n(t)$ dostiže maksimum u $t_n = 1/n$ i $\phi_n(1/n) = 1$. Za svako t zaključujemo $\phi_n(t) \to \phi(t)$, gde je

$$\phi(t) = \begin{cases} \frac{1}{2} - |t-2| \ , & 3/2 \le t \le 5/2 \\ 0 \ , & inače \end{cases}$$

Tako $\phi(t)$ dostiže maksimum u $t_0 = 2$; očito $t_n \to 0 \neq t_0$. ◊

Glavni problem u prethodnom primeru izgleda da je, iako $\phi_n(t) \to \phi(t)$ za svako t, konvergencija nije uniformna; tj. za svako $M > 0$,

$$\sup_{|t| \le M} |\phi_n(t) - \phi(t)| = 1 \quad \text{(za n dovoljno veliko)}$$

pa tako $\phi_n(t)$ ne konvergira uniformno ka $\phi(t)$ za $|t| \leq M$. Međutim, ova uniformna konvergencija nije po sebi dovoljna da garantuje konvergenciju tačaka u kojima $\phi_n(t)$ dostiže maksimum, ka tačkama u kojima $\phi(t)$ dostiže maksimum, što pokazuje sledeći primer.

Primer 10. Neka je $\{\phi_n(t)\}$ funkcija definisana sa

$$\phi_n(t) = \begin{cases} \frac{1}{2}(1-2|t|), & |t| \leq 1/2 \\ 1-2|t-n|, & n-1/2 \leq t \leq t+1/2 \\ 0\,, & inače \end{cases}$$

Lako je zaključiti da za svako $M > 0$

$$\sup_{|t|\leq M} |\phi_n(t) - \phi(t)| \to 0$$

gde je

$$\phi(t) = \begin{cases} \frac{1}{2}(1-2|t|)\,, & |t| \leq 1/2 \\ 0\,, & inače \end{cases}$$

Možemo primetiti da $\phi_n(t)$ dostiže maksimum u $t_n = n$, dok $\phi(t)$ dostiže maksimum u $t_0 = 0$; ponovo $t_n \to \infty \neq t_0$. ◊

Iako uniformna konvergencija $\phi_n(t)$ ka $\phi(t)$ važi za zatvorene i ograničene skupove u ovom primeru, niz tačaka u kojima se dostiže maksimum $\{t_n\}$ ne može biti sadržan unutar zatvorenog i ograničenog skupa pa tako niz ne konvergira ka tački maksimuma funkcije $\phi(t)$. Rezultat pokazuje da je dodavanje uslova da je niz tačaka maksimuma ograničen, dovoljan za konvergenciju t_n ka t_0, gde je t_0 tačka maksimuma za $\phi(t)$, štaviše, ovaj rezultat pokriva slučaj kada su $\{\phi_n(t)\}$ i $\phi(t)$ slučajni.

Teorema 4. Pretpostavimo da su $\{\phi_n(t)\}$ i $\phi(t)$ realne, slučajne funkcije definisane na realnoj pravoj. Pretpostavimo

(a) za svako $M > 0$,

$$\sup_{|t|\leq M} |\phi_n(t) - \phi(t)| \underset{p}{\to} 0;$$

(b) T_n maksimizira $\phi_n(t)$ a T_0 jedinstveno maksimizira $\phi(t)$;

(c) za svako ε> 0 , postoji M_ε takvo da $P(|T_n| > M_\varepsilon) < \varepsilon$ za svako n.

Tada $T_n \underset{p}{\to} T_0$.

Ovu teoremu je teško upotrebiti u praksi, jer iako nije previše teško dokazati uniformnu konvergenciju, može biti zaista teško pokazati $P(|T_n| > M_\varepsilon) < \varepsilon$. Međutim, ako pretpostavimo da je $\phi_n(t)$ konkavna funkcija za svako n, možemo značajno oslabiti uslove prethodne teoreme.

Teorema 5. Pretpostavimo da su $\{\phi_n(t)\}$ i $\phi(t)$ slučajne, konkavne funkcije. Ako

(a) za svako t, $\phi_n(t) \underset{p}{\to} \phi(t)$, i

(b) T_n maksimizira $\phi_n(t)$ a T_0 jedinstveno maksimizira $\phi(t)$

onda $T_n \underset{p}{\to} T_0$.

U većini slučajeva u kojima se prethodna teorema primenjuje, $\phi_n(t)$ će biti srednja vrednost n slučajnih veličina pa možemo koristiti slabi zakon velikih brojeva da pokažemo da $\phi_n(t) \underset{p}{\to} \phi(t)$, u takvim slučajevima, ograničenje funkcije $\phi(t)$ neće biti slučajna veličina, pa i tačka maksimuma T_0 ove funkcije takođe neće biti slučajna. Teoreme 4. i 5. takođe važe ako su funkcije $\{\phi_n\}$ definisane na R^p.

Teorema 5. takođe može biti upotrebljena ako su funkcije $\phi_n(t)$ konveksne. Ako je f konveksne funkcije, onda je $g = -f$ konkavna funkcija, pa se ova teorema može koristiti kako bi se ustanovila konvergencija u verovatnoći tačaka minimuma konveksnih funkcija. Primer 10. ilustruje primenu teoreme 5. u slučaju kada je $\phi_n(t)$ konveksna funkcija.

Primer 10. Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina sa neprekidnom funkcijom raspodele $F(x)$ i pretpostavimo da je μ jedinstvena medijana raspodele tako da je $F(\mu) = 0{,}5$. Medijana uzorka $X_1, \dots, X_n$, $\hat{\mu}_n$, može biti definisana kao minimum od

$$\sum_{i=1}^{n} |X_i - t|$$

ili ekvivalentno, kao minimum od

$$\phi_n(t) = \frac{1}{n}\sum_{i=1}^{n}[|X_i - t| - |X_i|].$$

Lako je videti da je $\phi_n(t)$ konveksna funkcija, pošto je $|a - t|$ konveksna u tački t za svako a. Prema slabom zakonu velikih brojeva imamo

$$\phi_n(t) \underset{p}{\to} E[|X_1 - t| - |X_1|] = \phi(t)$$

a $\phi(t)$ ima minimum u $t = \mu$. Tako $\hat{\mu}_n \underset{p}{\to} \mu$. (Ovaj rezultat važi i ako je $E[|X_i|] = \infty$ pošto $E[|X_i - t| - |X_i|] < \infty$ za svako t). ◊

Sada ćemo navesti teoremu koja govori o asimptotskom normalnom ponašanju MLE za niz nezavisnih slučajnih veličina sa istom funkcijom raspodele. Pretpostavićemo da je postojanost ocena dokazana.

Teorema 6. Pretpostavimo da je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina sa funkcijom gustine $f(x;\theta)$ koje zadovoljavaju uslove regularnosti (A1) – (A6) i pretpostavimo da MLE zadovoljavaju $\hat{\theta}_n \underset{p}{\to} \theta$, gde je

$$\sum_{i=1}^{n} l'\left(X_i; \hat{\theta}_n\right) = 0.$$

Onda

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{d}{\rightarrow} N(0, I(\theta)/J^2(\theta)).$$

Kada $I(\theta) = J(\theta)$, imamo

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{d}{\rightarrow} N(0, 1/I(\theta)).$$

Dokaz:

Već je bilo pokazano da

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \frac{-n^{-1/2} \sum_{i=1}^{n} l'(X_i; \theta)}{n^{-1} \sum_{i=1}^{n} l''(X_i; \theta) + (2n)^{-1}(\hat{\theta}_n - \theta) \sum_{i=1}^{n} l'''(X_i; \theta_n^*)}$$

S obzirom na prethodna razvijanja, potrebno je samo dokazati da

$$R_n = (\hat{\theta}_n - \theta) \frac{1}{2n} \sum_{i=1}^{n} l'''(X_i; \theta_n^*) \underset{p}{\rightarrow} 0.$$

Imamo da za svako $\varepsilon > 0$,

$$P(|R_n| > \varepsilon) = P\left(|R_n| > \varepsilon, \left|\hat{\theta}_n - \theta\right| > \delta\right) + P\left(|R_n| > \varepsilon, \left|\hat{\theta}_n - \theta\right| \leq \delta\right)$$

i

$$P\left(|R_n| > \varepsilon, \left|\hat{\theta}_n - \theta\right| > \delta\right) \leq P\left(\left|\hat{\theta}_n - \theta\right| > \delta\right) \rightarrow 0$$

kad $n \rightarrow \infty$. Ako je $\left|\hat{\theta}_n - \theta\right| \leq \delta$, imamo (prema uslovu (A6))

$$|R_n| \leq \frac{\delta}{2n} \sum_{i=1}^{n} M(X_i)$$

i pošto

$$\frac{1}{n} \sum_{i=1}^{n} M(X_i) \underset{p}{\rightarrow} E_\theta[M(X_1)] < \infty$$

(prema slabom zakonu velikih brojeva), sledi da

$$P\left(|R_n| > \varepsilon, \left|\hat{\theta}_n - \theta\right| \leq \delta\right)$$

može biti proizvoljno malo (za veliko n) uzimajući dovoljno malo δ. Tako $R_n \underset{p}{\rightarrow} 0$ i

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{d}{\rightarrow} N(0, I(\theta)/J^2(\theta))$$

primenjujući teoremu Sluckog. □

Uslovi regularnosti od (A1) do (A6) su, bez sumnje, minimalni uslovi. Konkretno, moguće je oslabiti uslove diferencijabilnosti po cenu povećanja tehničke teškoće dokaza. Na primer, mogli smo da zamenimo uslove (A5) i (A6) nešto slabijim uslovom

$$\frac{1}{n}\sum_{i=1}^{n} l''(X_i;T_n) \underset{p}{\rightarrow} -J(\theta)$$

za svaki niz slučajnih veličina $\{T_n\}$ takav da $T_n \underset{p}{\rightarrow} \theta$. Zaključujemo da se poslednji uslov podrazumeva pod uslovima (A5) i (A6). Uslovi (A1) do (A6) su ispunjeni za veliku većinu jednoparametarskih modela i ralativno su jednostavni za proveru.

6.1 Ocenjivanje standardnih grešaka

U slučajevima kada imamo $I(\theta) = J(\theta)$ prema teoremi 6. za dovoljno veliko n MLE $\hat{\theta}_n$ ima asimptotski normalnu raspodelu sa matematičkim očekivanjem θ i disperzijom $1/(nI(\theta))$. Ovo može biti upotrebljeno kako bi se standardna greška za $\hat{\theta}_n$ aproksimirala sa $[nI(\theta)]^{-1/2}$. Pošto $I(\theta)$ obično zavisi od θ, neophodno je oceniti $I(\theta)$ kako bi se ocenila standardna greška za $\hat{\theta}_n$. Postoje dva pristupa ocenjivanju $I(\theta)$, pa samim tim i standardne greške za $\hat{\theta}_n$.

- Ako $I(\theta)$ ima zatvorenu formu[16], možemo zameniti $\hat{\theta}_n$ sa θ; pa ocena standardne greške postaje

$$\widehat{se}(\hat{\theta}_n)= \frac{1}{\sqrt{nI(\hat{\theta}_n)}}$$

$nI(\hat{\theta}_n)$ se zove očekivana informaciona funkcija Fišera.

- Pošto $-E_\theta[l''(X_i;\theta)] = I(\theta)$, možemo oceniti $I(\theta)$ sa

$$\widehat{I(\theta)} = -\frac{1}{n}\sum_{i=1}^{n} l''(X_i;\hat{\theta}_n)$$

što dovodi do ocene standardne greške

$$\widehat{se}\big(\hat{\theta}_n\big) = \frac{1}{\sqrt{n\widehat{I(\theta)}}} = \left(-\textstyle\sum_{i=1}^{n} l''(X_i;\hat{\theta}_n)\right)^{-1/2}.$$

$n\widehat{I(\theta)}$ se zove posmatrana informaciona funkcija Fišera za θ.

[16] Može se predstaviti analitički, tj. kao konačan broj elementarnih funkcija.

Primer 11. Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz Košijeve[17] raspodele čija je funkcija gustine

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

Prvi i drugi izvod logaritma funkcije gustine su

$$l'(x;\theta) = \frac{2(x-\theta)}{1+(x-\theta)^2} \text{ i } l''(x;\theta) = -\frac{2(1-(x-\theta)^2)}{(1+(x-\theta)^2)^2}.$$

Kako je $I(\theta) = \frac{1}{2}$, onda je očekivana informaciona funkcija Fišera za θ jednaka $\frac{n}{2}$ i odgovarajuća ocena standardne greške je

$$\widehat{se}(\hat{\theta}_n) = \sqrt{\frac{2}{n}}.$$

Sa druge strane, posmatrana informaciona funkcija Fišera za θ je

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

(što nije jednako očekivanoj informacionoj funkciji Fišera) i odgovarajuća ocena standardne greške je

$$\widehat{se}(\hat{\theta}_n) = \left(\sum_{i=0}^{n} \frac{2(1-(X_i-\hat{\theta}_n)^2)}{(1+(X_i-\hat{\theta}_n)^2)^2}\right)^{-1/2} \qquad \Diamond$$

U prethodnom primeru vidimo da možemo različito oceniti standardne greške koristeći posmatranu i očekivanu informacionu funkciju Fišera. Postavlja se pitanje koja je od ove dve ocene superiornija. Rezultati koje su izneli Efron(Bradley Efron) i Hinkle (David V. Hinkley), 1978. godine, ukazuju da treba smatrati superiornijom ocenu na osnovu posmatrane informacione funkcije Fišera. Njihovo obrazloženje je da ova ocena zapravo ocenjuje $D_{\theta}^{1/2}(\hat{\theta}_n | S = s)$ gde je $S = S(\boldsymbol{X})$ drugorazredna statistika za θ i $s = S(\boldsymbol{x})$.

[17] Baron Augustin-Louis Cauchy (1789 – 1857), francuski matematičar, pionir matematičke analize.

6.2 Višeparametarski modeli

Sada ćemo postojanost i asimptotsku normalnost proširiti i na višeparametarske slučajeve. Pretpostavimo da su $X_1, X_2, \ldots$ nezavisne, jednako raspodeljene slučajne veličine, svaka sa funkcijom gustine $f(x;\boldsymbol{\theta})$ gde je $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. Ocena metodom maksimalne verodostojnosti $\hat{\theta}_n$ zasnovana na nezavisnim, jednako raspodeljenim slučajnim veličinama $X_1, X_2, \ldots, X_n$ zadovoljava jednakost

$$\sum_{i=1}^{n} l'\left(X_i; \widehat{\boldsymbol{\theta}}_n\right) = 0$$

gde je sada $l'(x;\boldsymbol{\theta})$ vektor parcijalnih izvoda od $l(x;\boldsymbol{\theta}) = \ln f(x;\boldsymbol{\theta})$ po komponentama $\boldsymbol{\theta}$.

Ideja za dokazivanje asimptotske normalnosti je zapravo ista kao u slučaju modela sa jednim parametrom: pravimo razvoj leve strane gornje jednakosti u Tejlorov red oko prave vrednosti parametra. Tako dobijamo

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'(X_i;\boldsymbol{\theta}) + \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \frac{1}{n} \sum_{i=1}^{n} l''(X_i;\boldsymbol{\theta}_n^*)$$

gde je

(a) $l''(x;\boldsymbol{\theta})$ je matrica drugih parcijalnih izvoda po $\theta_1, \theta_2, \ldots, \theta_p$; (j,k) element ove matrice je dat sa

$$l''_{jk}(x;\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j \partial\theta_k} l(x;\boldsymbol{\theta})$$

(b) $\boldsymbol{\theta}_n^*$ je granični segment koji objedinjuje $\boldsymbol{\theta}$ i $\widehat{\boldsymbol{\theta}}_n$.

Sada rešavajući gornju jednačinu imamo

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) = (-\frac{1}{n} \sum_{i=1}^{n} l''(X_i;\boldsymbol{\theta}_n^*))^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'(X_i;\boldsymbol{\theta})$$

pa bismo onda, pod odgovarajućim uslovima regularnosti, imali

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'(X_i;\boldsymbol{\theta}) \underset{d}{\rightarrow} N_p(0, I(\boldsymbol{\theta}))$$

(gde je $I(\boldsymbol{\theta}) = Cov_\theta[l'(X_i;\boldsymbol{\theta})]$) i

$$\frac{1}{n} \sum_{i=1}^{n} l''(X_i;\boldsymbol{\theta}_n^*) \underset{p}{\rightarrow} -J(\boldsymbol{\theta}) = E_\theta(l''(X_1;\boldsymbol{\theta}))$$

gde su $I(\boldsymbol{\theta})$ i $J(\boldsymbol{\theta})$ matrice formata $p \times p$. Sada pod uslovom da je $J(\boldsymbol{\theta})$ inverzna, imamo

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \underset{d}{\rightarrow} N_p(0, J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}).$$

Kao i u slučaju jednog parametra, za mnoge modele (uključujući eksponencijalne familije) imamo $I(\boldsymbol{\theta})$ = $J(\boldsymbol{\theta})$. U ovom slučaju ograničavajuća kovarijaciona matrica $J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}$ postaje $I(\boldsymbol{\theta})^{-1}$. Ukoliko je moguće diferencirati dva puta pod znakom inegrala (ili unutar sume) po svim elementima $\boldsymbol{\theta}$ na mestu p, onda važi $I(\boldsymbol{\theta})$ = $J(\boldsymbol{\theta})$. $I(\boldsymbol{\theta})$ nazivamo informacionom matricom Fišera.

Sada ćemo navesti uslove regularnosti koji važe za višeparametarski slučaj, a koji su analogni uslovima (A1) – (A6) koji su važili u slučaju jednog parametra:

(B1) Parametarski prostor Θ je otvoren podskup od R^p.
(B2) Skup $A = \{x: f(x; \boldsymbol{\theta}) > 0\}$ ne zavisi od **θ**.
(B3) $f(x; \boldsymbol{\theta})$ je tri puta neprekidno diferencijabilna po **θ** za svako $x \in A$.
(B4) Matematičko očekivanje $E_\theta[l'(X_i; \boldsymbol{\theta})] = 0$ za svako **θ** i disperzija $D_\theta[l'(X_i; \boldsymbol{\theta})] = I(\boldsymbol{\theta})$ gde je $I(\boldsymbol{\theta})$ pozitivno definitna matrica za svako **θ**.
(B5) $E_\theta[l''(X_i; \boldsymbol{\theta})] = -J(\boldsymbol{\theta})$ gde je $J(\boldsymbol{\theta})$ pozitivno definitna matrica za svako **θ**.
(B6) Neka je $l'''_{jkl}(x; \boldsymbol{\theta})$ mešoviti parcijalni izvod od $l(x; \boldsymbol{\theta})$ po $\theta_j, \theta_k, \theta_l$. Za svako $\boldsymbol{\theta}$, $\delta > 0$ i $1 \leq j, k, l \leq p$,

$$l'''_{jkl}(x; \boldsymbol{t}) \leq M_{jkl}(x)$$

za $\|\boldsymbol{\theta} - \boldsymbol{t}\| \leq \delta$ gde je $E_\theta\left[M_{jkl}(X_i)\right] < \infty$.

Teorema 7. Pretpostavimo da je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina sa funkcijom gustine $f(x; \boldsymbol{\theta})$ koje zadovoljavaju uslove regularnosti (B1) – (B6) i pretpostavimo da MLE zadovoljavaju $\widehat{\boldsymbol{\theta}}_n \underset{p}{\rightarrow} \boldsymbol{\theta}$, gde je

$$\sum_{i=1}^{n} l'\left(X_i; \widehat{\boldsymbol{\theta}}_n\right) = 0.$$

Onda

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{d}{\rightarrow} N_p(0, J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}).$$

Kada $I(\boldsymbol{\theta}) = J(\boldsymbol{\theta})$, imamo

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{d}{\rightarrow} N_p(0, I(\boldsymbol{\theta})^{-1}).$$

Dokaz ove teoreme je analogan dokazu teoreme 6. Primećujemo da teorema 7. podrazumeva postojanost $\widehat{\boldsymbol{\theta}}_n$.

Naredna dva primera ilustruju teoremu 7. i primenu uslova regularnosti (B1)-(B6).

Primer 12. Neka je $X_1, \ldots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz normalne raspodele sa matematičkim očekivanjem μ i disperzijom σ^2. Funkcija verodostojnosti u ovom slučaju ima oblik

$$L(x; \mu, \sigma^2) = (\frac{1}{2\pi\sigma^2})^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

Jednačine verodostojnosti imaju oblik

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

pa rešavanjem ovih jednačina dobijamo MLE za μ i σ^2

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \text{ i } \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Sada ćemo posmatrati

$$l(x; \mu, \sigma) = -\frac{1}{2\sigma^2}(x - \mu)^2 - \ln \sigma - \frac{1}{2}\ln(2\pi)$$

i nalaženjem parcijalnih izvoda od l po μ i σ, imamo da je

$$I(\mu, \sigma) = J(\mu, \sigma) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{pmatrix}.$$

Tako (posle provere uslova regularnosti, prema teoremi 7.) imamo

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_n - & \mu \\ \hat{\sigma}_n - & \sigma \end{pmatrix} \underset{d}{\rightarrow} N_2(0, I^{-1}(\mu, \sigma))$$

gde je $I^{-1}(\mu, \sigma)$ dijagonalna matrica sa vrednostima σ^2 i $\sigma^2/2$. ◊

Primer 13. Neka je $X_1, \ldots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz gama raspodele sa parametrima α i λ čija je funkcija gustine

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}\exp(-\lambda x) \quad x, \lambda, \alpha > 0$$

pa je funkcija verodostojnosti

$$L(x; \alpha, \lambda) = \prod_{i=1}^{n} \frac{\lambda^\alpha {x_i}^{\alpha-1}}{\Gamma(\alpha)}\exp(-\lambda x_i)$$

$$= \exp\left[(\alpha - 1)\sum_{i=1}^{n} \ln x_i - \lambda \sum_{i=1}^{n} x_i + n\alpha \ln \lambda - nln\Gamma(\alpha)\right]$$

Jednačine verodostojnosti su

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{\partial \ln L}{\partial \alpha} = nln\lambda + \sum_{i=1}^{n} \ln x_i - n\psi(\alpha) = 0$$

gde je $\psi(\alpha) = (\ln \Gamma(\alpha))'$, ($\psi(\alpha)$ se naziva digama funkcija, a njen izvod $\psi(\alpha)'$ trigama funkcija) pa njihovim rešavanjem dobijamo

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\hat{\alpha}_n}{\hat{\lambda}_n}$$

$$\frac{1}{n}\sum_{i=1}^{n} lnX_i = \psi(\hat{\alpha}_n) - \ln \hat{\lambda}_n$$

Pošto je ovo dvoparametarski model eksponencijalne familije raspodela, informaciona matrica se može izračunati preko drugog parcijalnog izvoda logaritma funkcije gustine, tako imamo

$$\frac{\partial^2}{\partial \alpha^2} \ln f(x; \alpha, \lambda) = -\psi'(\alpha)$$

$$\frac{\partial^2}{\partial \lambda^2} \ln f(x; \alpha, \lambda) = -\frac{\alpha}{\lambda^2}$$

$$\frac{\partial^2}{\partial \alpha \partial \lambda} \ln f(x; \alpha, \lambda) = \frac{1}{\lambda}$$

pa pošto nijedan od ovih izvoda ne zavisi od x lako se moža izračunati

$$I(\alpha, \lambda) = \begin{pmatrix} \psi'(\alpha) & -1/\lambda \\ -1/\lambda & \alpha/\lambda^2 \end{pmatrix}.$$

Inverz ove matrice ima oblik

$$I'(\alpha, \lambda) = \frac{\lambda^2}{\alpha\psi'(\alpha) - 1} \begin{pmatrix} \alpha/\lambda^2 & 1/\lambda \\ 1/\lambda & \psi'(\alpha) \end{pmatrix}.$$

Konkretno, to podrazumeva da

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \underset{d}{\rightarrow} N(0, \alpha/(\alpha\psi'(\alpha) - 1))$$

i

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \underset{d}{\rightarrow} N(0, \lambda^2\psi'(\alpha)/(\alpha\psi'(\alpha) - 1)).$$

Interesantno je uporediti raspodelu za $\sqrt{n}(\hat{\lambda}_n - \lambda)$ (pretpostavljajući da je α nepoznato) sa raspodelom za MLE od λ kada je α poznato.

U ovom drugom slučaju, MLE je

$$\tilde{\lambda}_n = \frac{\alpha}{\bar{X}_n}$$

pa imamo

$$\sqrt{n}(\tilde{\lambda}_n - \lambda) \underset{d}{\rightarrow} N(0, \lambda^2/\alpha).$$

Trebalo bi očekivati da $\tilde{\lambda}_n$ bude efikasnija ocena za λ pošto smo u mogućnosti da u ocenjivanje λ ugradimo i neka znanja o α. Moguće je pokazati da je $\alpha^{-1} < \psi'(\alpha)/(\alpha\psi'(\alpha) - 1)$ za svako $\alpha > 0$; sledeća tabela[18] daje vrednosti α^{-1} i $\psi'(\alpha)/(\alpha\psi'(\alpha) - 1) = g(\alpha)$ za nekoliko različitih vrednosti α.

α	α^{-1}	$g(\alpha)$
0.1	10	11.09
0.5	2	3.36
1	1	2.55
2	0.5	2.22
5	0.2	2.08
10	0.1	2.04
100	0.01	2.02

Primećujemo da što je α veće razlika u efikasnosti je značajnija. U praksi je retko α poznato, pa ne bismo imali izbora nego da koristimo $\hat{\lambda}_n$. ◊

Primer 13 ilustruje opštiji slučaj. Imajući u vidu dva podjednako dobra statistička modela, možemo dobiti efikasnije ocene u modelu sa manjim brojem parametara. Ali mogu nastati ozbiljni problemi ukoliko izaberemo model sa premalo parametara. U praksi je upravo izbor statističkog modela najteži i najznačajniji korak u analizi podataka.

[18] Preuzeto iz primera 5.15, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC, 2000.*

7. Pogrešno pretpostavljeni (određeni) modeli

Važno je imati u vidu da su statistički modeli obično aproksimacija realne slike. Ali izbor pogrešnog modela kojim ćemo opisati posmatrane podatke ne mora biti toliko problematičan sa praktične tačke gledišta. Recimo, pretpostavljeni model može biti jako blizu pravog modela, pa se gubi jako malo svojstava izborom pogrešnog modela. Drugo, ocenjene parametre za dat model često je korisno tumačiti čak iako je pretpostavljeni model pogrešan. Sledeća dva primera ilustruju ove tvrdnje.

Primer 14. Neka je $X_1, \dots, X_n$ niz nezavisnih slučajnih veličina iz eksponencijalne raspodele sa parametrom λ. Međutim, pretpostavimo da smo se odlučili da podatke opišemo modelom definisanim sledećom familijom gustina za X_i sa dva parametra

$$f(x;\alpha,\theta) = \frac{\alpha}{\theta}\left(1+\frac{x}{\theta}\right)^{-(\alpha+1)}, \qquad x > 0$$

$\alpha > 0 \; i \; \theta > 0$ su nepoznati parametri. Iako eksponencijalna raspodela ne pripada ovoj familiji raspodela, lako je uočiti da ako α i θ teže beskonačnosti tako da α/θ teži ka λ, imamo (za $x > 0$),

$$f(x;\alpha,\theta) \to \lambda \exp(-\lambda x) \qquad \left(\alpha,\theta \to \infty \;\; i \;\; \frac{\alpha}{\theta} \to \lambda\right).$$

Tako, dajući razumne ocene $\hat{\alpha}$ i $\hat{\theta}$, ocenjena gustina $f(x;\hat{\alpha},\hat{\theta})$ će biti blizu prave gustine za X_i. ◊

Primer 15. Neka su $X_1, \dots, X_n$ nezavisne slučajne veličine sa matematičkim očekivanjem i disperzijom

$$E(X_i) = \alpha + \beta t_i \;\; i \;\; D(X_i) = \sigma^2$$

gde su α, β, σ^2 nepoznati, a $t_1, \dots, t_n$ su poznate konstante. Pri ocenjivanju α, β, σ^2 često se pretpostavlja da su X_i iz normalne raspodele, i da su parametri ocenjeni metodom maksimalne verodostojnosti, a ove ocene važe čak i kada X_i ne pripadaju normalnoj raspodeli. ◊

Pretpostavimo da su $X_1, \dots, X_n$ nezavisne slučajne veličine sa funkcijom raspodele F. Pretpostavljamo, ipak, da X_i imaju istu funkciju gustine $f(x;\theta)$ za neko $\theta \in \Theta$ gde funkcija raspodele F ne mora obavezno da odgovara nekom $f(x;\theta)$. Pretpostavimo da ocena $\hat{\theta}_n$ zadovoljava jednakost

$$\sum_{i=1}^{n} l'\left(X_i; \hat{\theta}_n\right) = 0.$$

Šta je tačno ocenjivanje $\hat{\theta}_n$? Kakvo je ponašanje niza ocena $\{\hat{\theta}_n\}$ za veliko n? Razmotrimo funkcionalni parameter[19] $\theta(F)$ definisan sa

$$\int_{-\infty}^{\infty} l'\left(x; \theta(F)\right) dF(x) = 0.$$

Uticajna kriva[20] za $\theta(F)$ je

$$\phi(x; F) = \frac{l'\left(x; \theta(F)\right)}{\int_{-\infty}^{\infty} l''\left(x; \theta(F)\right) dF(x)},$$

pa odavde imamo da

$$\sqrt{n}(\hat{\theta}_n - \theta(F)) \underset{d}{\to} N(0, \sigma^2)$$

gde je

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} \left[l'\left(x; \theta(F)\right)\right]^2 dF(x)}{(\int_{-\infty}^{\infty} l''\left(x; \theta(F)\right) dF(x))^2}.$$

Sledeća teorema daje precizne uslove pod kojima je prethodna tvrdnja tačna. Ovi uslovi su analogni uslovima (A1) – (A6).

---

[19] U statistici se često javlja potreba za ocenjivanjem parametara koji u osnovi zavise od funkcije raspodele podataka. Takve parametre nazivamo funkcionalnim parametrima (u literaturi se sreće i termin statistički funkcional). Npr., očekivanje slučajne veličine čija je funkcija raspodele F može biti predstavljena kao $\mu(F) = \int_{-\infty}^{\infty} x dF(x)$. Vrednost $\mu(F)$ očito zavisi od funkcija raspodele F, tako da možemo razmišljati o $\mu(\cdot)$ kao o realnoj funkciji na prostoru funkcija raspodele, na isti način na koji je funkcija x^2 realna funkcija na realnoj pravoj. Još neki primeri funkcionalnih parametara su: $\sigma^2(F) = \int_{-\infty}^{\infty} \left(x - \mu(F)\right)^2 dF(x)$, $med(F) = F^{-1}\left(\frac{1}{2}\right) = inf\{x: F(x) \geq 1/2\}$, mera lokacije θ(F) definisana kao $\int_{-\infty}^{\infty} \psi\left(x - \theta(F)\right) dF(x) = 0$.

[20] Pojam koji omogućava heuristički metod za određivanje raspodele za $\sqrt{n}(\theta\left(\hat{F}_n\right) - \theta(F))$.
Definicij: Uticajna kriva za $\theta(F)$ u F je funkcija $\phi(x; F) = \lim_{t \downarrow 0} \frac{\theta\left((1-t)F + t\Delta_x\right) - \theta(F)}{t}$ koja obezbeđuje postojanje limesa. Uticajna kriva takođe može biti ocenjena sa $\phi(x; F) = \frac{d}{dt}\theta\left((1-t)F + t\Delta_x\right)|_{t=0}$ uvek kada postoji ovaj limes. Δ_x je degenerativna funkcija raspodele koja stavlja svu verovatnoću u x tako da je$\Delta_x(y) = 0 \; za \; y < x \; i \; \Delta_x(y) = 1 \; za \; y \geq x,$ za $t \in [0,1]$ $(1-t)F + t\Delta_x$ je funkcija raspodele. Uticajna kriva dopušta linearnu aproksimaciju razlike $\theta\left(\hat{F}_n\right) - \theta(F)$. Koristeći Centralnu graničnu teoremu i Slutsky-evu teoremu imamo $\sqrt{n}(\theta\left(\hat{F}_n\right) - \theta(F)) \underset{d}{\to} N(0, \sigma^2(F))$, gde je $\sigma^2(F) = \int_{-\infty}^{\infty} \sigma^2(x; F) dF(x)$.

Teorema 8. Pretpostavimo da su $X_1, \dots, X_n$ nezavisne, jednako raspodeljene slučajne veličine sa funkcijom raspodele F i pretpostavimo da ocena $\hat{\theta}_n$ zadovoljava jednakost

$$\sum_{i=1}^{n} l'(X_i; \hat{\theta}_n) = 0.$$

za neko $\hat{\theta}_n$ koje pripada otvorenom skupu Θ.

Ako:

(a) $l'(x;\theta)$ je strogo opadajuća (ili strogo rastuća) funkcija parametra θ za svako x

(b) $\int_{-\infty}^{\infty} l'(x;\theta)dF(x) = 0$ ima jedinstveno rešenje $\theta = \theta(F)$ gde je $\theta(F) \in \Theta$

(c) $\int_{-\infty}^{\infty} \left[l'(x;\theta(F))\right]^2 dF(x) = I(F) < \infty$

(d) $-\int_{-\infty}^{\infty} l''(x;\theta(F))dF(x) = J(\theta) < \infty$

(e) $|l'''(x;t)| \leq M(x)\ za\ \theta(F) - \delta \leq t \leq \theta(F) + \delta\ \ i\ neko\ \delta > 0$

$gde\ je\ \int_{-\infty}^{\infty} M(x)dF(x) < \infty$

onda $\hat{\theta}_n \underset{p}{\to} \theta(F)\ i\ \sqrt{n}(\hat{\theta}_n - \theta(F)) \underset{d}{\to} N(0, I(F)/J^2(F))$.

Pretpostavku da je $l'(x;\theta)$ strogo monotona u θ možemo zameniti monotonošću za svako x tako što ćemo dodati pretpostavku da $\hat{\theta}_n \underset{p}{\to} \theta(F)$. U određenim slučajevima nećemo imati

$$\int l'(x;\theta)dF(x) = 0$$

za svako $\theta \in \Theta$, ali zato možemo imati

$$\lim_{\theta \to a} \int l'(x;\theta)dF(x) = 0$$

za neko a koje leži na granici skupa Θ (ali nije unutar skupa Θ); npr. a može biti $\pm\infty$. U ovakvim slučajevima obično je moguće pokazati da niz ocena $\{\hat{\theta}_n\}$ konvergira u verovatnoći ka a.

Takođe je moguće proširiti rezultate na slučaj sa više parametara.

Pretpostavimo da su $X_1, \dots, X_n$ nezavisne, jednako raspodeljene slučajne veličine sa funkcijom raspodele F i neka ocena $\hat{\theta}_n$ zadovoljava jednakost

$$\sum_{i=1}^{n} l'(X_i; \hat{\boldsymbol{\theta}}_n) = 0$$

gde je $l'(x;\boldsymbol{\theta})$ vektor parcijalnih izvoda (po komponentama $\boldsymbol{\theta}$) od $\ln f(x;\boldsymbol{\theta})$. Tako pod odgovarajućim uslovima imamo

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}(F)) \underset{d}{\to} N_p(0, J(F)^{-1} I(F) J(F)^{-1})$$

gde $\boldsymbol{\theta}(F) = (\theta_1(F), \dots, \theta_p(F))$ zadovoljava

$$\int_{-\infty}^{\infty} l'(x;\boldsymbol{\theta}(F))dF(x) = 0$$

a matrice $I(F)$ i $J(F)$ su definisane kao

$$I(F) = \int_{-\infty}^{\infty} \left[l'\left(x;\boldsymbol{\theta}(F)\right)\right][l'(x;\boldsymbol{\theta})]^T \, dF(x)$$

$$J(F) = -\int_{-\infty}^{\infty} l''(x;\boldsymbol{\theta}(F)) dF(x)$$

(Kao i pre, $l''(x;\boldsymbol{\theta})$ je matrica drugih parcijalnih izvoda od $l(x;\boldsymbol{\theta})$ po $\theta_1,\dots,\theta_p$).

Ilustrujmo ova teorijska razmatranja sa nekoliko primera.

Primer 16. Pretpostavimo da su $X_1,\dots,X_n$ nezavisne, jednako raspodeljene slučajne veličine iz normalne raspodele sa očekivanjem θ_0i disperzijom σ^2. Međutim, mi pogrešno pretpostavljamo da je gustina za X_i data kao

$$f(x;\theta) = \frac{\exp(x-\theta)}{[1+\exp(x-\theta)]^2}$$

za neko θ. MLE za θ za ovaj model na osnovu uzorka $(X_1,\dots,X_n)$ je rešenje jednačine

$$\sum_{i=1}^{n} \frac{\exp\left(X_i-\hat{\theta}_n\right)-1}{\exp\left(X_i-\hat{\theta}_n\right)+1} = 0$$

Moguće je pokazati da $\hat{\theta}_n \underset{p}{\rightarrow} \theta_0$ pošto

$$\int_{-\infty}^{\infty} \left(\frac{\exp(x-\theta_0)-1}{\exp(x-\theta_0)+1}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta_0)^2}{2\sigma^2}\right) dx = 0.$$

Takođe imamo da $\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) \underset{d}{\rightarrow} N\left(0,\gamma^2(\sigma^2)\right)$.
Vrednosti za $\gamma^2(\sigma^2)/\sigma^2$ (kao funkcije od σ) date su u sledećoj tabeli[21]

σ	0.5	1	2	5	10	20	50
$\gamma^2(\sigma^2)/\sigma^2$	1.00	1.02	1.08	1.24	1.37	1.46	1.53

◊

[21] Preuzeto iz primera 5.18, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC, 2000.*

Primer 17. Pretpostavimo da su $X_1, \dots, X_n$ nezavisne slučajne veličine iz uniformne raspodele na intervalu $[0, b]$. Međutim, mi pogrešno pretpostavljamo da X_i imaju gama raspodelu sa gustinom

$$f(x;\alpha,\lambda) = \frac{\lambda^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\lambda x) \quad x > 0$$

Ocene $\hat{\alpha}_n$ i $\hat{\lambda}_n$ zadovoljavaju jednačine

$$n\frac{\hat{\alpha}_n}{\hat{\lambda}_n} - \sum_{i=0}^{n} X_i = 0$$

$$nln\hat{\lambda}_n + \sum_{i=1}^{n} \ln X_i - n\psi(\hat{\alpha}_n) = 0$$

gde je $\psi(\alpha) = (\ln \Gamma(\alpha))'$. Sledi da $\hat{\alpha}_n \underset{p}{\to} \alpha(b)$ i $\hat{\lambda}_n \underset{p}{\to} \lambda(b)$ gde $\alpha(b)$ i $\lambda(b)$ zadovoljavaju jednačine

$$\ln \lambda(b) + \frac{1}{b}\int_0^b \ln x \, dx - \psi\big(\alpha(b)\big) = 0$$

$$\frac{\alpha(b)}{\lambda(b)} - \frac{1}{b}\int_0^b x dt = 0.$$

Iz druge jednačine dobijamo da je $\alpha(b) = b\lambda(b)/2$ i tako $\lambda(b)$ zadovoljava

$$\ln \lambda(b) + \ln b - 1 - \psi(b\lambda(b)/2\ ) = 0$$

Numeričkim rešavanjem (Preuzeto iz primera 5.19, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC,2000.*) ove jednačine imamo

$$\lambda(b) = \frac{3.55585}{b}$$

pa je tako $\alpha(b) = b\lambda(b)/2 = 1.77793$. Sada koristeći $\alpha(b)$ i $\lambda(b)$ možemo izračunati matrice

$$I(F) = \begin{pmatrix} 1 & -b/4 \\ -b/4 & b^2/12 \end{pmatrix}$$

$$J(F) = \begin{pmatrix} 0.74872 & -b/3.55585 \\ -b/3.55585 & b^2/7.11171 \end{pmatrix}.$$

A odavde imamo da

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_n - & \alpha(b) \\ \hat{\lambda}_n - & \lambda(b) \end{pmatrix} \underset{d}{\to} N_2(0, J(F)^{-1} I(F) J(F)^{-1})$$

gde je

$$J(F)^{-1}I(F)J(F)^{-1} = \begin{pmatrix} 9.61 & 16.03/b \\ 16.03/b & 29.92/b^2 \end{pmatrix} \qquad \lozenge$$

Primer 18. Pretpostavimo da su $X_1, \dots, X_n$ nezavisne slučajne veličina iz gama raspodele sa parametrima α i λ, koje predstavljaju uzorak prihoda iz neke populacije. Koristeći $X_1, \dots, X_n$ želimo da ocenimo tzv. Gini indeks[22] raspodele. Za gama raspodelu ovaj indeks zavisi samo od α i dat je izrazom

$$\pi^{-1/2}\frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(\alpha+1)}.$$

Međutim, mi pogrešno pretpostavljamo da su $X_1, \dots, X_n$ iz log-normalne raspodele sa funkcijom gustine

$$f(x;\mu,\sigma) = \frac{1}{\sigma x\sqrt{2\pi}}\exp(-\frac{(\ln x-\mu)^2}{2\sigma^2}) \qquad za\; x > 0$$

($\ln X_i$ će imati normalnu raspodelu sa očekivanjem μ i disperzijom σ^2). Gini indeks za log – normalnu raspodelu zavisi samo od σ i dat je sa $\gamma(\sigma) = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$, gde je $\Phi(x)$ normirana normalna raspodela $N(0,1)$. MLE za σ je

$$\hat{\sigma}_n = (\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2)^{1/2} \quad \text{gde je } Y_i = \ln X_i$$

a odgovarajuća ocena Gini indeksa (pod pretpostavkom log – normalnog modela) je $\gamma(\hat{\sigma}_n)$. Prema slabom zakonu velikih brojeva

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 \underset{p}{\rightarrow} D(\ln X_i) = \psi'(\alpha)$$

gde je ψ' izvod funkcije ψ, iz prethodnog primera, tj. drugi izvod logaritma gama funkcije. Tako primenjujući teoremu neprekidnog preslikavanja[23] imamo

$$\gamma(\hat{\sigma}_n) \underset{p}{\rightarrow} 2\Phi\left(\left(\frac{\psi'(\alpha)}{2}\right)^{\frac{1}{2}}\right) - 1.$$

[22] Ekonomisti često žele da utvrde raspodelu ličnih prihoda u populaciji. Preciznije, oni žele da izmere „nejednakost" ove raspodele. Jedan način da ovo urade jeste da razmatraju tzv. Lorencovu krivu koja daje procenat prihoda najsiromašnijih u populaciji. Naime, neka je F funkcija raspodele(F(x)=0 za x< 0) čija je očekivana vrednost μ(F). Za t između 0 i 1 definišimo $q_F(t) = \frac{\int_0^t F^{-1}(s)ds}{\int_0^1 F^{-1}(s)ds} = \frac{\int_0^t F^{-1}(s)ds}{\mu(F)}$. Jedna od mera nejednakosti zasnovana na Lorencovoj krivoj jeste Gini indeks definisan kao $\theta(F) = 2\int_0^1 \big(t - q_F(t)\big)dt = 1 - \int_0^1 q_F(t)dt$. Gini indeks je zapravo dva puta oblast između funkcija t i $q_F(t)$, pa je $\theta(F) \in [0,1]$. Kada postoji savršena jednakost tj. $q_F(t) = t$ onda je $\theta(F) = 0$, dok se razlika između prihoda najbogatijih i najsiromašnijih pripadnika populacije povećava, tako se i Gini indeks povećava.

[23] Pretpostavimo da je $g(x)$ neprekidna realna funkcija.

(a) Ako $X_n \underset{p}{\rightarrow} X \; tada \; g(X_n) \underset{p}{\rightarrow} g(X)$.

(b) Ako $X_n \underset{d}{\rightarrow} X \; tada \; g(X_n) \underset{d}{\rightarrow} g(X)$.

Granica se sada može uporediti sa pravom vrednošću Gini indeksa u cilju procene asimptotske pristrasnosti prilikom postavke pogrešnog modela i te vrednosti su date u sledećoj tabel (Preuzeto iz primera 5.20, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC,2000.*)

α	$pravi$	$pogrešan$
0.1	0.883	1.000
0.2	0.798	1.000
0.5	0.637	0.884
1	0.500	0.636
2	0.375	0.430
5	0.246	0.261
10	0.176	0.181
20	0.125	0.127
50	0.080	0.080

Poređenje prave vrednosti Gini indeksa (iz gama raspodele)sa pogrešnom verzijom na osnovu log – normalnog modela

Tabela iznad pokazuje da je ocena $\gamma(\hat{\sigma}_n)$ jako slabo pristrasna za malo α, ali i da pristrasnost postepeno nestaje kako α raste. ◊

Teorema 8. ukazuje da je ocena standardne greške za $\hat{\theta}_n$ (kada se ocenjuje jedan parametar) data sa

$$\widehat{se}(\hat{\theta}_n) = (-\sum_{i=1}^{n} l''(X_i;\hat{\theta}_n))^{-1}(\sum_{i=1}^{n}\left[l'(X_i;\hat{\theta}_n)\right]^2)^{1/2}.$$

U slučaju više parametara, kovarijaciona matrica za $\widehat{\boldsymbol{\theta}}_n$ se može oceniti sa

$$\widehat{Cov}(\widehat{\boldsymbol{\theta}}_n) = \hat{J}(\widehat{\boldsymbol{\theta}}_n)^{-1}\hat{I}(\widehat{\boldsymbol{\theta}}_n)\hat{J}(\widehat{\boldsymbol{\theta}}_n)^{-1}$$

gde je

$$\hat{J}(\widehat{\boldsymbol{\theta}}_n) = -\frac{1}{n}\sum_{i=1}^{n} l''(X_i;\widehat{\boldsymbol{\theta}}_n)$$

$$\hat{I}(\widehat{\boldsymbol{\theta}}_n) = \frac{1}{n}\sum_{i=1}^{n}\left[l'(X_i;\widehat{\boldsymbol{\theta}}_n)\right]\left[l'(X_i;\widehat{\boldsymbol{\theta}}_n)\right]^T.$$

Ova ocena je postala poznata pod nazivom „sendvič“ ocena.

8. Neparametarsko ocenjivanje metodom maksimalne verodostojnosti

U prethodnom razmatranju ocenjivanja metodom maksimalne verodostojnosti pretpostavili smo da posmatrani uzorak ima zajedničku funkciju gustine koja zavisi od realnog parametra ili parametra vektora, i odatle smo dobijali funkcije verodostojnosti. Ova formulacija odmah isključuje ocenjivanje metodom maksimalne verodostojnosti za neparametarske modele, jer za ove modele obično nemamo dovoljno jake pretpostavke da bismo definisali funkciju verodostojnosti u uobičajenom smislu.

Moguće je definisati pojam neparametarskog ocenjivanja metodom maksimalne verodostojnosti iako je ta formulacija donekle slabija. Pretpostavimo da su $X_1, \ldots, X_n$ nezavisne slučajne veličine sa nepoznatom funkcijom raspodele F. Želimo da definišemo (neparametarsku) MLE za F. Kako bismo omogućili valjano definisanje, razmatraćemo samo raspodele čija funkcija raspodele ima pozitivne vrednosti u tačkama $X_1, \ldots, X_n$. Jednostavnosti radi, ovde ćemo pretpostaviti da se X_i razlikuju (što bi bio slučaj da su X_i iz neprekidne raspodele). Ako je p_i funkcija raspodele za X_i onda je neparametarska log- funkcija verodostojnosti

$$\ln L(p_1, \ldots, p_n) = \sum_{i=1}^{n} \ln p_i$$

gde je $p_i \geq 0 \ (i = 1, \ldots, n)$ i $p_1 + \cdots + p_n = 1$. Maksimizirajući neparametarsku log-funkciju verodostojnosti dobijamo $\hat{p}_i = 1/n \ (i = 1, \ldots, n)$, pa je tako neparametarska MLE za F empirijska funkcija raspodele $\hat{F}$

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x).$$

Imajući u vidu našu neparametarsku MLE za F, možemo ustanoviti neparametarsku MLE za proizvoljan funkcionalan parametar $\theta(F)$, $\hat{\theta} = \theta(\hat{F})$.

9. Bajesove ocene

Zaključci o obeležju X se donose na osnovu realizovanog uzorka. Postoje dve osnovne grupe zaključaka: o parametrima uraspodeli (tačkasto i intervalno ocenjivanje parametara) i oraspodeli za X (statistički testovi). Svaki zaključak predstavlja jednu vrstu odluke.

U svakoj od posmatranih procedura uzorku se pridružuje odgovarajuća odluka, a odluke se mogu razlikovati za različite uzorke istog obima i iz iste populacije. Odlike takođe možemo videti kao aktivnosti, akcije. U tom smislu što posle donošenja odluke nastavljamo dalji rad u vezi sa posmatranim obeležjem, može se reći da svaka odluka donosi neki gubitak ili dobitak. U nastavku će biti reči samo o gunkciji gubitka vezanoj za neku odluku (akciju).

Ako su u pitanju odluke o parametrima, onda je funkcija gubitka nenegativna funkcija koja svakoj odluci (akciji) i svakoj vrednosti parametra θ dodeljuje broj iz [0,+∞). Odluka, označićemo je sa $d(\boldsymbol{x})$, je funkcija realizovanog uzorka $\boldsymbol{x} = (x_1, \dots, x_n)$, ali pre nego što se uzorak dobije posmatra se zapravo $d(\boldsymbol{X})$, gde je $\boldsymbol{X}$ =$(X_1, \dots, X_n)$ prost slučajan uzorak za obeležje X. S toga je $d(\boldsymbol{X})$ neka sučajna veličina (statistika), pa je i funkcija gubitka $U(d(\boldsymbol{X}), \theta)$. Statistika $d(\boldsymbol{X})$ je zapravo ocena parametra θ, pa može biti označena sa $\tilde{\theta}(\boldsymbol{X})$. Zbog jednostavnosti računanja često se posmatra funkcija gubitka

$$U\big(\tilde{\boldsymbol{\theta}}(\boldsymbol{X}), \theta\big) = c\big[\theta - \tilde{\theta}(\boldsymbol{X})\big]^2$$

gde je c neka pogodno izabrana konstanta, često $c = 1$.
Pošto je funkcija gubitka slučajna veličina, može se odrediti njeno očekivanje

$$E(U\big(\tilde{\theta}(\boldsymbol{X}), \theta\big))$$

koje je funkcija od $\tilde{\theta}$ i θ, i naziva se funkcija rizika, u oznaci $R(\tilde{\theta}, \theta)$. Dakle,

$$R(\tilde{\theta}, \theta) = E(U\big(\tilde{\theta}(\boldsymbol{X}), \theta\big)).$$

U slučaju da obeležje X ima neprekidnu raspodelu imamo da je

$$R(\tilde{\theta}, \theta) = \int_{\boldsymbol{R}^n} U\big(\tilde{\theta}(\boldsymbol{X}), \theta\big) f(\boldsymbol{x}, \theta) d\boldsymbol{x}$$

gde je $f(\boldsymbol{x}, \theta)$ funkcija raspodele uzorka $\boldsymbol{X}$, tj. funkcija verodostojnosti.

Statistike $\tilde{\theta}(\boldsymbol{X})$ kojima se ocenjuje jedan parametar θ u raspodeli obeležja mogu biti raznovrsne po obliku isloženosti izraza. Odluka o izboru statistike se zasniva na nekom (povoljnom) kvalitetu te statistike.

Parametri u raspodeli obeležja se obično posmatraju kao skalarne, neslučajne veličine. U nekim situacijama se dešava da parametar bude slučajna veličina.

Primer 19. Neka su date slučajne veličine $X_1 : B(10, 0.2)$ i $X_2 : B(10, 0.9)$ i neka je p_1 dati broj iz (0,1) i

$$X = \begin{cases} X_1, & sa\ verovatnoćom\ p_1 \\ X_2, & sa\ verovatnoćom\ p_2 = 1 - p_1 \end{cases}$$

Tada se raspodela za X može zapisati i u obliku $X: B(10, p)$, gde je p slučajna veličina sa raspodelom

$$p: \begin{pmatrix} 0.2 & 0.9 \\ p_1 & 1 - p_1 \end{pmatrix}. \qquad \lozenge$$

Ukoliko je parametar θ u raspodeli obeležja X slučajna veličina, onda se postavlja pitanje kako oceniti tu veličinu na osnovu realizovanog uzorka za obeležje X, jer uzorak za parametar θ ne postoji. Zadatak je rešiv uz primenu Bajesove teoreme (formule)[24] i uvođenje funkcije rizika i takav pristup „ocenjivanju“ je zasnovan na aposteriornim verovatnoćama koje se računaju upravo po Bajesovoj formuli.

U opštem slučaju pretpostavimo sledeće:

- za obeležje X je ustanovljen statistički model u obliku familije raspodela sa poznatim analitičkim izrazom $\{f(x,\theta), \theta \in \Theta\}$, gde je Θ parametarski prostor
- parametar θ je slučajna veličina sa datom apriornom raspodelom (tj. njenom funkcijom gustine) $\pi(\theta)$
- poznat je skup mogućih ocena $\tilde{\theta}(\boldsymbol{X})$ za θ, a obično se uzima da se skup vrednosti tih ocena poklapa sa parametarskim prostorom Θ
- kriterijum kvaliteta ocene izražava se funkcijom gubitka $U\big(\tilde{\theta}(\boldsymbol{X}), \theta\big)$, često oblika $c\big[\theta - \tilde{\theta}(\boldsymbol{X})\big]^2$.

Kada se dobije realizovana vrednost uzorka $\boldsymbol{x}$, moguće je posmatrati aposteriornu raspodelu $\pi(\theta|\boldsymbol{x})$. Po Bajesovoj formuli se dobija

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)\pi(\theta)}{\int_\Theta f(\boldsymbol{x},t)\pi(t)dt}.$$

Funkcija rizika za ocenu $\tilde{\theta}(\boldsymbol{X})$ je očekivana vrednost funkcije gubitka

$$R(\tilde{\theta}(\boldsymbol{X}), \theta) = E\Big(U\big(\tilde{\theta}(\boldsymbol{X}), \theta\big)\Big) = \int_{\boldsymbol{R}^n} U\big(\tilde{\theta}(\boldsymbol{x}), \theta\big) f(\boldsymbol{x},\theta) d\boldsymbol{x}.$$

Funkcija rizika je funkcija od θ, a kako je ovde θ slučajna veličina, onda je funkcija rizika slučajna veličina i njeno matematičko očekivanje se naziva apriorni srednji rizik

$$r(\pi, \tilde{\theta}) = E\Big(R(\tilde{\theta}, \theta)\Big) = \int_R R(\tilde{\theta}, \theta)\pi(\theta)d\theta$$

a aposteriorni srednji rizik je

$$\int_R U\big(\tilde{\theta}, \theta\big)\pi(\theta|x)d\theta.$$

Optimalna ocena θ^* je ona koja obezbeđuje infimum apriorne funkcije rizika.

[24] Pretpostavimo sa su $B_1, B_2, \ldots$ disjunktni skupovitakvi da važi $P(B_i) > 0\ za\ svako\ k\ \bigcup_{i=1}^{\infty} B_i = \Omega$. Tada za svaki događaj A

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{\infty} P(B_i)P(A|B_i)}.$$

Primer 20. Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz Bernulijeve raspodele sa nepoznatim parametrom θ i pretpostavimo da je apriorna raspodela za nepoznati parametar beta raspodela, tj.

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 < \theta < 1$$

gde su α i β poznati. Na osnovu datog realizovanog uzorka $\boldsymbol{x}$ imamo da je posteriorna gustina za parametar θ

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)\pi(\theta)}{\int_0^1 f(\boldsymbol{x},t)\pi(t)dt} = \frac{\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}{\int_0^1 t^{y+\alpha-1}(1-t)^{n-y+\beta-1}dt}$$

(gde je $y = x_1 + \dots + x_n$)

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

Primećujemo da je aposteriorna raspodela takođe iz familije beta raspodele, sa parametrima koji zavise od podataka iz uzorka. ◊

U prethodnom primeru apriorna raspodela zavisi od parametara α i β. Ovi parametri se često nazivaju hiperparametri.Pobornici Bajesovog pristupa ocenjivanju tvrde da vrednosti hiperparametara moraju biti navedeni nezavisno od posmatranih podataka, tako da apriorna raspodela odražava prava *a priori* uverenja. Međutim, ako nismo spremni da *a priori* preciziramo vrednosti hiperparametara, moguće je koristiti takozvani empirijski Bajesov pristup, po kome su hiperparametri ocenjeni iz posmatranih podataka i potom zamenjeni u aposteriornoj raspodeli. Precitnije, ako je $\pi(\theta;\alpha)$ apriorna gustina za θ koja zavisi od hiperparametra α, možemo definisati „marginalnu" zajedničku funkciju gustine

$$g(\boldsymbol{x};\alpha) = \int_\Theta f(\boldsymbol{x};\theta)\pi(\theta;\alpha)d\theta.$$

Tada hiperparametar α može biti ocenjen metodom maksimalne verodostojnosti ili metodom zamene, pretpostavljajući da je zajednička funkcija gustine podataka $g(\boldsymbol{x};\alpha)$. Empirijski Bajesov pristup je često jako koristan, mada uprkos svom imenu, on u suštini i nije bajesovski pošto je apriorna raspodela, tj. njeni parametri, ocenjena na osnovu podataka.

Primer 21. Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz Puasonove raspodele sa očekivanjem λ gde je apriorna raspodela za λ data sa

$$\pi(\lambda;\alpha,\beta) = \frac{\alpha^\beta}{\Gamma(\beta)}\lambda^{\beta-1}\exp(-\alpha\lambda), \quad \lambda > 0$$

gde su α,β hiperparametri. Postavljajući da je $y = x_1 + \cdots + x_n$, „marginalna“ zajednička funkcija gustine je data kao

$$g(\boldsymbol{x};\alpha,\beta) = \int_0^\infty \frac{\exp(-n\lambda)\lambda^y}{x_1!\ldots x_n!}\,\pi(\lambda;\alpha,\beta)d\lambda$$

$$= \left(\frac{\alpha}{n+\alpha}\right)^\beta \frac{\Gamma(y+\beta)}{\Gamma(\alpha)x_1!\ldots x_n!}\left(\frac{1}{n+\alpha}\right)^y.$$

Ocene za α i β se mogu dobiti metodom maksimalne verodostojnosti, gde je funkcija verodostojnosti $g(\boldsymbol{x};\alpha,\beta)$. Kao alternativu za dobijanje ocena možemo koristiti metod momenata. Npr. ako o parametru λ razmišljamo kao o slučajnoj veličini sa gustinom $\pi(\lambda;\alpha,\beta)$, imamo da je $E(X_i|\lambda) = \lambda$ i $D(X_i|\lambda) = \lambda$. Tada je

$$E(X_i) = E\big(E(X_i|\lambda)\big) = \frac{\alpha}{\beta}$$

$$D(X_i) = E\big(D(X_i|\lambda)\big) + D\big(E(X_i|\lambda)\big) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}$$

Koristimo da je očekivanje i disperzija za gama raspodelu α/β i α/β^2, respektivno. Ovene $\hat{\alpha}$ i $\hat{\beta}$ se mogu dobiti izjednačavanjem očekivanja i disperzije uzorka sa $\hat{\alpha}/\hat{\beta}$ i $\hat{\alpha}/\hat{\beta} + \hat{\alpha}/\hat{\beta}^2$, respektivno. ◊

Kako prelazimo sa aposteriorne gustine za θ na tačkastu ocenu za θ? Opšti princip je uzeti očekivanje aposteriorne raspodele (ili medijanu) da bude Bajesova tačkasta ocena. Druga, često korišćena ocena je aposteriorni mod, tj. vrednost θ koja maksimizira $\pi(\theta|\boldsymbol{x})$. Ilustrujmo ovo primerom.

Primer 22. U primeru 20. aposteriorna raspodela za θ je

$$\pi(\theta|\boldsymbol{x}) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

gde je $y = x_1 + \cdots + x_n$. Aposteriorno očekivanje je

$$\hat{\theta} = \int_0^1 \theta\pi(\theta|\boldsymbol{x})d\theta = \frac{y+\alpha}{n+\alpha+\beta}$$

dok je aposteriorni mod

$$\breve{\theta} = \frac{y+\alpha-1}{n+\alpha+\beta-2}.$$

Kada su y i n dovoljno veliki u odnosu na α i β, onda su obe ocene $\hat{\theta}$ i $\breve{\theta}$ aproksimativno y/n (što je ocena za θ metodom maksimalne verodostojnosti). ◊

9.1 Konjugovane apriorne raspodele i nepoznavanje apriorne raspodele

Iako su apriorne raspodele u suštini proizvoljne, pogodno je izabrati apriornu raspodelu tako da se posteriorna raspodela može lako izračunati. Klasičan primer takvih apriornih raspodela su konjugovane[25] familije raspodela. Ako aposteriorna raspodela pripada istoj parametarskoj familiji raspodela kao i apriorna raspodela, tada apriornu i aposteriornu raspodelu nazivamo konjugovanim raspodelama.

Pretpostavimo da je za obeležje ustanovljen model iz eksponencijalne familije raspodela, tako da zajednička funkcija gustine ima oblik

$$f(\boldsymbol{x};\theta) = exp\{c(\theta)T(\boldsymbol{x}) + S(\boldsymbol{x}) - d(\theta)\}.$$

i pretpostavimo da apriorna gustina za parametar θ ima oblik

$$\pi(\theta) = K(\alpha,\beta)exp[\alpha c(\theta) - \beta d(\theta)]$$

za neko α i β (gde je $K(\alpha,\beta)$ izabrano tako da je $\pi(\theta)$ gustina). Tada bi aposteriorna raspodela imala oblik

$$\pi(\theta|\boldsymbol{x}) = \frac{exp\{c(\theta)T(\boldsymbol{x}) + S(\boldsymbol{x}) - d(\theta)\} \cdot K(\alpha,\beta)exp[\alpha c(\theta) - \beta d(\theta)]}{\int_{\Theta} exp\{c(t)T(\boldsymbol{x}) + S(\boldsymbol{x}) - d(t)\} \cdot K(\alpha,\beta)exp[\alpha c(t) - \beta d(t)]dt}$$

a kako imenilac razlomka na desnoj strani jednakosti ne zavisi od parametra θ, imamo da je

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)f(\boldsymbol{x};\theta) \propto exp[(T(\boldsymbol{x}) + \alpha)c(\theta) - (\beta + 1)d(\theta)].$$

Iz ovoga se, u opštem slučaju, može zaključiti da

$$\pi(\theta|\boldsymbol{x}) = K(T(\boldsymbol{x}) + \alpha, \beta + 1)exp[(T(\boldsymbol{x}) + \alpha)c(\theta) - (\beta + 1)d(\theta)],$$

što ima potpuno istu formu kao i $\pi(\theta)$.

Apriorne raspodele $\{\pi(\theta)\}$ indeksirane hiperparametrima α i β su konjugovana familija raspodela za eksponencijalnu familiju raspodela.

Primer 23. Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz geometrijske raspodele

$$f(x;\theta) = \theta(1-\theta)^x$$

za $x = 0,1,2,\dots$ gde je $0 < \theta < 1$. Zajednička funkcija raspodele pripada eksponencijalnoj familiji raspodela

$$f(\boldsymbol{x};\theta) = exp\left[\ln(1-\theta)\sum\nolimits_{i=1}^{n} x_i + nln\,\theta\right].$$

[25] Pojam konjugovanih familija raspodela uveli su Rafa (Raiffa H.) i Sklajfer (Schlaifer R.), 1961. godine.

Odavde sledi da će konjugovana familija apriornih raspodela za θ imati oblik

$$\pi(\theta) = K(\alpha,\beta)exp[\alpha \ln(1-\theta) + \beta \ln\theta] = K(\alpha,\beta)(1-\theta)^{\alpha}\theta^{\beta}$$

za $0 < \theta < 1$. Za $\alpha > 0$ i $\beta > 0$, $\pi(\theta)$ zapravo predstavlja beta raspodelu sa

$$K(\alpha,\beta) = \frac{\Gamma(\alpha,\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

Primećujemo da su iste familije raspodela konjugovane u primeru 20. ◊

Primer 24. Neka je $X_1, \ldots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz eksponencialne raspodele sa parametrom λ. Zajednička funkcija gustine je ponovo iz eksponencijalne familije raspodela

$$f(\boldsymbol{x};\theta) = exp\left[-\lambda \sum_{i=1}^{n} x_i + nln\,\lambda\right].$$

Odavde sledi da će konjugovana familija apriornih raspodela za λ imati oblik

$$\pi(\lambda) = K(\alpha,\beta)exp[-\alpha\lambda + \beta \ln\lambda] = K(\alpha,\beta)\lambda^{\beta}\exp(-\alpha\lambda)$$

za $\lambda > 0$. Za $\alpha > 0$ i $\beta > -1$, $\pi(\lambda)$ zapravo predstavlja gama raspodelu sa

$$K(\alpha,\beta) = \frac{\alpha^{\beta+1}}{\Gamma(\beta+1)}.$$

◊

Konjugovane apriorne raspodele se često koriste zbog jednostavnosti pri računanju aposteriornih raspodela, a ne zbog njihove sposobnosti da tačno opišu *a priori* ubeđenja. U suštini, pod pretpostavkom konjugovane apriorne raspodele za θ imamo dosta pojednostavljenu procenu integrala

$$\int_{\Theta} \pi(\theta) f(\boldsymbol{x},\theta)d\theta.$$

Međutim, sa razvojem kompjuterske tehnike, korišćenje konjugovanih raspodela, kako bi se pojednostavio račun pri dobijanju aposteriornih raspodela, je postalo manje bitno, jer se sada tehnike numeričke integracije mogu koristiti za procenu aposteriornih raspodela, i to sa minimalnim teškoćama. Numeričke zehnike su posebno značajne u slučaju višeparametarskih problema, gde korisne konjugovane raspodele više nisu na raspolaganju. Monte-Karlo integracija[26] je često korisna u ovom kontekstu, a u poslednjih nekoliko godina, tehnike kao što su

[26] Generalno, metoda Monte-Karlo se može okarakterisati kao numerička metoda rešavanja matematičkih problema pomoću modeliranja slučajnih promenljivih i statističkog ocenjivanja karakteristika tih promenljivih. Ova metoda je primenljiva na sve matematičke zadatke, ne samo na one vezane za slučajne promenljive. Naziv metode potiče od članka The Monte-Carlo method, N.Metropolis, S.M.Ulam, J.Amer.Stat.Assoc., 44, No. 247, 335-341, iz 1949.godine.

Gibsovo uzorkovanje[27] su korišćene efikasno kod problema sa velikim brojem parametara. Više o ovoj metodi se može naći u Gliksovoj monografiji[28] iz 1996. godine.

Primer 25. Pretpostavimo da želimo da procenimo integral

$$K(\boldsymbol{x}) = \int_{\Theta} \pi(\theta) f(\boldsymbol{x}, \theta) d\theta.$$

Neka je g funkcija gustine na parametarskom prostoru Θ, onda posmatrajmo

$$K(\boldsymbol{x}) = \int_{\Theta} \frac{\pi(\theta) f(\boldsymbol{x}, \theta)}{g(\theta)} g(\theta) d\theta.$$

Da bismo ocenili $K(\boldsymbol{x})$, neka su $T_1, \ldots, T_m$ nezavisne, jednako raspodeljene slučajne veličine sa gustinom g. Za dovoljno veliko m, imaćemo (iz slabog zakona velikih brojeva)

$$\widehat{K}(\boldsymbol{x}) = \frac{1}{m} \sum_{i=1}^{m} \frac{\pi(T_i) f(\boldsymbol{x}, T_i)}{g(T_i)} \approx K(\boldsymbol{x})$$

pa tako $\widehat{K}(\boldsymbol{x})$ služi kao ocena za $K(\boldsymbol{x})$. ◊

Primer 26. U primeru 24. smo videli da su eksponencijalna i gama raspodela konjugovane. Posmatrajmo sada sledeći slučaj. Neka je $X_1, \ldots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz eksponencialne raspodele sa parametrom λ. Pretpostavimo međutim da je apriorna raspodela za parametar λ data standardnom Košijevom raspodelom, pa će apriorna gustina biti oblika

$$\pi(\lambda) = \frac{1}{\pi(1 + \lambda^2)}, \quad -\infty < \lambda < \infty$$

Kako zajednička gustina ima oblik

$$f(\boldsymbol{x}; \lambda) = \lambda^n exp\left[-\lambda \sum_{i=1}^{n} x_i\right]$$

[27] Tehnika je dobila ime po fizičaru J.W.Gibsu (J.W.Gibbs), po analogiji između metode uzorkovanja i fizičke statistike. Algoritam su opisali braća Geman (Stuart i Donald Geman), 1984.godine, oko osam decenija posle Gibsove smrti.

Ovaj algoritam se koristi za dobijanje niza slučajnih uzoraka iz zajedničke raspodele verovatnoća više slučajnih veličina (ili kada ta raspodela nije eksplicitno poznata, ali je uslovna raspodela svake promenljive poznata), kada je direktno uzorkovanje teško sprovesti.

[28] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

zaključujemo da će biti nezgodno izračunati aposteriornu gustinu za parametar λ. Imamo

$$\pi(\lambda|\boldsymbol{x}) = \frac{\lambda^n exp[-\lambda \sum_{i=1}^n x_i](\pi(1+\lambda^2))^{-1}}{\int_{-\infty}^{\infty} \frac{\lambda^n exp[-\lambda \sum_{i=1}^n t_i]}{\pi(1+t^2)} dt}$$

pa će se teškoće pojaviti prilikom računanja integrala u imeniocu. Naravno, možemo pribeći proceni ovog integrala metodom Monte-Karlo, kako je pokazano u prethodnom primeru. U praktičnom delu rada biće obrađen ovaj primer za konkretne uzorke, kako bi se ilustrovala aposteriorna gustina u slučaju kada apriorna i aposteriorna raspodela nisu konjugovane. ◊

Bajesov pristup teoriji odlučivanja je često smatran neobjektivnim. U slučaju kada je parametarski prostor diskretan i konačan, ako imamo m elemenata parametarskog prostora, možemo odrediti apriornu raspodelu uzimajući da je verovatnoća svakog elementa jednaka $1/m$. Međutim, teškoće se javljaju kada parametarski prostor nije konačan.

Skoncentrisaćemo se na slučaj kada je parametarski prostor podskup realne prave. Pretpostavimo prvo da je parametarski prostor Θ interval (a, b). Logičan izbor apriorne raspodele je uniformna raspodela data gustinom

$$\pi(\theta) = \frac{1}{b-a}, \quad a < \theta < b.$$

Međutim, apriorna gustina za $g(\theta)$ neće biti uniformna ako $g(\cdot)$ nije linearna. Ako je parametarski prostor beskonačan interval, onda „ispravna" uniformna apriorna gustina ne postoji, u smislu da je $\int_\Theta k d\theta = \infty$ za svako $k > 0$. Ipak, ova „neispravna" apriorna gustina će često dati validnu aposteriornu gustinu. Aposteriorna gustina je

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x};\theta)}{\int_\Theta f(\boldsymbol{x};t)dt}.$$

Uniformne „neispravne" apriorne raspodele imaju iste probleme kao i uniformne „ispravne" apriorne raspodele, a to je nedostatak svojstva invarijantnosti kod nelinearnih transformacija.

Primer 27 . Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz normalne raspodele sa očekivanjem μ i disperzijom 1. Pretpostavićemo uniformnu apriornu raspodelu za μ, na realnoj pravoj. Onda je aposteriorna gustina za μ data sa

$$\pi(\mu|\boldsymbol{x}) = k(\boldsymbol{x}) exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

gde je

$$k(\boldsymbol{x}) = \left(\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - t)^2\right]\right)^{-1} = \sqrt{\frac{n}{2\pi}}\, exp\left[\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right].$$

Tako imamo da je aposteriorna gustina

$$\pi(\mu|\boldsymbol{x}) = \sqrt{\frac{n}{2\pi}}\, exp\left[-\frac{n}{2}(\mu - \bar{x})^2\right]$$

Pa je aposteriorna raspodela zapravo normalna raspodela sa očekivanjem $\bar{x}$ i disperzijom $1/n$. ◊

Jedan pristup definisanju apriornih raspodela predložio je H. Džefri[29] u svojoj knjizi Theory of Probability iz 1961. godine. Neka je g monotona funkcija na Θ i definišimo $\nu = g(\theta)$. Informaciju o ν možemo definisati kao i obično

$$I(\nu) = D\left[\frac{\partial}{\partial\nu}\ln f(\boldsymbol{X}; g^{-1}(\nu))\right] = \frac{1}{(g'(\theta))^2} D\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X};\theta)\right].$$

Sada izaberimo funkciju g tako da je $I(\nu)$ konstantna i stavimo uniformnu apriornu raspodelu na g(Θ). To podrazumeva da je apriorna raspodela na Θ

$$\pi(\theta) = k|g'(\theta)| \propto D\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X};\theta)\right]^{\frac{1}{2}} = I^{1/2}(\theta)$$

gde je k neka pozitivna konstanta. Ove apriorne raspodele se nazivaju Džefrijeve apriorne raspodele.

Primer 28 . Neka je $X_1, \dots, X_n$ niz nezavisnih, jednako raspodeljenih slučajnih veličina iz Puasonove raspodele sa očekivanjem λ. Informaciona funkcija za λ je $I(\lambda) = n/\lambda$.

Transformacija funkcije g mora da zadovoljava

$$\lambda(g'(\lambda))^2 = const.$$

Tako je Džefrijeva apriorna raspodela za λ

$$\pi(\lambda) = \frac{k}{\sqrt{\lambda}}, \quad \lambda > 0.$$

Pošto je $\int_0^\infty \lambda^{-1/2} d\lambda = \infty$, apriorna raspodela je „neispravna", pa se k može izabrati proizvoljno. Koristeći Džefrijevu apriornu raspodelu, aposteriorna raspodela za λ posmatrajući uzorak $\boldsymbol{x} = (x_1, \dots, x_n)$ je gama raspodela sa gustinom

$$\pi(\lambda|\boldsymbol{x}) = \frac{n^{y+1/2}\lambda^{y-1/2}\exp(-n\lambda)}{\Gamma(y + 1/2)}$$

gde je $y = x_1 + \cdots + x_n$. Ova aposteriorna gustina je uvek „ispravna", iako je apriorna gustina „neispravna". ◊

[29] Sir Harold Jeffreys (1891 – 1989), engleski matematičar i geofizičar.

9.2 Osobine Bajesovih ocena

1. Ako je funkcija gubitka ograničena i ako se za ocenu θ^* postiže infimum aposteriorne funkcije rizika, onda se za isto θ^* postiže i infimum apriorne funkcije rizika. (Znači da se kao optimalna ocena uzima ona koja obezbeđuje infimum aposteriorne funkcije rizika)
2. Ako je funkcija gubitka strogo konveksna na Θ, onda je Bajesova ocena θ^* jedinstvena.
3. Ako je funkcija gubitka kvadratna funkcija onda je θ^* jedinstvena (na osnovu 2.) i jednaka uslovnom matematičkom očekivanju $E(\theta|X)$.
4. Primena Bajesovih ocena na velikim uzorcima opravdana je činjenicom da, uz dovoljno opšte pretpostavke, Bajesove ocene predstavljaju dovoljne statistike, asimptotski su nepristrasne i postojane, imaju asimptotski normalnu raspodelu, a od ocena dobijenih metodom maksimalne verodostojnosti se razlikuju za red veličine $1/\sqrt{n}$.
5. Problemi i nedostaci ove teorije proizilaze iz pretpostavke da parametar θ jeste slučajna veličina i da treba odabrati raspodelu za θ.Takođe, izračunavanja su složenija ukoliko je oblik funkcije gubitaka složen.
   Metoda se često primenjuje za uzorak obima $n = 1$, što po nekim autorima označava ne statistički, već subjektivni pristup ocenjivanju.

10. Numeričko izračunavanje ocena metodom maksimalne verodostojnosti

U mnogim problemima ocenjivanja, teško je dobiti formu analitičkih izraza za ocenjivanje metodom maksimalne verodostojnosti. U ovakvim situacijama je obično neophodno numerički izračunati ocene metodom maksimalne verodostojnosti. Mnoge numeričke metode mogu se koristiti za dobijanje ocena, a ovde će biti razmotrene metode: Njutn-Rapson[30] (i njena Fišerova modifikacija) i EM[31] algoritam.

10.1 Algoritam Njutn-Rapson

Ovaj algoritam je opšte namene za pronalaženje rešenja nelinearnih jednačina ili sistema nelinearnih jednačina. Njutn-Rapson algoritam je dobar izbor za računanja MLE, pošto su ove ocene često rešenje sistema jednačina (jednačina verodostojnosti).

Pretpostavimo da želimo da nađemo rešenje jednačine $g(x_0) = 0$ gde je g diferencijabilna funkcija. S obzirom na broj x koji je blizu broja x_0, iz razvoja u Tejlorov red oko x sledi da

$$0 = g(x_0) \approx g(x) + g'(x)(x_0 - x)$$

i rešavanjem po x_0 dobijamo

$$x_0 \approx x - \frac{g(x)}{g'(x)}.$$

Tako pomoću x_k, možemo dobiti x_{k+1}

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

i ova procedura prolazi kroz iteracije za $k = 1,2,3,\ldots$ dokle god $|g(x_k)|$ (ili $|g(x_k)/g'(x_k)|$) nije dovoljno malo.

Pretpostavimo da je zajednička gustina za $\boldsymbol{X} = (X_1, \ldots, X_n)$ data sa $f(\boldsymbol{x};\theta)$ i neka je $L(\theta) = f(\boldsymbol{x};\theta)$ funkcija verodostojnosti zasnovana na $\boldsymbol{X} = \boldsymbol{x}$. Pretpostavimo da MLE $\hat{\theta}$ zadovoljava $S(\hat{\theta}) = 0$, gde je $S(\theta)$ izvod od log-funkcije verodostojnosti, $\ln L(\theta)$. ($S(\theta)$ se često zove funkcija „pogotka“) Neka je $\hat{\theta}^{(k)}$ ocena za θ posle k iteracija algoritma; onda

[30] Sir Isaac Newton (1642–1727), engleski fizičar, matematičar, astronom.
Joseph Raphson (1648–1715), engleski matematičar.

[31] Skraćenica od expectation–maximization.

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{S(\hat{\theta}^{(k)})}{H(\hat{\theta}^{(k)})}$$

gde je

$$H(\theta) = -\frac{\partial^2}{\partial\theta^2}\ln L(\theta).$$

Procedura prolazi kroz iteracije dok kriterijum konvergencije ne bude ispunjen, tj. dok $\left|S(\hat{\theta}^{(k)})\right|$ ili $\left|\hat{\theta}^{(k)} - \hat{\theta}^{(k+1)}\right|$ nije dovoljno malo.

Kako bi bilo moguće koristiti Njutn-Rapson algoritam potrebno je imati inicijalnu vrednost ocene za θ, $\hat{\theta}^{(0)}$. Zapravo, u nekim slučajevima, ova inicijalna ocena je problematična jer algoritam neće uvek konvergirati za dato $\hat{\theta}^{(0)}$. Takođe je moguće da jednačina $S(\hat{\theta}) = 0$ ima nekoliko rešenja, svako od rešenja odgovara ili lokalnom minimumu, ili lokalnom maksimumu ili prevojnoj tački log-funkcije verodostojnosti, zato je moguće da će niz ocena $\{\hat{\theta}^{(k)}\}$ konvergirati ka pogrešnom rešenju za $S(\hat{\theta}) = 0$. (Pitanje konvergencije je mnogo važnije kada ocenjujemo tri ili više parametara, kada je u pitanju jedan ili dva parametra moguće je iz grafika log-funkcije verodostojnosti odrediti odgovarajuću inicijalnu ocenu). Ako nije očigledno da će algoritam konvergirati ka maksimumu funkcije verodostojnosti onda treba probati sa nekoliko drugih inicijalnih ocena ili treba probati drugu metodu za nalaženje inicijalne ocene za algoritam. Ako je $\hat{\theta}_n^{(0)}$ zasnovana na n opservacija i $\sqrt{n}(\hat{\theta}_n^{(0)} - \theta)$ konvergira u raspodeli onda $\sqrt{n}(\hat{\theta}_n^{(1)} - \hat{\theta}_n) \underset{p}{\to} 0$, gde je $\hat{\theta}_n$ MLE, a $\hat{\theta}_n^{(1)}$ je ocena dobijena u prvoj iteraciji, koristeći $\hat{\theta}_n^{(0)}$ kao početnu vrednost. Tako možemo uzeti da $\hat{\theta}_n^{(0)}$ bude ocena metodom momenata (ili ocena nekom drugom metodom zamene) za θ.

Primer 29. Pretpostavimo da su $X_1, \dots, X_n$ nezavisne slučajne veličina iz Košijeve raspodele sa funkcijom gustine

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

Logaritam funkcije verodostojnosti je

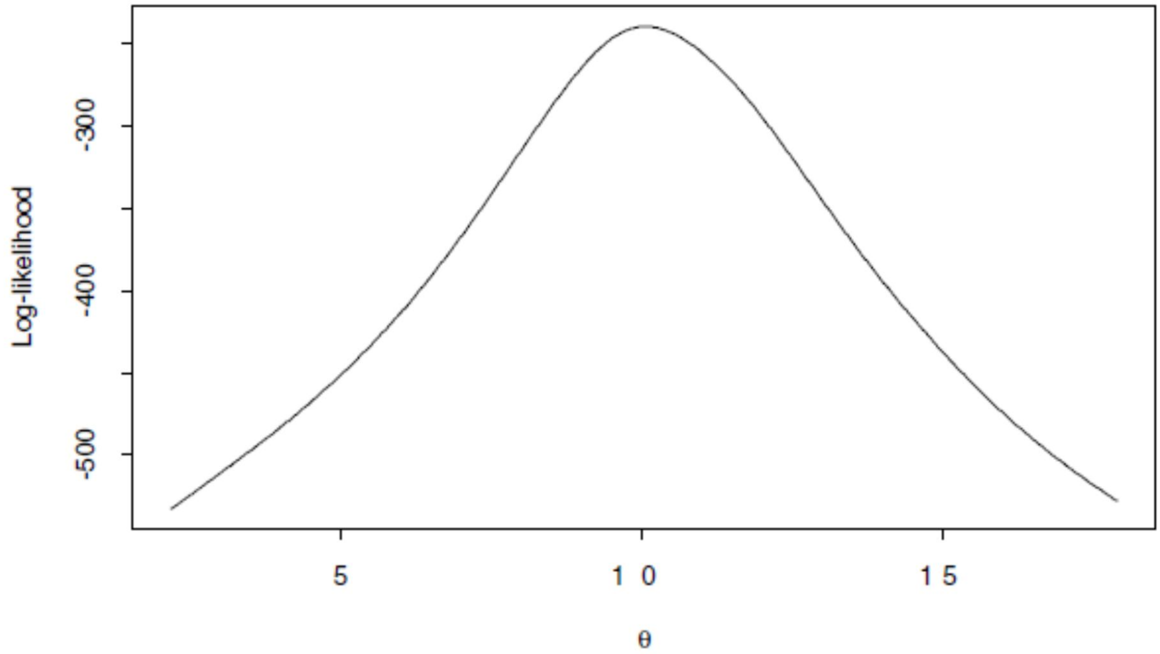
$$\ln L(\theta) = -\sum_{i=1}^{n} \ln[1 + (x_i - \theta)^2] - n\ln\pi.$$

Ocena metodom maksimalne verodostojnosti $\hat{\theta}$ zadovoljava jednačinu

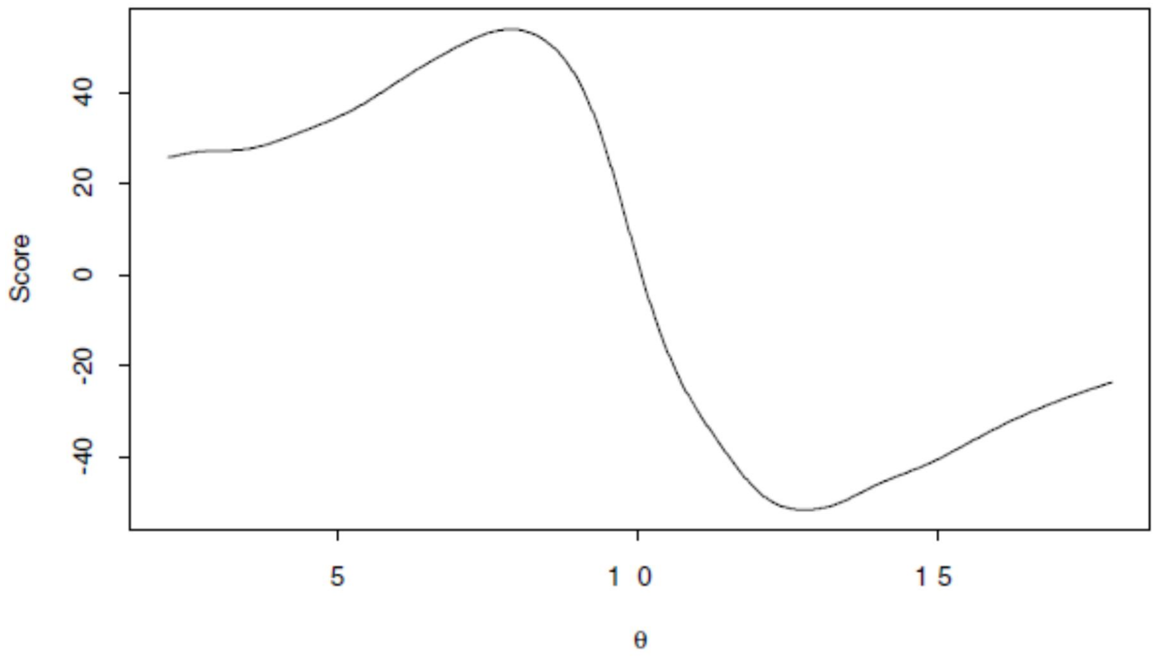
$$S(\hat{\theta}) = \sum_{i=1}^{n} \frac{2(x_i - \hat{\theta})}{1 + (x_i - \hat{\theta})^2} = 0.$$

Funkcija $S(\theta)$ nije monotona i otuda jednačina $S(\hat{\theta}) = 0$ može imati više od jednog rešenja za dati uzorak $(x_1, \dots, x_n)$.

Kako bismo ilustrovali Njutn-Rapson algoritam posmatraćemo uzorak sa 100 elemenata gde je vrednost parametra $\theta = 10.$[32] Sledeći grafici prikazuju logaritam funkcije verodostojnosti $\ln L(\theta)$ i funkciju „pogotka“ $S(\theta)$ za posmatrane podatke iz Košijeve raspodele.



Izgled grafika logaritma funkcije verodostojnosti za podatke iz Košijeve raspodele



Izgled grafika funkcije „pogotka“ za podatke iz Košijeve raspodele

[32]Podaci i rezultati preuzeti iz primera 5.21, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC, 2000.*

Kako bismo odredili $\hat{\theta}^{(0)}$, prvo treba naći adekvatnu inicijalnu vrednost ocene za θ. Pošto je gustina za X_i simetrična oko θ, ima smisla razmatrati ili srednju vrednost ili medijanu uzorka kao inicijalne ocene. Međutim, kako matematičko očekivanje za X_i iz Košijeve raspodele nije definisano, možemo pretpostaviti da srednja vrednost uzorka ne mora biti dobra ocene za θ. Zato ćemo kao inicijalnu ocenu koristiti medijanu uzorka. Sukcesivne vrednosti za $\hat{\theta}^{(k)}$ su definisane sa

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{S(\hat{\theta}^{(k)})}{H(\hat{\theta}^{(k)})}$$

gde je

$$H(\theta) = 2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{(1+(x_i-\theta)^2)^2}.$$

Vrednosti za $\hat{\theta}^{(k)}$ i $\ln L(\hat{\theta}^{(k)})$ u različitim iteracijama date su u sledećoj tebeli

k	$\hat{\theta}^{(k)}$	$\ln L(\hat{\theta}^{(k)})$
0	10.04490	-239.6569
1	10.06934	-239.6433
2	10.06947	-239.6433
3	10.06947	-239.6433

Najkritičniji deo je izbor $\hat{\theta}^{(0)}$, npr. ako je uzeto da je $\hat{\theta}^{(0)}$ manje od 8.74 ili veće od 11.86 onda niz $\{\hat{\theta}^{(k)}\}$ neće konvergirati. ◊

Njutn-Rapson algoritam može takođe biti proširen i na slučajeve sa više parametara. Neka je $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ i pretpostavimo da je MLE za θ data jednačinom $S(\hat{\boldsymbol{\theta}}) = 0$ gde je

$$S(\boldsymbol{\theta}) = (\frac{\partial}{\partial\theta_1}\ln L(\boldsymbol{\theta}), \dots, \frac{\partial}{\partial\theta_p}\ln L(\boldsymbol{\theta}))^T.$$

Tada za dato $\hat{\boldsymbol{\theta}}^{(k)}$,definišemo $\hat{\boldsymbol{\theta}}^{(k+1)}$ na sledeći način

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \hat{\boldsymbol{\theta}}^{(k)} + \left[H\left(\hat{\boldsymbol{\theta}}^{(k)}\right)\right]^{-1} S(\hat{\boldsymbol{\theta}}^{(k)})$$

gde je $H(\boldsymbol{\theta})$ matrica negativnih drugih parcijalnih izvoda od $\ln L(\boldsymbol{\theta})$, a element (i,j) matrice H je dat sa

$$H''_{ij}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L(\boldsymbol{\theta}).$$

Ocene za standardnu grešku ocena parametra se mogu dobiti iz Njutn-Rapson algoritma. Za slučaj jednog parametra, disperzija za $\hat{\theta}$ se može oceniti sa $\left[H(\hat{\theta})\right]^{-1}$, dok se u slučaju više parametara kovarijaciona matrica za $\hat{\theta}$ može oceniti matricom $\left[H(\widehat{\boldsymbol{\theta}})\right]^{-1}$. Ove ocene disperzije podrazumevaju da je model korektno definisan, tj. da podaci pripadaju familiji raspodela koja se ocenjuje.

10.2 Fišer scoring algoritam

Ovaj algoritam predstavlja modifikaciju Njutn-Rapson algoritma, tako što je H zamenjeno sa H^* gde je (i,j) element od H^*

$$H_{ij}^*(\boldsymbol{\theta}) = E_\theta\left[H_{ij}(\boldsymbol{\theta})\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f(\boldsymbol{X};\boldsymbol{\theta})\right]$$

A ovo matematičko očekivanje je izračunato uz pretpostavku da je $\boldsymbol{\theta}$ prava vrednost parametra. H se naziva posmatrana informaciona matrica Fišera, a H^* je očekivana informaciona matrica Fišera. Ukoliko je $\widehat{\boldsymbol{\theta}}^{(k)}$ ocena za $\boldsymbol{\theta}$ posle k iteracija, definišemo $\widehat{\boldsymbol{\theta}}^{(k+1)}$ sa

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \left[H\left(\widehat{\boldsymbol{\theta}}^{(k)}\right)\right]^{-1} S\left(\widehat{\boldsymbol{\theta}}^{(k)}\right).$$

Bitna razlika između Njutn-Rapson algoritma i Fišer scoring algoritma je u činjenici da $H^*(\boldsymbol{\theta})$ zavisi od posmatrane vrednosti za $\boldsymbol{X},\ \boldsymbol{x},$ samo kroz vrednost parametra $\boldsymbol{\theta}$, dok $H(\boldsymbol{\theta})$, generalno, zavisi i od $\boldsymbol{\theta}$ i od $\boldsymbol{x}.$

Primer 30. Pretpostavimo da su $X_1,\dots,X_n$ nezavisne slučajne veličina iz Košijeve raspodele, kao u primeru 19. Već smo u prethodnom primeru videli da je

$$H(\theta) = 2\sum_{i=1}^{n}\frac{1-(x_i-\theta)^2}{(1+(x_i-\theta)^2)^2}$$

pa je onda

$$H^*(\theta) = \frac{n}{\pi}\int_{-\infty}^{\infty}\frac{1-(x-\theta)^2}{(1+(x-\theta)^2)^3}dx = \frac{n}{2}.$$

Otuda imamo da je Fišer scoring algoritam

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{4}{n}\sum_{i=1}^{n}\frac{x_i-\hat{\theta}^{(k)}}{1+(x_i-\hat{\theta}^{(k)})^2}.$$

Vrednosti za $\hat{\theta}^{(k)}$ i $\ln L(\hat{\theta}^{(k)})$ u zavisnosti od iteracije date su u sledećoj tebeli[33]

k	$\hat{\theta}^{(k)}$	$\ln L(\hat{\theta}^{(k)})$
0	10.04490	-239.6569
1	10.06710	-239.6434
2	10.06923	-239.6433
3	10.06945	-239.6433
4	10.06947	-239.6433

Iteracije za Fišer scoring algoritam za podatke iz Košijeve raspodele

Kao i u primeru 19., za $\hat{\theta}^{(0)}$ je uzeta vrednost medijane uzorka. Jedna prednost Fišer scoring algoritma, u ovom primeru, je činjenica da $\hat{\theta}^{(k)}$ konvergira za mnogo širi interval od početne vrednosti $\hat{\theta}^{(0)}$. ◊

Razlike između Njutn-Rapson alogoritma i Fišer scoring algoritma su male, ali ipak važne. Iako je teško mnogo generalizovati, moguće je zaključiti sledeće:

- Konvergencija Njutn-Rapson alogoritma je često brža (pod uslovom da nizovi aproksimacija u oba algoritma konvergiraju).
- Radijus konvergencije za Fišer scoring algoritam je često veći, što ukazuje da je izbor inicijalne ocene manje bitan za ovaj algoritam.

U slučaju modela iz eksponencijalne familije, ova dva algoritma su skoro ekvivalentna, ako su $(\theta_1, \ldots, \theta_p)$ parametri iz skupa prirodnih brojeva onda je $H(\boldsymbol{\theta}) = H^*(\boldsymbol{\theta})$, pa su i ova dva algoritma identična.

[33] Podaci i rezultati preuzeti iz primera 5.22, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC, 2000.*

10.3 EM algoritam

Ovaj algoritam pruža koristan okvir za izračunavanje ocena metodom maksimalne verodostojnosti u tzv. situacijama nepotpunih podataka, tj. kada podaci nedostaju ili nisu verodostojno zabeleženi. Formalizovan je u radu Demstera (Arthur Pentland Dempster), Lirda (Nan M. Laird) i Rubina (Donald Bruce Rubin), 1977. godine, mada su specijalni slučajevi ovog algoritma korišćeni u pojedinim problemima mnogo godina pre 1977. godine. Danas se EM algoritam koristi u rešavanju raznih problema. Sledeći primer ilustruje problem nekompletnih podataka.

Primer 31. Pretpostavimo da su $X_1, \ldots, X_n$ nezavisne slučajne veličina iz eksponencijalne raspodele sa parametrom λ. Imajući u vidu realizovane vrednosti $(x_1, \ldots, x_n)$ uzorka $(X_1, \ldots, X_n)$, MLE za θ je $1/\bar{x}$, gde je $\bar{x}$ realizovana vrednost uzoračke sredine. Međutim, pretpostavimo da umesto da posmatramo ceo uzorak, posmatramo donje i gornje granice, u_i i v_i, za X_i, tako da je $u_i < X_i < v_i$. Imajući u vidu $(u_1, v_1), \ldots, (u_n, v_n)$, funkcija verodostojnosti za λ je

$$L(\lambda) = \prod_{i=1}^{n} P_\lambda(u_i \le X_i \le v_i) = \prod_{i=1}^{n} (\exp(-\lambda u_i) - \exp(-\lambda v_i)).$$

Ovo je jednostavan primer cenzurisanih intervalnih podataka. ◊

Glavna ideja kod EM algoritma je konstrukcija ocene celokupnih podataka, koji posle mogu biti maksimizirani korišćenjem tradicionalnih numeričkih metoda (kao što je Njutn-Rapson algoritam). Tako EM algoritam zapravo i nije numerički algoritam, već procedura opšte namene za računanje ocena parametara nekompletnih podataka tako što se iteracijama računaju ocene parametara metodom maksimalne verodostojnosti na osnovu funkcije verodostojnosti formirane pomoću kompletnih podataka. U osnovi EM algoritma je pretpostavka da se ocene parametara metodom maksimalne verodostojnosti za celokupne podatke mogu lako izračunati, dok je sa funkcijom verodostojnosti nekompletnih podataka teško raditi. EM algoritam ponavlja dva koraka (koji se zovu E i M korak) do konvergencije. Pre detaljnog opisa algoritma, ilustrujmo ga na prethodnom primeru.

Primer 32. U primeru 21. logaritam funkcije verodostojnosti bazirane na realizovanim vrednostima uzorka $(x_1, \ldots, x_n)$ je

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i.$$

Prvi korak EM algoritma je nalaženje ocene od $\ln L(\lambda)$ uzimajući u obzir nekompletne podatke $(u_1, v_1), \dots, (u_n, v_n)$ i pretpostavljajući da je prava vrednost parametra $\hat{\lambda}^{(k)}$, gde je $\hat{\lambda}^{(k)}$ ocena parametra λ posle k iteracija algoritma. Ovo se postiže nalaženjem očekivanja za

$$n \ln \lambda - \lambda \sum_{i=1}^{n} X_i$$

ako je dato $u_1 < X_1 < v_1, \dots, u_n < X_n < v_n$ i pretpostavljajući da je vrednost parametra $\hat{\lambda}^{(k)}$. Dobijamo

$$E(X_i | u_i < X_i < v_i; \lambda) = \frac{1}{\lambda} + \frac{u_i \exp(-\lambda u_i) - v_i \exp(-\lambda v_i)}{\exp(-\lambda u_i) - \exp(-\lambda v_i)}$$

i kako bismo završili korak E algoritma, zamenjujemo

$$\hat{x}_i^{(k)} = \frac{1}{\hat{\lambda}^{(k)}} + \frac{u_i \exp\left(-\hat{\lambda}^{(k)} u_i\right) - v_i \exp(-\hat{\lambda}^{(k)} v_i)}{\exp\left(-\hat{\lambda}^{(k)} u_i\right) - \exp(-\hat{\lambda}^{(k)} v_i)}$$

sa x_i u $\ln L(\lambda)$. Korak M algoritma uključuje nalaženje maksimuma novodobijene log-funkcije verodostojnosti. Maksimum se postiže u

$$\hat{\lambda}^{(k+1)} = (\frac{1}{n} \sum_{i=1}^{n} \hat{x}_i^{(k)})^{-1}.$$

Koraci E i M se ponavljaju dok kriterijum za konvergenciju ne bude ispunjen. ◊

Sada možemo opisati EM algoritam. Pretpostavimo da su $\boldsymbol{X} = (X_1, \dots, X_n)$ neprakidne slučajne veličine sa zajedničkom funkcijom gustine $f_X(\boldsymbol{x}; \theta)$ (gde parametar θ može biti realan ili zadat vektorom) i neka su $\boldsymbol{Y} = (Y_1, \dots, Y_m)$ slučajne veličine takve da $Y_i = g_i(\boldsymbol{X})$ gde su $g_1, \dots, g_m$ poznate funkcije. Preslikavanje koje X_i slika u Y_i obično nije jedan-jedan, što znači da svaki dati ishod za Y_i, $y_1, \dots, y_m$, može biti dobijen od više različitih $x_1, \dots, x_n$. Na taj način na X_i gledamo kao na kompletne podatke, a na Y_i kao na nekompletne. Zajednička gustina za $\boldsymbol{Y}$ je

$$f_Y(y; \theta) = \int_{A(y)} f_X(\boldsymbol{x}; \theta)\, d\boldsymbol{x}$$

gde je

$$A(y) = \{\boldsymbol{x}: y_i = g_i(\boldsymbol{x})\ za\ i = 1, \dots, m\}.$$

Ako su slučajne veličine $X_1, \dots, X_n$ diskretnog tipa, znak integrala se zamenjuje sumom po $A(y)$.

Da smo posmatrali $\boldsymbol{X} = \boldsymbol{x}$, mogli smo oceniti θ maksimiziranjem funkcije verodostojnosti kompletnih podataka $L(\theta) = f_X(\boldsymbol{x}; \theta)$. Umesto toga posmatramo $\boldsymbol{Y} = \boldsymbol{y}$. U principu, zajednička funkcija gustine za $\boldsymbol{Y}$ se teško izračunava direktno, pa će zato biti teško izračunati MLE iz nekompletnih podataka koristeći standardne algoritme. S obzirom na preliminarnu ocenu parametra, EM algoritam prvo konstruiše ocenu na osnovu funkcije verodostojnosti kompletnih podataka, a potom maksimizira tu funkciju kako bi se dobila nova ocena parametra, a ova dvoetapna procedura se ponavlja dok kriterijum za konvergenciju ne bude ispunjen.

- (E korak) E korak u EM algoritmu uključuje nalaženje ocene parametra θ pomoću funkcije verodostojnosti za kompletne podatke, s obzirom na posmatrane (ili nepotpune) podatke. Imajući u vidu ocenu za θ, $\hat{\theta}^{(k)}$, posle k iteracija algoritma, definišemo
$$\ln L^{(k)}(\theta) = E\left[\ln f_X(\boldsymbol{X};\theta)\middle|\boldsymbol{Y} = \boldsymbol{y}; \hat{\theta}^{(k)}\right]$$
gde se očekivanje uzima pod pretpostavkom da je prava vrednost parametra $\hat{\theta}^{(k)}$.

- (M korak) Ažurirane ocene za θ se dobijaju u koraku M algoritma EM. Ažurirana ocena, $\hat{\theta}^{(k+1)}$, je izabrana tako da maksimizira ocenu logaritma funkcije verodostojnosti kompletnih podataka, $\ln L^{(k)}(\theta)$, koja je dobijena u koraku E. Izračunavanje $\hat{\theta}^{(k+1)}$ će često zahtevati upotrebu neke numetičke metode kao što je Njutn-Rapson algoritam, često je korisno uzimati $\hat{\theta}^{(k)}$ kao inicijalnu ocenu za $\hat{\theta}^{(k+1)}$ u ovom slučaju. Kada se izračuna $\hat{\theta}^{(k+1)}$ vraćamo se na korak E.

Koraci E i M se ponavljaju dok kriterijum za konvergenciju ne bude ispunjen.

U mnogim slučajevima, npr. modeli iz eksponencijalne familije, $f_X(\boldsymbol{x};\theta)$ je proizvod dva nekonstantna faktora, od kojih jedan ne zavisi od θ. Dalje možemo pisati
$$\ln f_X(\boldsymbol{x};\theta) = h_1(\boldsymbol{x};\theta) + h_2(\boldsymbol{x}).$$

U takvim slučajevima, možemo redefinisati $\ln L^{(k)}(\theta)$ u koraku E kao
$$\ln L^{(k)}(\theta) = E\left[h_1(\boldsymbol{X};\theta)\middle|\boldsymbol{Y} = \boldsymbol{y}; \hat{\theta}^{(k)}\right].$$
Ova modifikacija algoritma može značajno pojednostaviti računanje. M korak algoritma ostaje isti kao pre, $\hat{\theta}^{(k+1)}$ maksimizira $\ln L^{(k)}(\theta)$.

Šta se može reći o svojstvima konvergencije algoritma EM? Ako je $L(\theta)$ funkcija verodostojnosti na osnovu $(y_1, \ldots, y_m)$, važi
$$\ln L\left(\hat{\theta}^{(k+1)}\right) \geq \ln L\left(\hat{\theta}^{(k)}\right)$$
za svaku inicijalnu ocenu $\hat{\theta}^{(0)}$. Nažalost, ovo ne mora podrazumevati da niz $\{\hat{\theta}^{(k)}\}$ konvergira ka MLE. Ovo sugeriše da izbor inicijalne ocene može biti veoma značajan. U praksi je dobra ideja uzeti nekoliko različitih inicijalnih ocena. Ako je $\hat{\theta}^{(k+1)}$ izračunata numerički u koraku M, važno je osigurati da je odgovarajuća inicijalna ocena izabrana kako bi se garantovalo da će $\hat{\theta}^{(k+1)}$ biti maksimum za $L^{(k)}(\theta)$.

Primer 33. Pretpostavimo da su $X_1, \ldots, X_n$ nezavisne slučajne veličine čija je raspodela mešavina dve Puasonove raspodele

$$f_X(x; \lambda, \theta) = \theta \frac{\exp(-\lambda)\lambda^x}{x!} + (1-\theta)\frac{\exp(-\mu)\mu^x}{x!}$$

za $x = 1,2, \ldots$ .Ova funkcija se dobija ako posmatramo slučajne veličine iz Puasonove raspodele sa matematičkim očekivanjem λ i verovatnoćom θ i slučajne veličine iz Puasonove raspodele sa matematičkim očekivanjem μ i verovatnoćom $1-\theta$.

Logaritam funkcije verodostojnosti za $\theta, \lambda\ i\ \mu$ je dat sa

$$\ln L(\theta, \lambda, \mu) = \sum_{i=1}^{n} \ln(\theta \frac{\exp(-\lambda)\lambda^{x_i}}{x_i!} + (1-\theta)\frac{\exp(-\mu)\mu^{x_i}}{x_i!}).$$

S obzirom na $(x_1, \ldots, x_n)$, moguće je oceniti parametre koristeći Njutn-Rapson algoritam, međutim, implementacija ovog algoritma u datom primeru je teška pa je prirodna alternativa EM algoritam.

Kako bismo koristili EM algoritam, moramo naći odgovarajući problem „kompletnih" podataka. Pretpostavimo da su (X_1, Y_1),..., (X_n, Y_n) parovi nezavisnih slučajnih veličina sa verovatnoćama

$$P(Y_i = y) = \theta^y (1-\theta)^{1-y} \quad za\ y = 0,1$$

$$P(X_i = x | Y = 0) = \frac{\exp(-\mu)\mu^x}{x!} \quad za\ x = 0,1,2, \ldots$$

$$P(X_i = x | Y = 1) = \frac{\exp(-\lambda)\lambda^x}{x!} \quad za\ x = 0,1,2, \ldots$$

Funkcija verodostojnosti za kompletne podatke zasnovana na (x_1, y_1),..., (x_n, y_n) ima oblik

$$\begin{aligned}\ln L_c(\theta, \lambda, \mu) = &\sum_{i=1}^{n} y_i[\ln\theta + x_i \ln\lambda - \lambda] \\ &+ \sum_{i=1}^{n} (1-y_i)[\ln(1-\theta) + x_i \ln\mu - \mu] \\ &- \textstyle\sum_{i=1}^{n} \ln x_i! \,.\end{aligned}$$

Sada moramo naći očekivanje logaritma funkcije verodostojnosti kompletnih podataka s obzirom na $(x_1, \ldots, x_n)$. Kako bismo ovo uradili dovoljno je naći očekivanje za Y_i pri uslovu $X_i = x_i$ za bilo koje vrednosti parametra $\theta, \lambda\ i\ \mu$, što će se lako izračunati korišćenjem Bajesove teoreme

$$E(Y_i | X_i = x; \theta, \lambda, \mu) = \frac{\theta \exp(-\lambda)\lambda^x}{\theta \exp(-\lambda)\lambda^x + (1-\theta)\exp(-\mu)\mu^x}.$$

Prema tome, sa datim ocenama $\hat{\theta}^{(k)}, \hat{\lambda}^{(k)}$ i $\hat{\mu}^{(k)}$ dobijamo $\hat{\theta}^{(k+1)}, \hat{\lambda}^{(k+1)}$ i $\hat{\mu}^{(k+1)}$ maksimizirajući

$$\ln L^{(k)}(\theta,\lambda,\mu) = \sum_{i=1}^{n} \hat{y}_i^{(k)}[\ln\theta + x_i \ln\lambda - \lambda] + \sum_{i=1}^{n} \left(1-\hat{y}_i^{(k)}\right)[\ln(1-\theta) + x_i \ln\mu - \mu]$$

gde je

$$\hat{y}_i^{(k)} = E\left(Y_i \middle| X_i = x; \hat{\theta}^{(k)}, \hat{\lambda}^{(k)}, \hat{\mu}^{(k)}\right).$$

Dobijamo da je

$$\hat{\theta}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i^{(k)}$$

$$\hat{\lambda}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i \hat{y}_i^{(k)}}{\sum_{i=1}^{n} \hat{y}_i^{(k)}}$$

$$\hat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i (1-\hat{y}_i^{(k)})}{\sum_{i=1}^{n} (1-\hat{y}_i^{(k)})}.$$

Kako bismo ilustrovali konvergenciju EM algoritma u ovom primeru, posmatraćemo broj postignutih golova na utakmicama Prve divizije engleske fudbalske lige tokom sezone 1978 – 79.[34] U tom periodu bilo je 22 tima, svaki tim igra 42 utakmice, što je ukupno 924 utakmica koje posmatramo.

Podaci su predstavljeni u sledećoj tabeli.

Golovi	0	1	2	3	4	5	6	7
Frekvencija	252	344	180	104	28	11	2	3

Frekvencija raspodele golova u mečevima Prve divizije

EM algoritam je jako spor u konvergenciji u ovom primeru iako su inicijalne vrednosti ocena uzete da budu jako blizu vrednostima maksimuma. Npr., počevši od inicijalnih ocena $\hat{\theta}^{(0)} = 0.95$, $\hat{\lambda}^{(0)} = 1.23$ i $\hat{\mu}^{(0)} = 3.04$, EM algoritmu treba oko 300 iteracija kako bi konvergirao ka ocenama metodom maksimalne verodostojnosti $\hat{\theta} = 0.954$, $\hat{\lambda} = 1.232$ i $\hat{\mu} = 3.043$. Može se takođe primeniti i Njutn–Rapson algoritam ali za ovaj konkretan primer algoritam dosta „luta“, koristeći iste početne vrednosti date iznad, ovaj algoritam će zapravo divergirati. ◊

[34] Podaci i rezultati preuzeti iz primera 5.25, knjige *Mathematical statistics, Keith Knight, Chapman & Hall/CRC, 2000.*

10.4 Poređenje Njutn–Rapson i EM algoritma

Često je izvodljivo koristiti i jedan i drugi algoritam za izračunavanje ocena metodom maksimalne verodostojnosti. Zato se postavlja pitanje koji algoritam je bolje koristiti. Evo dva ubedljiva teoretska razloga zašto se treba odlučiti za Njutn-Rapson algoritam:

- ocene standardne greške su „nus“ pojava Njutn-Rapson algoritma a ne EM algoritma;
- Njutn-Rapson algoritam ima bržu konvergenciju, u smislu da $\hat{\theta}^{(k)} \to \hat{\theta}$ mnogo brže kada je $\hat{\theta}^{(k)}$ u okolini od $\hat{\theta}$.

Ipak, postoje i mnogi razlozi zašto treba izabrati EM algoritam, a možda je najvažniji razlog laka implementacija. Na primer, EM algoritam štedi mnogo više vremena potrebnog za programiranje nego Njutn-Rapson algoritam ili neki drugi sofisticiraniji algoritam. Štaviše, u mnogim problemima, broj iteracija do konvergencije jednog i drugog algoritma je približno jednak, što zavisi od bliskosti problema kompletnih i nekompletnih podataka.

11. Primeri ocenjivanja u statističkom programu R

U statističkom programu R, koji se široko koristi u obradi i analizi podataka zbog svojih atraktivnih karakteristika, postoji nekoliko paketa čije funkcije možemo koristiti prilikom optimizacije funkcija verodostojnosti. U ovom odeljku biće razmatrano nekoliko takvih funkcija.

Primer A: Prvo ćemo razmatrati slučaj kada je nepoznat jedan parametar. Pretpostavimo da imamo slučajan uzorak iz Puasonove raspodele i želimo da ocenimo nepoznati parametar λ. Kako smo u prethodnim segmentima rada videli, ocenjivanje metodom maksimalne verodostojnosti počinje definisanjem funkcije verodostojnosti i nalaženjem njenog logaritma. U R-u se ovo može postići na dva načina, deklarisanjem odgovarajuće analitičke forme funkcije verodostojnosti, a potom i njenog logaritma, za datu raspodelu ili pozivanjem funkcije *dpois(x, lambda, log = FALSE)* iz paketa *stats,* koja računa gustinu za datu raspodelu. Kodovi u R-u za deklarisanje logaritma funkcije verodostojnosti su:

-
```
poisson.lik<-function(lambda,y)
 {
  n<-length(y)
  logl<-sum(y)*log(lambda)-n*lambda-sum(log(factorial(y)))
  return(logl)
 }
```

Ovim kodom smo deklarisali log-funkciju verodostojnosti kao jednu R funkciju koja za argumente prima parametar λ i podatke iz uzorka. U telu funkcije je predstavljena analitička forma log-funkcije verodostojnosti i funkcija vraća njenu vrednost u odnosu na zadati uzorak.

-
```
logl <- function(lambda, a)   sum(dpois(a,lambda, log = TRUE))
```

Pozivom funkcije dpois(a,lambda, log = TRUE) dobijamo vrednosti logaritma funkcije gustine, čijim sumiranjem potom dobijamo i log-funkciju verodostojnosti.

Sledeći kod u R-u služi za predstavljanje grafika log-funkcije verodostojnosti za ovaj primer:

```
npoint <- 101
lambdas <- seq(min(a), max(a), length = npoint)
logls <- double(npoint)
for (i in 1:npoint)
 logls[i] <- logl(lambdas[i], a)
plot(lambdas, logls, type = "l", xlab = expression(lambda), ylab ="Log-funkcija verodostojnosti")
```

Kada je definisana funkcija verodostojnosti i njen logaritam prelazimo na njihovu optimizaciju kako bismo dobili MLE. Za ovo mogu poslužiti različite funkcije, a ovde će biti razmatrane dve funkcije iz paketa *stats,* funkcije *nlm* i *optim,* i funkcija iz paketa *stats4, mle.*

- *nlm(f, p, ..., hessian = FALSE, typsize = rep(1, length(p)), fscale = 1, print.level = 0, ndigit = 12, gradtol = 1e-6, stepmax = max(1000 *sqrt(sum((p/typsize)^2)), 1000),steptol = 1e-6, iterlim = 100, check.analyticals = TRUE)* - funkcija koja minimizira funkciju f koja je prosleđena kao parameter, koristeći neku od metoda Njutnovog tipa .
- *mle(minuslogl, start = formals(minuslogl), method = "BFGS",fixed = list(), nobs, ...)* - ocenjivanje parametara metodom maksimalne verodostojnosti.
- *optim(par, fn, gr = NULL, ...,method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN", "Brent"), lower = -Inf, upper = Inf, control = list(), hessian = FALSE)* – funkcija koja se koristi u svrhe generalne optimizacije na osnovu Nelder–Mid algoritma, kvazi-Njutn i konjugovanog-gradijent algoritma.[35]

Sada ćemo pogledati primenu navedenih funkcija na konkretnom uzorku.

\# vector koji prima vrednosti 100 slučajnih veličina iz Puasonove raspodele

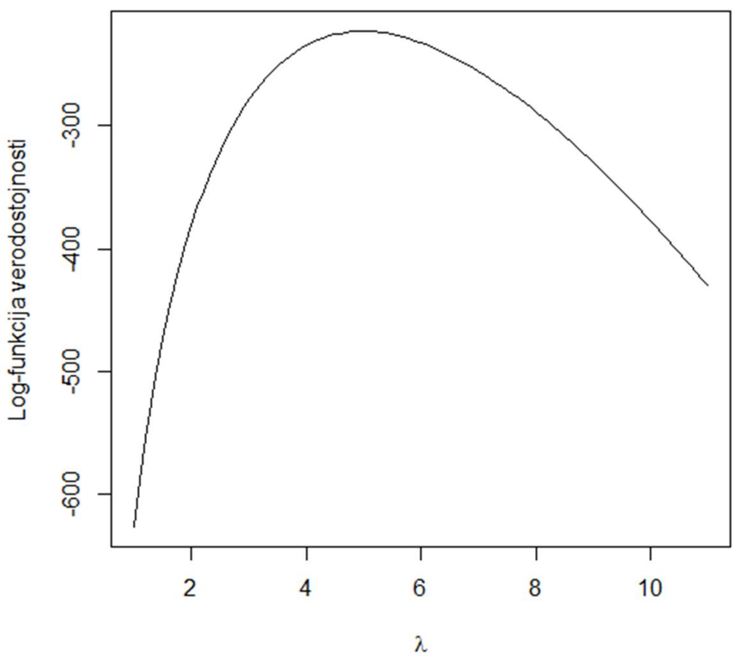
```
> a<-rpois(100,5)
```

\# inicijalizacija log-funlcije verodostojnosti

```
> logl <- function(lambda,a)    sum(dpois(a,lambda, log = TRUE))
```

\# grafik log-funkcije verodostojnosti

```
> npoint <- 101
> lambdas <- seq(min(a), max(a), length = npoint)
> logls <- double(npoint)
> for (i in 1:npoint)
+    logls[i] <- logl(lambdas[i], a)
> plot(lambdas, logls, type = "l",xlab = expression(lambda), ylab ="Log-funkcija verodostojnosti")
```



[35] Iako ove metode optimizacije nisu opisane u radu, u nekim primerima je korišćena funkcija optim, u želji da se prikaže raznolikost funkcija za optimizaciju i računanje MLE koje postoje u R-u.

definisanje $-lnL$ [36]
```
> mlogl <- function(lambda,a)   -sum(dpois(a,lambda, log = TRUE))
```

poziv funkcije nlm
prvi parametar koji prima f-ja nlm je funkcija koju treba minimizirati
drugi parametar je inicijalna vrednost ocene našeg parametra, pa ćemo u tu svrhu
koristiti uzoračku sredinu,koja je ocena metodom momenta
parametar hessian = TRUE vraća drugi izvod funkcije koju treba minimizirati, to je
matrica formata $k \times k$, gde je k dimenzija parametarskog prostora
parametar fscale pomaže funkciji nlm da odredi da li je konvergirala do rešenja
```
> out <- nlm(mlogl, mean(a),a,hessian = TRUE,fscale = length(a))
> print(out)
```

izlaz
```
$minimum
[1] 223.0609
$estimate
[1] 4.99
$gradient
[1] -3.333428e-07
$hessian
        [,1]
[1,] 20.03607
$code
[1] 3
$iterations
[1] 1
```

Zaljučujemo da je ocenjena vrednost parametra jako bliska stvarnoj vrednosti parametra, a izlaz parametra hessian pomaže da utvrdimo da li je dobijeno rešenje lokalni minimum, tj. maksimum, a što je mnogo važnije služi kao ocena za posmatranu informacionu funkciju Fišera.

poziv funkcije mle
```
> mlogl1 <- function(lambda)   -sum(dpois(a,lambda, log = TRUE))
> mle(mlogl1,start = list(lambda=mean(a)),nobs = length(a))
```

izlaz
```
Call:
mle(minuslogl = mlogl1, start = list(lambda = mean(a)), nobs = length(a))
Coefficients:
lambda
  4.99
```

[36] Funkcije za optimizaciju koje su navedene zapravo minimiziraju zadatu funkciju, a kako je nama za MLE potrebno maksimizirati lnL, to postižemo minimiziranjem –lnL.

inicijalizacija log-funkcije verodostojnosti analitičkim putem

```
> data<-a
> poisson.lik<-function(lambda,y)
+ {
+   n<-length(y)
+   logl<-sum(y)*log(lambda)-n*lambda-sum(log(factorial(y)))
+   return(logl)
+ }
```

poziv funkcije optim
poslednji parametar u ovoj funkciji podešen je tako da maksimizira prosleđenu f-ju

```
> optim(mean(data),poisson.lik,y=data,method="BFGS",control=list(fnscale=-1))
```

izlaz

```
$par
[1] 4.99
$value
[1] -223.0609
$counts
function gradient
      15        1
$convergence
[1] 0
$message
NULL
```

računanje standardne greške ocene

```
> as<-optim(mean(data),poisson.lik,y=data,hessian=TRUE,method="BFGS",
            control=list(fnscale=-1))
> OI<-solve(-1*as$hessian)
> se<-sqrt(diag(OI))
> se
```

izlaz

```
[1] 0.2233831
```

Primer B: U ovom primeru ćemo posmatrati podatke iz gama raspodele i oceniti nepoznate parametre α i λ. Podaci su preuzeti sa http://www.stat.umn.edu/geyer/5102/data/ex3-1.txt. U ovom primeru glavna razlika je zahtev da se ocene dva parametra, pa će funkcija koja definiše log-funkciju verodostojnosti primati kao argument vektor koji se sastoji od parametara koje želimo da ocenimo.

učitavanje podataka iz tekstualnog fajla
```
> pg<-read.table("C:/as1/pg.txt",header=T)
```

iščitavanje podataka
```
> pg
      x
1  1.99
2  1.46
3  2.01
4  0.18
5  4.35
6  0.30
7  0.34
8  0.91
9  0.24
10 0.27
11 0.82
12 5.86
13 1.63
14 1.35
15 1.26
16 1.40
17 2.59
18 1.34
19 0.70
20 1.39
21 1.40
22 0.83
23 2.87
24 3.62
25 0.52
26 4.15
27 1.72
28 1.23
29 1.60
30 1.61
```

definisanje f-je koja će računati log-funkciju verodostojnosti za ovaj primer
```
> mlogl <- function(theta, x) {
+     if (length(theta) != 2) stop("Theta mora biti vektor duzine 2")
+     alpha <- theta[1]
+     lambda <- theta[2]
+     if (alpha <= 0) stop("theta[1] mora biti pozitivno")
+     if (lambda <= 0) stop("theta[2] mora biti pozitivno ")
+     return(- sum(dgamma(x, shape = alpha, rate = lambda, log = TRUE)))
+ }
```

definisanje inicijalnih vrednosti za ocene željenih parametara
u tu svrhu ćemo koristiti ocene metodom momenta

```
> alpha.start <- mean(pg$x)^2 / var(pg$x)
> lambda.start <- mean(pg$x) / var(pg$x)
> theta.start <- c(alpha.start, lambda.start)
```

poziv f-je nlm za optimizaciju

```
> out <- nlm(mlogl, theta.start, x = pg$x, hessian = TRUE,
+    fscale = length(pg$x))
> print(out)
```

izlaz

```
$minimum
[1] 43.19981
$estimate
[1] 1.680531 1.009529     # ocenjene vrednosti za α i λ.
$gradient
[1]  1.647685e-05 -2.484541e-05
$hessian
          [,1]          [,2]
[1,]  24.15181  -29.71534
[2,] -29.71534  49.45877
$code
[1] 1
$iterations
[1] 5
```

vidimo da je ocene parametara dobijena posle pet iteracija
ukoliko želimo da pogledamo rezultate svake iteracije do konačne konvergencije
pozivamo f-ju nlm sa dodatkom argumenta *print.level* = 2

```
> out <- nlm(mlogl, theta.start, x = pg$x, hessian = TRUE,
+    fscale = length(pg$x),print.level = 2)
```

izlaz

```
iteration = 0
Step:
[1] 0 0
Parameter:
[1] 1.5398404   0.9250142
Function Value
[1] 43.26619
Gradient:
[1] -9.740582e-01  2.698641e-05

iteration = 1
Step:
[1]  3.663168e-02   -1.014885e-06
```

Parameter:
[1] 1.5764720   0.9250132
Function Value
[1] 43.2485
Gradient:
[1] 0.002880992   -1.188064552

iteration = 2
Step:
[1] 0.05390210   0.04441225
Parameter:
[1] 1.6303741   0.9694255
Function Value
[1] 43.21047
Gradient:
[1] -0.01951597   -0.51380027

iteration = 3
Step:
[1] 0.04626223   0.03693969
Parameter:
[1] 1.676636   1.006365
Function Value
[1] 43.19987
Gradient:
[1] -0.0000313266   -0.0409296108

iteration = 4
Step:
[1] 0.003755283   0.003045666
Parameter:
[1] 1.680392   1.009411
Function Value
[1] 43.19981
Gradient:
[1] 0.0001639238   -0.0017317164

iteration = 5
Parameter:
[1] 1.680531   1.009529
Function Value
[1] 43.19981
Gradient:
[1] 1.647685e-05   -2.484541e-05

Relative gradient close to zero.
Current iterate is probably solution.

inicijalne vrednosti ocena parametara

```
> print(theta.start)
```

izlaz

```
[1] 1.5398404   0.9250142
```

računanje sopstvenih vrednosti matrice hessian

```
> eigen(out$hessian, symmetric = TRUE, only.values = TRUE)
```

izlaz

```
$values
[1] 69.102539  4.508044
$vectors
NULL
```

računanje posmatrane informacione matrice Fšera

```
> n <- length(pg$x)
> alpha.hat <- out$estimate[1]
> lambda.hat <- out$estimate[2]
> fish <- n * matrix(c(trigamma(alpha.hat), - 1 / lambda.hat,
+      - 1 / lambda.hat, alpha.hat / lambda.hat^2), nrow = 2)
> fish
```

izlaz

```
         [,1]        [,2]
[1,]  24.15495   -29.71683
[2,] -29.71683    49.46866
```

Zaključujemo da su ocene dobijene metodom maksimalne verodostojnosti jako bliske ocenama po metodi momenta, koje smo uzeli kao inicijalne vrednosti. Činjenica da su sve sopstvene vrednosti matrice drugih parcijalnih izvoda od $-lnL$ pozitivne, ukazuje da su dobijene MLE lokalni maksimum funkcije verodostojnosti.

Primer C: U ovom primeru ćemo posmatrati podatke iz mešovite normalne raspodele sa funkcijom gustine zadatom kao

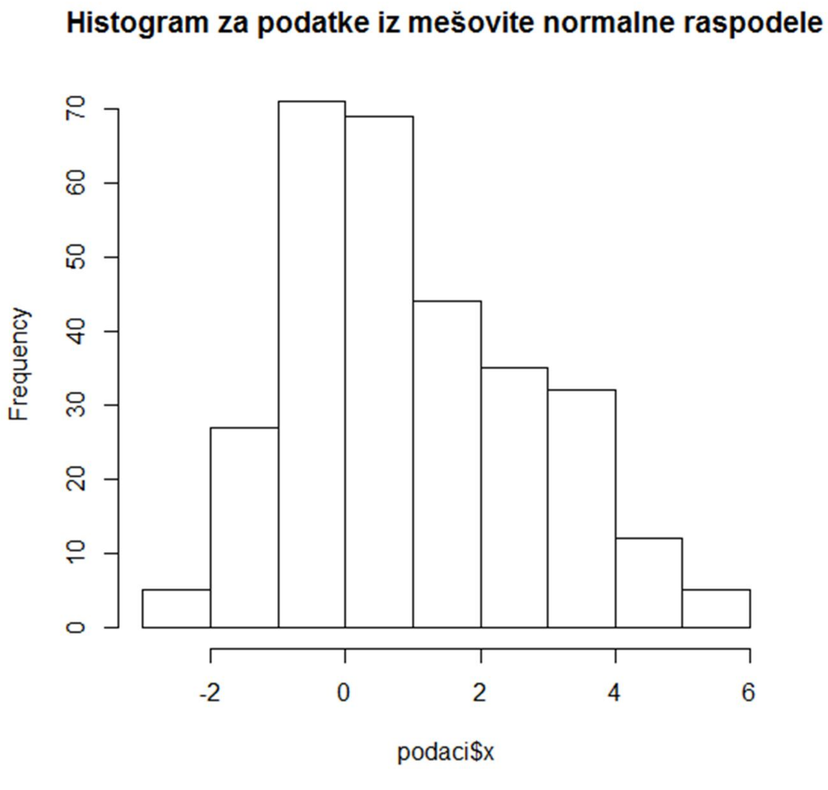
$$f(x;\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,p) = p\varphi(x;\mu_1,\sigma_1^2) + (1-p)\varphi(x;\mu_2,\sigma_2^2)$$

gde $\varphi(x;\mu,\sigma^2)$ označava normalnu raspodelu sa matematičkim očekivanjem μ i disperzijom σ^2. Napomenimo ovde da se, kada je u pitanju mešavina raspodela, pa čak i normalnih, javlja problem prilikom računanja MLE, jer metoda maksimalne verodostojnosti dovodi do jednačine koja se ne može rešiti bez primene metoda numeričke analize. Podaci koje koristimo kao uzorak su preuzeti sa http://www.stat.umn.edu/geyer/5102/data/mix.txt. U ovom primeru je potrebno naći ocene za čak pet nepoznatih parametara $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p$, pa će funkcija za računanje logaritma funkcije verodostojnosti kao argument primati vektor dužine pet.

Glavni problem u ovom primeru je kako odabrati dobre inicijalne vrednosti za ocene parametara koje želimo da izračunamo. Pošto u ovom slučaju ne postoji metod za ocenjivanje inicijalne vrednosti koji je jednostavan, moramo pribeći gruboj oceni inicijalnih vrednosti. Za parametar p uzećemo kao inicijalnu vrednost $p = 1/2$, a zatim ćemo podatke iz uzorka podeliti na dva dela i iz svakog dela naći srednju vrednost i disperziju, kao inicijalne vrednosti za ocene ostalih parametara.

```
# učitavanje podataka iz uzorka
 > podaci<-read.table("C:/as/podaci.txt",header=T)
```

```
# osnovne informacije o podacima iz uzorka
> length(podaci$x)
[1] 300
> summary(podaci$x)
   Min.    1st Qu.   Median    Mean     3rd Qu.    Max.
-2.9930  -0.3162   0.6282   0.9782   2.2990    5.4600
> hist(podaci$x, main = paste("Histogram za podatke iz mešovite normalne raspodele" ))
```




```
# definisanje f-je koja će računati log-funkciju verodostojnosti za ovaj primer
> mlogl <- function(theta) {
+     stopifnot(is.numeric(theta))
+     stopifnot(length(theta) == 5)
+     mu1 <- theta[1]
+     mu2 <- theta[2]
+     v1 <- theta[3]
```

```
+     v2 <- theta[4]
+     p <- theta[5]
+     logl <- sum(log(p * dnorm(podaci$x, mu1, sqrt(v1)) +
+         (1 - p) * dnorm(podaci$x, mu2, sqrt(v2))))
+     return(- logl)
+ }
```

definisanje inicijalnih vrednosti za ocene željenih parametara

```
> n <- length(podaci$x)
> p.start <- 1 / 2
> mu1.start <- mean(sort(podaci$x)[seq(along = podaci$x) <= n / 2])
> mu2.start <- mean(sort(podaci$x)[seq(along = podaci$x) > n / 2])
> v1.start <- var(sort(podaci$x)[seq(along = podaci$x) <= n / 2])
> v2.start <- var(sort(podaci$x)[seq(along = podaci$x) > n / 2])
> theta.start <- c(mu1.start, mu2.start, v1.start,
+     v2.start, p.start)
```

poziv f-je nlm za optimizaciju

```
> out <- nlm(mlogl, theta.start, print.level = 2,
+     fscale = length(podaci$x))
```

izlaz

```
iteration = 0
Step:
[1] 0 0 0 0 0
Parameter:
[1] -0.4345200  2.3909647  0.5340638  1.4761858  0.5000000
Function Value
[1] 578.9291
Gradient:
[1] -24.221222   3.829967 -33.645415  -7.743632 -10.308731

iteration = 1
Step:
[1]  0.37395121 -0.05913083  0.51945123  0.11955386  0.15915639
Parameter:
[1] -0.06056879  2.33183384  1.05351507  1.59573967  0.65915639
Function Value
[1] 576.888
Gradient:
[1]   1.528432 -22.022011   4.471106  -3.801064  62.280484

iteration = 2
Step:
[1]  0.03730838  0.02319929  0.04868772  0.01765874 -0.06645932
Parameter:
[1] -0.02326041  2.35503313  1.10220279  1.61339841  0.59269708
```

Function Value
[1] 574.9797
Gradient:
[1] 11.425042 -14.374406   7.829206 -4.913388 -1.323404

iteration = 3
Step:
[1] -0.132932732 0.142319134 -0.108313640 0.039354175 -0.002508294
Parameter:
[1] -0.1561931 2.4973523 0.9938892 1.6527526 0.5901888
Function Value
[1] 571.9802
Gradient:
[1] -3.5139701 -5.6512817 8.1667974 -0.4425899 1.9574198

iteration = 4
Step:
[1] -0.012697183 0.119567677 -0.177571370 0.012003020 0.005383833
Parameter:
[1] -0.1688903 2.6169199 0.8163178 1.6647556 0.5955726
Function Value
[1] 571.1776
Gradient:
[1] -2.5609788 0.1576641 -3.8129466 2.4040823 6.4208575

iteration = 5
Step:
[1] 0.023837809 -0.016034800 0.055663895 -0.014519101 -0.002972054
Parameter:
[1] -0.1450525 2.6008851 0.8719817 1.6502365 0.5926006
Function Value
[1] 571.028
Gradient:
[1] -0.1755207 -1.1534099 0.4704294 1.6781086 -1.0286218

iteration = 6
Step:
[1] 0.003988521 0.015186311 -0.001574308 -0.020233284 0.002246955
Parameter:
[1] -0.1410640 2.6160715 0.8704074 1.6300032 0.5948475
Function Value
[1] 570.974
Gradient:
[1] 0.005483002 -0.978518212 0.131819320 1.711024473 -1.387966222

iteration = 7
Step:
[1] 0.05172824 0.19493324 0.01018253 -0.28944290 0.03490704

Parameter:
[1] -0.08933575 2.81100470 0.88058990 1.34056032 0.62975455
Function Value
[1] 570.3688
Gradient:
[1] -0.06588334 0.60937831 -2.73051728 1.52674672 -1.66339976

iteration = 8
Step:
[1] 0.03482953 0.10138775 0.02156434 -0.16146705 0.02012846
Parameter:
[1] -0.05450622 2.91239245 0.90215424 1.17909327 0.64988302
Function Value
[1] 570.1614
Gradient:
[1] -0.2302532 1.1483486 -3.7756233 0.7723296 -0.9828010

iteration = 9
Step:
[1] 0.04923960 0.09283847 0.05541817 -0.15747469 0.02155543
Parameter:
[1] -0.005266622 3.005230918 0.957572409 1.021618587 0.671438450
Function Value
[1] 570.0279
Gradient:
[1] 0.4934329 0.5429539 -2.9173082 -0.7397089 -1.0028978

iteration = 10
Step:
[1] 0.0040748312 -0.0137530487 0.0269834340 0.0280006296 -0.0001847823
Parameter:
[1] -0.001191791 2.991477869 0.984555843 1.049619216 0.671253668
Function Value
[1] 569.9647
Gradient:
[1] 0.7560124 -0.1905389 -0.8710482 -0.2063503 -1.1811013

iteration = 11
Step:
[1] 0.003946482 0.015422609 0.019676636 -0.010989239 0.004285831
Parameter:
[1] 0.00275469 3.00690048 1.00423248 1.03862998 0.67553950
Function Value
[1] 569.9543
Gradient:
[1] -0.152237021 -0.071251794 -0.017372286 -0.088242346 -0.006700247

iteration = 12

```
Step:
[1]  0.0023200693  0.0026906300  0.0024528727 -0.0004125557  0.0007165602
Parameter:
[1] 0.00507476 3.00959111 1.00668535 1.03821742 0.67625606
Function Value
[1] 569.954
Gradient:
[1] -0.005453671 -0.048882765  0.002160838 -0.014976801  0.121645712

iteration = 13
Step:
[1]  1.764352e-04  8.164357e-04  2.715500e-04  3.803224e-06 -3.280708e-05
Parameter:
[1] 0.005251195 3.010407543 1.006956902 1.038221224 0.676223251
Function Value
[1] 569.954
Gradient:
[1] -0.0026469706 -0.0011596753  0.0090737721 -0.0005088537 -0.0186598754

iteration = 14
Step:
[1]  2.096708e-05  4.962925e-05 -9.017388e-05 -4.842425e-05  1.969806e-05
Parameter:
[1] 0.005272162 3.010457173 1.006866728 1.038172800 0.676242949
Function Value
[1] 569.954
Gradient:
[1]  0.0001388116 -0.0001973923  0.0005436689  0.0001303129  0.0003213927

iteration = 15
Parameter:
[1] 0.005267125 3.010454061 1.006854273 1.038172543 0.676241202
Function Value
[1] 569.954
Gradient:
[1]  2.501110e-06 -8.005972e-06  2.596997e-06 -3.285201e-07  2.989964e-05

Relative gradient close to zero.
Current iterate is probably solution.

Warning messages:
1: In log(p * dnorm(podaci$x, mu1, sqrt(v1)) + (1 - p) * dnorm(podaci$x,  :
  NaNs produced
2: In nlm(mlogl, theta.start, print.level = 2, fscale = length(podaci$x)) :
  NA/Inf replaced by maximum positive value
```

opet pozivamo f-ju nlm, ali ovog puta malo modifikovanu, da vidimo da li je dala
tačno rešenje iako imamo poruke upozorenja

```
> out <- nlm(mlogl, out$estimate, print.level = 2,
+     fscale = length(podaci$x), hessian = TRUE)
```

izlaz

```
iteration = 0
Parameter:
[1] 0.005267125 3.010454061 1.006854273 1.038172543 0.676241202
Function Value
[1] 569.954
Gradient:
[1]  2.501110e-06 -8.005972e-06  2.596997e-06 -3.285201e-07  2.989964e-05
Relative gradient close to zero.
Current iterate is probably solution.
> print(out)
```

izlaz

```
$minimum
[1] 569.954
$estimate
[1] 0.005267125 3.010454061 1.006854273 1.038172543 0.676241202
$gradient
[1]  2.501110e-06 -8.005972e-06  2.596997e-06 -3.285201e-07  2.989964e-05
$hessian
           [,1]       [,2]       [,3]      [,4]       [,5]
[1,]  158.807393 -24.139843 -32.853691  5.586298 -103.42268
[2,]  -24.139843  63.447290  -9.839708 18.661225  -84.43796
[3,]  -32.853691  -9.839708  68.430939 -2.357628  -65.80344
[4,]    5.586298  18.661225  -2.357628 28.203433   38.00752
[5,] -103.422678 -84.437959 -65.803437 38.007516 1079.07185
$code
[1] 1
$iterations
[1] 0
```

štampamo inicijalne ocene za parametre

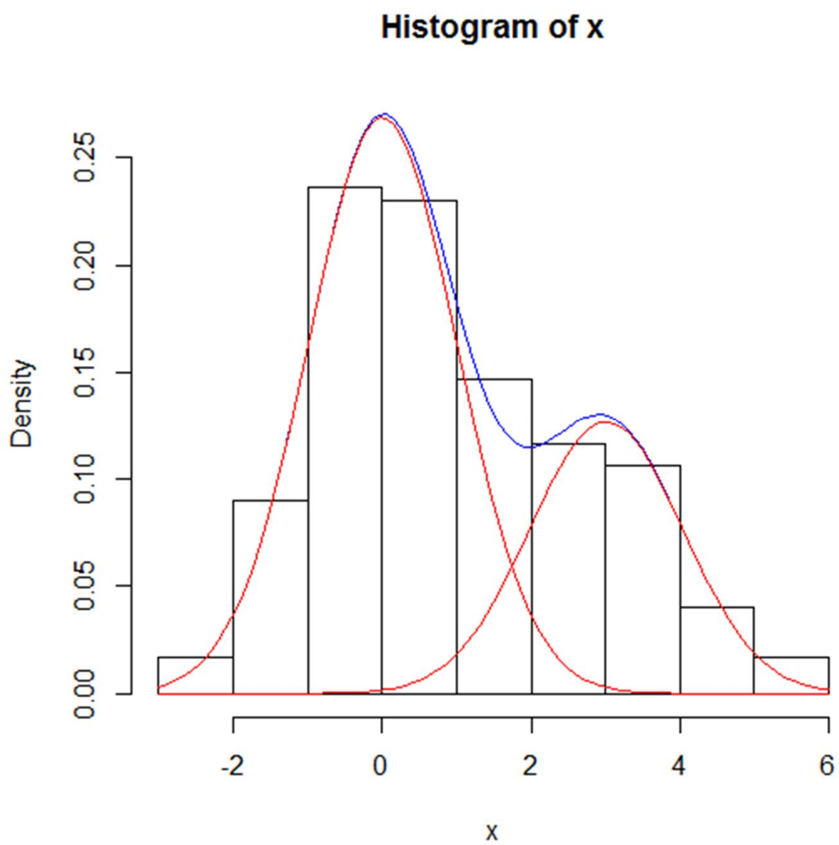
```
> print(theta.start)
[1] -0.4345200  2.3909647  0.5340638  1.4761858  0.5000000
```

računanje sopstvenih vrednosti matrice hessian

```
> eigen(out$hessian, symmetric = TRUE)
$values
[1] 69.102539  4.508044
```

$vectors
```
           [,1]       [,2]
[1,] -0.5514608 -0.8342008
[2,]  0.8342008 -0.5514608
```

\# crtanje histograma posmatranih podataka na osnovu ocenjenih parametara
```
> mu1.hat <- out$estimate[1]
> mu2.hat <- out$estimate[2]
> sigma1.hat <- sqrt(out$estimate[3])
> sigma2.hat <- sqrt(out$estimate[4])
> p.hat <- out$estimate[5]
> x<-podaci$x
> fred <- function(x) p.hat * dnorm(x, mu1.hat, sigma1.hat) +
+     (1 - p.hat) * dnorm(x, mu2.hat, sigma2.hat)
> hist(x, freq = FALSE,
+     ylim = range(0, fred(mu1.hat), fred(mu2.hat)))
> curve(fred, add = TRUE,col = "blue")
> curve(p.hat * dnorm(x, mu1.hat, sigma1.hat),
+     add = TRUE, col = "red")
> curve((1 - p.hat) * dnorm(x, mu2.hat, sigma2.hat),
+     add = TRUE, col = "red")
```
\# izlaz



Ovaj histogram predstavlja podatke iz uzorka, dok je plavom krivom iscrtan grafik funkcije gustine mešovite normalne raspodele na osnovu dobijenih vrednosti za ocene parametara, a crvenim krivama su predstavljeni grafici gustine normalnih raspodela koje daju ovu mešovitu normalnu raspodelu, opet pomoću ocenjenih parametara.

Važni rezultati ovog primera su MLE za posmatrane parametre:

```
$estimate
[1] 0.005267125   3.010454061   1.006854273   1.038172543   0.676241202
```

i posmatrana informaciona matrica Fišera

```
$hessian
           [,1]          [,2]          [,3]         [,4]          [,5]
[1,]  158.807393 -24.139843 -32.853691  5.586298 -103.42268
[2,]  -24.139843  63.447290  -9.839708 18.661225  -84.43796
[3,]  -32.853691  -9.839708  68.430939 -2.357628  -65.80344
[4,]    5.586298  18.661225  -2.357628 28.203433   38.00752
[5,] -103.422678 -84.437959 -65.803437 38.007516 1079.07185
```

Činjenica da su sve sopstvene vrednosti matrice drugih parcijalnih izvoda od $-lnL$ pozitivne, ukazuje da su dobijene MLE lokalni maksimum funkcije verodostojnosti.

Primer D: U ovom primeru ćemo razmotriti dve funkcije iz paketa *maxLik,* koji se koristi za ocenjivanje metodom maksimalne verodostojnosti. Prva funkcija je *maxLik(),* a razmatraćemo ocenjivanje parametra kod eksponencijalnog i normalnog modela.

```
# generišemo slučajne veličine iz eksponencijalnog modela
> t <- rexp(100, 2)
> summary(t)
   Min.     1st Qu.    Median     Mean     3rd Qu.    Max.
0.005007   0.164900   0.410300   0.521100   0.749100   2.153000

# definišemo analitički log-funkciju verodostojnosti
> loglik <- function(theta)   log(theta) - theta*t

# poziv funkcije maxLik()
> a <- maxLik(loglik, start=1, print.level=2)

# izlaz
----- Initial parameters: -----
fcn value: -52.10586
```

```
    parameter initial   gradient       free
[1,]        1           47.89414        1
Condition number of the (active) hessian: 1
-----Iteration 1 -----
-----Iteration 2 -----
-----Iteration 3 -----
-----Iteration 4 -----
-----Iteration 5 -----
--------------
gradient close to zero
5  iterations
estimate: 1.91917
Function value: -34.81072

> print( a )
# izlaz
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 1: gradient close to zero
Log-Likelihood: -34.81072 (1 free parameter(s))
Estimate(s): 1.91917
```

Vidimo da je ocenjena vrednost parametra 1.91917 jako bliska zadatoj vrednosti parametra.

Sada ćemo istu funkciju upotrebiti na primeru uzorka iz normalne raspodele, kada treba oceniti dva nepoznata parametra.

```
# definisanje log-funkcije verodostojnosti za normalnu raspodelu
> loglik <- function(param) {
+ mu <- param[1]
+ sigma <- param[2]
+ ll <- -0.5*N*log(2*pi) - N*log(sigma) - sum(0.5*(x - mu)^2/sigma^2)
+ ll
+ }
```

\# vektor x koji prima 1000 slučajnih veličina iz normalne raspodele $N(1,2)$

```
> x <- rnorm(1000, 1, 2)
> N <- length(x)

# poziv funkcije maxLik
> res <- maxLik(loglik, start=c(0,1))
> print( res )

# izlaz
Maximum Likelihood estimation
Newton-Raphson maximisation, 8 iterations
```

```
Return code 1: gradient close to zero
Log-Likelihood: -2120.295 (2 free parameter(s))
Estimate(s): 1.074895  2.016486
```

Druga funkcija *maxNR()* koristi Njutn-Rapson algoritam za nalaženje ocena metodom maksimalne verodostojnosti. Koristićemo pethodno generisane podatke, i za eksponencijalni i za normalni model.

```
# definicija log-funkcije verodostojnosti za eksponencijalnu raspodelu
> loglik1 <- function(theta)  sum(log(theta) - theta*t)

# poziv funkcije maxNR
> az <- maxNR(loglik1, start=1, print.level=2)
----- Initial parameters: -----
fcn value: -52.10586
    parameter initial      gradient     free
[1,]         1                   47.89414     1
Condition number of the (active) hessian: 1
-----Iteration 1 -----
-----Iteration 2 -----
-----Iteration 3 -----
-----Iteration 4 -----
-----Iteration 5 -----
--------------
gradient close to zero
5  iterations
estimate: 1.91917
Function value: -34.81072

>  summary(az)
--------------------------------------------
Newton-Raphson maximisation
Number of iterations: 5
Return code: 1
gradient close to zero
Function value: -34.81072
Estimates:
     estimate      gradient
[1,]  1.91917 -7.105427e-09
--------------------------------------------

# poziv funkcije maxNR za podatke iz normalne raspodele
> res <- maxNR(loglik, start=c(0,1))
> summary(res)
--------------------------------------------
Newton-Raphson maximisation
```

Number of iterations: 8
Return code: 1
gradient close to zero
Function value: -2120.295
Estimates:
```
     estimate      gradient
[1,] 1.074895  4.547474e-07
[2,] 2.016486 -4.547474e-07
```

Primer E. U ovom primeru ćemo praktično pokazati Bajesovu teoriju odlučivanja i to u slučaju kada apriorna i aposteriorna raspodela nisu konjugovane. Ovom problematikom smo se teorijski bavili u primerima 25. i 26.

Dakle, pretpostavimo da imamo uzorak iz eksponencijalne raspodele sa parametrom λ i neka je apriorna raspodela za parametar Košijeva. Kako bismo odredili aposteriornu raspodelu prvo je potrebno proceniti integral do koga smo došli u primeru 26. Koristićemo metodu prikazanu u primeru 25.

\# generišemo slučajan uzorak iz eksponencijalne raspodele obima $n = 20$
```
> b<-rexp(20)
> b
 [1] 1.25499468 0.44387814 1.41638486 0.23568402 0.96406637 1.03296862
 [7] 3.44956043 2.06659516 0.04275698 0.36570474 0.12331841 0.28435676
[13] 4.02601434 1.49758777 0.79198875 0.35379120 0.75188038 0.10150570
[19] 1.41045972 0.65071314
```

\# generišemo niz slučajnih veličina iz normalne raspodele $N(0,1)$, koji će nam
\# pomoći prilikom računanja procene integrala
```
> a<-rnorm(20)
> a
 [1] -1.0159827 -0.4216357  0.5704563  0.4300427 -1.1700024 -0.3016150
 [7]  0.4455039 -0.9571611  0.1070569 -0.1366743 -1.2469396  2.1328264
[13]  1.2451090 -1.3345625  1.3986234  0.5144300  0.8699490 -0.9310695
[19]  0.3952445 -1.1690328
```

\# sledeća funkcija daje procenu približne vrednosti integrala
```
>  procena<-function()
{
+  sum=0
+  for(i in 1:length(a))
{
+  g<-(dcauchy(a[i])*dexp(b[i]))/dnorm(a[i])
+ sum=sum+g
+ }
+ cat("Priblizna vrednost integrala je:",(1/length(a))*sum,"\n")
+ }
```

poziv funkcije i izlaz

> procena()

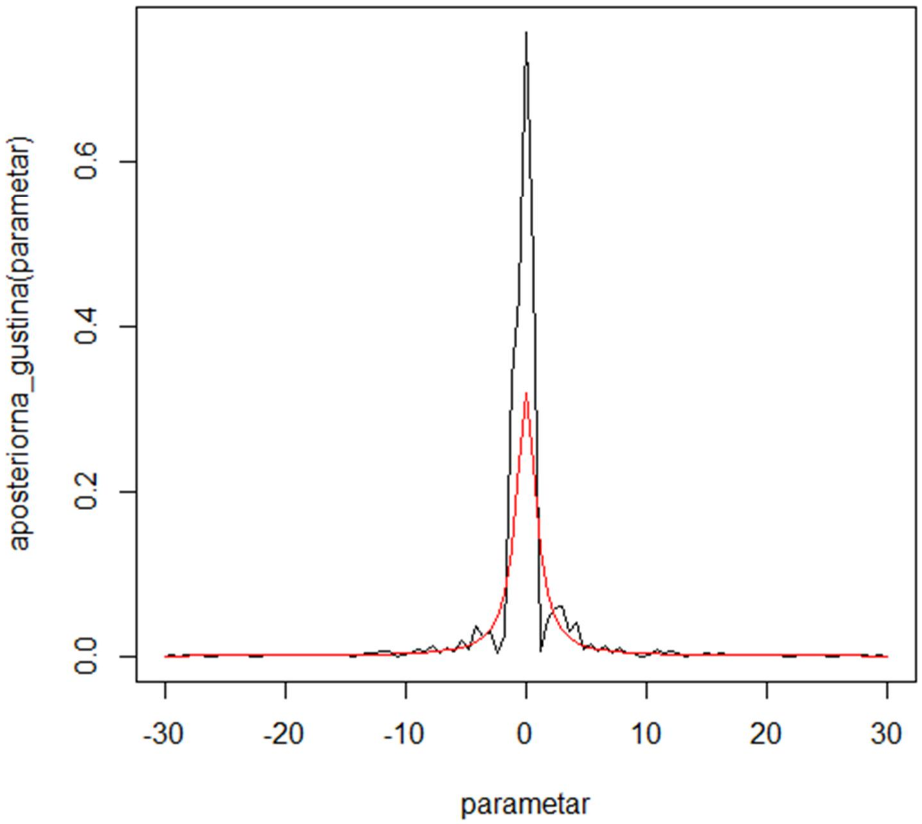
Priblizna vrednost integrala je: 0.3725418

#kada smo procenili integral, grafički ćemo predstaviti apriornu i aposteriornu #funkciju gustine

> aposteriorna_gustina<-function(x) (dexp(b)(1/(pi*(1+x^2))))/0.3725418*

> curve(aposteriorna_gustina,from = -30, to =30, lwd = 1, xname="parametar")

> curve(1/(pi(1+x^2)), add = TRUE, col = "red")*



Na grafiku je crvenom krivom predstavljena apriorna gustina za naš parametar, dok je crnom krivom predstavljena aposteriorna gustina.

12. Zaključak

Videli smo koliki značaj ima funkcija verodostojnosti i uopšte princip verodostojnosti u matematičkoj statistici, a ovaj rad se posebno osvrnuo na metod maksimalne verodostojnosti i Bajesov metod i njegovu primenu za dobijanje ocena parametara. Zaključili smo da ocene dobijene metodom maksimalne verodostojnosti imaju mnoga atraktivna svojstva kada se radi o uzorku velikog obima. Govorili smo o asimptotskoj postojanosti ovih ocena, što znači da sa povećanjem obima uzorka, ocena konvergira ka pravoj vrednosti. One su i asimptotski efikasne i nepristrasne, što znači da je za uzorak velikog obima ocena ovom metodom jako precizna. Videli smo i da ukoliko je uzorak dovoljno velikog obima, ove ocene imaju normalnu raspodelu. Iz svega se može zaključiti da ocene metodom maksimalne verodostojnosi za uzorke velikog obima imaju odlična svojstva, te se tehnika maksimalne verodostojnosti puno koristi (npr. u ekonometriji).
Glavna zamerka Bajesovoj metodi jeste subjektivnost prilikom izbora apriorne raspodele za parametar.

Nažalost, veličina uzorka koja je potrebna da bi se postigla željena svojstva mora da bude zaista velika. Sa problemima pri ocenjivanju parametara ovom metodom možemo se sresti kod raznih tipova raspodela, pa bismo se zato trebali pridržavati nekih opštih saveta kad određujemo ocene parametara:

- treba da tačno znamo šta ocenjujemo,
- treba da utvrdimo preciznost koja je dopustiva u našem istraživanju, ili preciznost koju uopšte možemo dobiti, kao i to gde ćemo primenjivati dobijeni rezultat,
- treba se uveriti u saglasnost izabranog postupka i podataka kojima raspolažemo,
- treba odrediti neku grubu, ali sigurnu procenu rezultata,
- treba popravljati ocenu, ostajući u granicama koje smo prethodno ustanovili,
- ne treba prihvatati ocene koje se značajno menjaju ako izbacimo nekoliko podataka iz uzorka,
- ako se koriste tvrđenja u kojima treba da obim uzorka teži $+\infty$, onda ih ne možemo primeniti na uzorke malog obima,
- ne treba preterivati sa brojem parametara u modelu, jer iako ćemo sa većim brojem parametara lakše opisati proučavanu pojavu, za posledicu dobijamo problem pri određivanju ocena tih parametara.

Literatura:

1. Keith Knight, Mathematical statistics, Chapman & Hall/ CRC, 2000.
2. Robert V. Hogg, Joseph W.McKean, Allen T.Craig, Introduction to Mathematical Statistics, Pearson Education, Inc.New Jersey, USA, 2005.
3. Peter Kongdon, Bayesian Statistical Modelling, John Wiley & Sons, Ltd, 2006.
4. Stevan M. Stojanović, Matematička statistika, Naučna knjiga, Beograd, 1980.
5. Zoran A. Ivković, Teorija verovatnoća sa metematičkom statistikom, Građevinska knjiga, Beograd, 1976.
6. Vesna Jevremović, Jovan Mališić, Statističke metode u meteorologiji i inženjerstvu, Savezni hidrometeorološki zavod, Beograd, 2002.
7. Internet:
   - http://www.weibull.com/hotwire/issue33/relbasics33.htm
   - http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Readings/Greene-Chapter-16.pdf
   - http://www.maths.bris.ac.uk/~MAZJCR/ToI/minSuffStat.pdf
   - http://www.stats.ox.ac.uk/~reinert/stattheory/theoryshort09.pdf
   - http://www.cc.gatech.edu/~lebanon/notes/sufficiency.pdf
   - http://www2.math.umd.edu/~slud//s705/LecNotes/Sec3NotF09.pdf
   - http://www.cran.r-project.org/
   - ftp://rm.mirror.garr.it/mirrors/CRAN/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf
   - http://www.ms.uky.edu/~mai/sta321/MLEexample.pdf
   - http://www.docstoc.com/docs/119953547/Notes-on-R-Language-for-Statistical-Computing
   - http://www.unc.edu/~monogan/computing/r/MLE_in_R.pdf

Sadržaj