Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij
matematike

Antonio Nuić

Bayesovsko statističko zaključivanje

Diplomski rad

Osijek, 2019.

Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni preddiplomski studij
matematike

Antonio Nuić
Bayesovsko statističko zaključivanje
Diplomski rad

Voditelj: prof.dr.sc. Mirta Benšić

Osijek, 2019.

Sadržaj

Uvod

Jedna od osnovnih razlika bayesovske u odnosu na klasičnu statistiku je da se u bayesovskoj statistici i parametri u statističkom modelu modeliraju kao slučajne varijable, odnosno vektori, dok se u klasičnom pristupu tretiraju kao fiksne vrijednosti. U bayesovskoj statistici i prije samog promatranja podataka imamo neka vjerovanja o parametru koja su iskazana u inicijalnoj, odnosno apriornoj, distribuciji parametra. Na temelju podataka, a korištenjem Bayesove formule, ažuriramo vjerovanja o parametru. Na taj način dobivamo aposteriornu distribuciju parametra. Još jedna prednost bayesovske statistike je u intuitivnom pristupu testiranju hipoteza o parametru.

Dugi niz godina bayesovska statistika je bila zapostavljena zbog nemogućnosti računanja izraza koji se pojavljuju u Bayesovoj formuli, ali razvojem računalne znanosti ova je mana uklonjena. Naime, pri ažuriranju apriorne distribucije često se pojavljuju integrali koje je nemoguće egzaktno izraziti i teško izračunati. Razvojem Monte Carlo Markovljevih lanaca (Monte Carlo Markov Chain, često se označava kraticom MCMC) ovaj problem se rješava jer postaje moguće simuliranje uzoraka iz aposteriorne distribucije bez eksplicitnog izračunavanja problematičnih izraza. Jedna od MCMC metoda je Metropolis-Hastings algoritam koji će se koristiti u ovom radu.

1 Osnovni pojmovi

Bayesovski prisup bazira se na Bayesovom teoremu. Stoga, kako bismo mogli proučavati bayesovski pristup statističkom zaključivanju potrebno je poznavati osnove teorije vjerojatnosti. Zato ćemo kratko definirati i iskazati ključne pojmove i tvrdnje koje se koriste za bayesovski pristup.

Neka je $\Omega \neq \emptyset$ neki neprazan skup. S $\mathcal{P}(\Omega)$ označavamo partitivni skup od Ω.

Definicija 1.1. *Familija $\mathcal{F}$ podskupova skupa Ω ($\mathcal{F} \subseteq \mathcal{P}(\Omega)$) naziva se σ-algebra na skupu Ω ako vrijedi:*

1. $\emptyset \in \mathcal{F}$,
2. *ako je $A \in \mathcal{F}$, onda je i $A^C \in \mathcal{F}$,*
3. *ako je $(A_n, n \in \mathbb{N}) \in \mathcal{F}$ prebrojiva familija skupova iz $\mathcal{F}$, onda vrijedi i*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Definicija 1.2. *Neka je $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ σ-algebra na skupu Ω. Uređeni par $(\Omega, \mathcal{F})$ naziva se izmjeriv prostor.*

Definicija 1.3. *Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor. Funkcija $P : \mathcal{F} \to \mathbb{R}$ je vjerojatnost na Ω ako vrijedi*

1. $P(A) \geq 0$, *za svaki $A \in \mathcal{F}$,*
2. $P(\Omega) = 1$,
3. *Ako su $A_i \in \mathcal{F}$, $i \in \mathbb{N}$ i $A_i \cap A_j = \emptyset$ za $i \neq j$ (A_i međusobno disjunktni). Tada vrijedi*

$$P(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i).$$

Definicija 1.4. *Uređena trojka $(\Omega, \mathcal{F}, P)$, gdje su $\mathcal{F}$ σ-algebra na Ω i P vjerojatnost na $\mathcal{F}$, naziva se vjerojatnosni prostor.*

Primijetimo nekoliko osnovnih karakteristika funkcije vjerojatnosti P koje lako dobivamo iz definicije vjerojatnosnog prostora. Elemente σ-algebre $\mathcal{F}$ nazivamo događajima, a broj $P(A)$, gdje je $A \in \mathcal{F}$, nazivamo vjerojatnost događaja A.

Teorem 1.1. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor. Vrijedi sljedeće:*

1. $P(\emptyset) = 0$,
2. *Ako su $A_1, ..., A_n$ međusobno disjunktni skupovi, odnosno $A_i \cap A_j = \emptyset$, za $i \neq j$ tada je*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i),$$

3. $A, B \in \mathcal{F}$ *i* $A \subseteq B \Rightarrow P(A) \leq P(B)$,

4. $A_n \in \mathcal{F},\ n \in \mathbb{N} \Rightarrow P(\bigcup\limits_{n=1}^{\infty} A_n) \leq \sum\limits_{n=1}^{\infty} P(A_n)$,

5. $A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

6. $A \in \mathcal{F} \Rightarrow P(A^C) = 1 - P(A)$.

Dokaz se može naći u [1].

Definicija 1.5. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i neka su* A, B *događaji iz* $\mathcal{F}$ $(A, B \in \mathcal{F})$. A *i* B *su nezavisni ako vrijedi* $P(A \cap B) = P(A)P(B)$.

Definicija 1.6. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i neka je* $A \in \mathcal{F}$ *takav da vrijedi* $P(A) > 0$. *Funkcija* $P_A : F \to [0, 1]$ *definirana s*

$$P_A(B) = P(B|A) = \frac{P(A \cap B)}{P(A)},\ B \in \mathcal{F}$$

je vjerojatnost na $\mathcal{F}$ *i naziva se uvjetna vjerojatnost uz uvjet* A. *Broj* $P(B|A)$ *naziva se vjerojatnost od* B *uz uvjet da se dogodio događaj* A.

Dodatno, definiciju uvjetne vjerojatnosti uz uvjet A $(A \in \mathcal{F})$ možemo proširiti na slučaj kada je $P(A) = 0$. U tom slučaju uzimamo da je $P(B|A) = P(B)$ ako je $P(A) = 0$, $B \in \mathcal{F}$. Primijetimo, $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$, $A, B \in \mathcal{F}$. Obzirom da je P_A vjerojatnost na σ-algebri $\mathcal{F}$, iz teorema 1.1 slijedi:

Propozicija 1.1. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor i* $A \in \mathcal{F}$ *takav da je* $P(A) > 0$. *Vrijedi sljedeće:*

1. $P(\emptyset|A) = 0$,

2. $P(B^C|A) = 1 - P(B|A)$, $B \in \mathcal{F}$,

3. $A_1, A_2 \in \mathcal{F} \Rightarrow P(A_1 \cup A_2|A) = P(A_1|A) + P(A_2|A) - P(A_1 \cap A_2|A)$,

4. $A_1, A_2 \in \mathcal{F}$ *i* $A_1 \subseteq A_2 \Rightarrow P(A_1|A) \leq P(A_2|A)$,

5. $A_n \in \mathcal{F},\ n \in \mathbb{N} \Rightarrow P(\bigcup\limits_{n=1}^{\infty} A_n|A) \leq \sum\limits_{n=1}^{\infty} P(A_n|A)$.

Dokaz se može naći u [1]. Za uvođenje pojma slučajne varijable potrebno je definirati nekoliko pojmova.

Definicija 1.7. *Neka je* $\mathbb{R}$ *skup svih realnih brojeva. Neka je* $\mathcal{B}$ σ*-algebra generirana familijom svih otvorenih skupova na* $\mathbb{R}$. σ*-algebru* $\mathcal{B}$ *nazivamo Borelovom* σ*-algebrom na* $\mathbb{R}$, *a elemente Borelove* σ*-algebre* $\mathcal{B}$ *nazivamo Borelovim skupovima.*

Borelova σ-algebra na $\mathbb{R}$ se često označava i s $\mathcal{B}(\mathbb{R})$. Analogno, ako imamo k-dimenzionalni realni prostor, Borelovu σ-algebru koja je generirana svim otvorenim skupovima na R^k označavamo s $\mathcal{B}(R^k)$.

Definicija 1.8. *Funkciju* $g : \mathbb{R} \to \mathbb{R}$ *nazivamo Borelovom funkcijom ako vrijedi da je* $g^{-1}(B) \in \mathcal{B}$, $\forall B \in \mathcal{B}$, *odnosno ako je* $g^{-1}(B) \subset \mathcal{B}$.

Sada možemo definirati slučajnu varijablu.

Definicija 1.9. *Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor. Slučajna varijabla (za $k = 1$), odnosno slučajni vektor (za $k \geq 2$), u paru σ-algebri $(\mathcal{F}, \mathcal{B}(\mathbb{R}^k))$ je izmjeriva funkcija $X : \Omega \to \mathbb{R}^k$, odnosno funkcija za koju vrijedi $X^{-1}(B) \in \mathcal{F}$ za svako $B \in \mathcal{B}(\mathbb{R}^k)$, tj. $X^{-1}(\mathcal{B}(\mathbb{R}^k)) \subseteq F$.*

Budući da Ω može biti vrlo kompliciran i da je ponekad teško zadati vjerojatnost P na $(\Omega, \mathcal{F})$, želimo skup $\mathbb{R}$ opskrbiti dovoljno bogatom σ-algebrom i na njoj definirati novu vjerojatnost P_X. Tada $P(X \in A)$ računamo kao $P_X(A)$. Kako smo na $\mathbb{R}$ definirali Borelovu σ-algebru, ostaje nam još definirati odgovarajuću vjerojatnost P_X. To ćemo napraviti na sljedeći način:

$$P_X(B) := P(X \in B) = P(X^{-1}(B)),\ B \in \mathcal{B}.$$

Vjerojatnost P_X još nazivamo vjerojatnosna mjera inducirana slučajnom varijablom X. Za vjerojatnosni prostor $(\mathbb{R}, \mathcal{B}, P_X)$ kažemo da je vjerojatnosni prostor induciran slučajnom varijablom X.

Definicija 1.10. *Neka je X slučajna varijabla na skupu Ω. Funkcija distribucije od X je funkcija $F_X : \mathbb{R} \to [0, 1]$ definirana kao $F_X(x) = P_X(\big\langle -\infty, x\big]) = P(X \in \big\langle -\infty, x\big]) = P(X \leq x)$, $x \in \mathbb{R}$.*

Pristup računanju vjerojatnosti događaja ovisi o tipu slučajne varijabli. Osnovni tipovi slučajnih varijabli su diskretne i neprekidne slučajne varijable.

Definicija 1.11. *Za slučajnu varijablu X kažemo da je diskretna ako postoji konačan (ili prebrojiv) skup $D \subseteq \mathbb{R}$ takav da je $P_X(D) = P(X \in D) = 1$.*

Diskretne slučajne varijable najčešće se zadaju tablicom distribucije

$$X = \begin{pmatrix} x_1 & x_2 & ... & x_n & ... \\ p_1 & p_2 & ... & p_n & ... \end{pmatrix},$$

pri čemu su $x_1, ..., x_n$ moguće realizacije slučajne varijable X, a $p_i = P(X = x_i)$, za $i \in \{1, ..., n, ...\}$.

Definicija 1.12. *Neka je X slučajna varijabla na $(\Omega, \mathcal{F}, P)$ s funkcijom distribucije F_X. Kažemo da je X neprekidna slučajna varijabla ako postoji nenegativna realna Borelova funkcija $f : \mathbb{R} \to \mathbb{R}_+$ takva da vrijedi*

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(t)dt,\ x \in \mathbb{R}.$$

Funkciju f nazivamo funkcijom gustoće slučajne varijable X.

Ostali temeljni pojmovi iz područja vjerojatnosti i statistike potrebni za praćenje i razumijevanje ovog rada mogu se naći u [1].

2 Funkcija vjerodostojnosti i Bayesov teorem

Neka su dani podaci $x_1, x_2, ..., x_n$ koji su realizacije jednako distribuiranih slučajnih varijabli $X_1, X_2, ..., X_n$, ali pri čemu ne znamo kojoj distribuciji pripadaju. S $\mathcal{P}_0$ označimo familiju svih distribucija kojima $X_1, X_2, ...X_n$ mogu pripadati. Primjerice, $\mathcal{P}_0$ može biti familija svih normalnih distribucija, svih Poissonovih distribucija itd. Za $\mathcal{P}_0$ mora vrijediti da može biti indeksiran konačno dimenzionalnom realnom veličinom, odnosno parametrom. Npr. ako $\mathcal{P}_0$ sadrži sve Poissonove distribucije, tada parametar možemo označiti s λ, gdje λ predstavlja očekivani broj realizacija promatranog događaja u fiksnom vremenskom intervalu. Skup svih vrijednosti koje parametar može poprimiti nazivamo parametarski prostor, kojeg ćemo označavati s Φ. Vjerojatnost i očekivanje slučajnih varijabli s distribucijom iz $\mathcal{P}_0$, ako znamo $\Theta = \theta$, označavamo s P_θ i E_θ.

2.1 Funkcija vjerodostojnosti

Definicija 2.1. *Neka je $X = (X_1, ..., X_n)$ slučajni vektor i $x = (x_1, ..., x_n)$ jedna realizacija slučajnog vektora X. Neka je $f(x|\theta)$ njegova funkcija gustoće. Funkciju po parametru θ: $L(\theta|x) := f(x|\theta)$, $\theta \in \Phi$ nazivamo funkcijom vjerodostojnosti (likelihood) od X na osnovi realizacije x.*

Navest ćemo dva primjera funkcije vjerodostojnosti: u prvom će podaci biti iz Poissonove distribucije, a u drugom iz normalne.

Primjer 2.1. *Slučajnom varijablom iz Poissonove distribucije se modelira vjerojatnost pojave promatranog događaja u fiksnom vremenskom intervalu. Poissonova distribucija je određena parametrom $\lambda \geq 0$, koji predstavlja očekivani broj realizacija događaja tijekom fiksnog vremenskog intervala. Vjerojatnost da se u tom vremenskom intervalu slučajna varijabla X iz Poissonove distribucije realizira s k (odnosno da se promatrani događaj realizira k puta) dana je s*

$$P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, 3, ...$$

Pretpostavimo da je dana realizacija jednostavnog slučajnog vektora $(x_1, x_2, ..., x_n)$ iz Poissonove distribucije. Tada za funkciju vjerodostojnosti za parametar λ vrijedi:

$$L(\lambda|x_1, ..., x_n) = \prod_{i=1}^{n} P_\lambda(X = x_i) = \prod_{i=1}^{n} \frac{\lambda_i^{x}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!} e^{-n\lambda},$$

odnosno $L(\lambda|x_1, ..., x_n)$ je proporcionalno $\lambda^{\sum\limits_{i=1}^{n} x_i} e^{-n\lambda}$.

Primjer 2.2. *Normalna (Gaussova) distribucija jedna je od najvažnijih i najkorištenijih distribucija u vjerojatnosti i statistici. Normalna slučajna varijabla određena je s dva parametra: μ i σ^2 koji su očekivanje i varijanca slučajne varijable s normalnom distribucijom. Za funkciju gustoće slučajne varijable X s normalnom distribucijom vrijedi:*

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

Koristimo oznaku $X \sim \mathcal{N}(\mu, \sigma^2)$. Pretpostavimo da je dana realizacija jednostavnog slučajnog vektora $(x_1, x_2, ..., x_n)$ iz normalne distribucije. Tada funkcija vjerodostojnosti za parametre μ i σ^2 glasi:

$$L(\mu, \sigma^2 | x_1, ..., x_n) = \sum_{i=1}^{n} f(x_i|\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum\limits_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}},$$

odnosno $L(\mu, \sigma^2 | x_1, ..., x_n)$ *je proporcionalno s* $\sigma^{-n} e^{-\sum\limits_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$.

2.2 Bayesov teorem

Bayesov teorem jedan je od osnovnih rezultata u teoriji vjerojatnosti i temelj je bayesovske statistike. Kao što je već naglašeno, u bayesovskoj statistici parametri se promatraju kao slučajne varijable. Taj postupak se provodi na sljedeći način: odaberemo distribuciju koja najbolje opisuje naša vjerovanja o parametru (ova distribucija se naziva apriorna distribucija i jedan od najvećih problema u bayesovskom pristupu je adekvatan odabir apriorne distribucije; više o apriornoj distribuciji i nekim načinima njenog odabira će biti riječi u nastavku poglavlja), a zatim iskoristimo podatke kako bismo ažurirali ta vjerovanja. Ovo ažuriranje vrši se korištenjem Bayesovog teorema.

Teorem 2.1. *(Bayesov teorem) Neka su A,B događaji na Ω i neka vrijedi $P(B) > 0$. Tada vrijedi sljedeća jednakost koja se naziva Bayesova formula:*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Iz Bayesove formule očigledno slijedi:

$$P(A|B) \propto P(B|A)P(A).$$

Objasnimo način na koji se vrši ažuriranje vjerovanja o parametru Θ. Pretpostavimo da je Θ diskretna slučajna varijabla. Nadalje, neka E označava skup nekih informacija o Θ. Ono što nas zanima je distribucija parametra Θ u svjetlu novih informacija E, odnosno zanima nas vjerojatnost $P(\Theta = \theta|E)$. Koristimo Bayesov teorem:

$$P(\Theta = \theta|E) = P(\Theta = \theta)\frac{P(E|\Theta = \theta)}{P(E)}.$$

Primijetimo nekoliko činjenica:

1. $P(\Theta = \theta)$ je inicijalana funkcija gustoće od Θ, funkcija gustoće koja se odnosi na vjerovanja o parametru Θ prije uzimanja u obzir novih infomacija E. Ovu funkciju gustoće nazivamo apriornom funkcijom gustoće.

2. Nakon uzimanja u obzir informacija E, ažuriramo naša vjerovanja o Θ koja su sada sadržana u funkciji gustoće $P(\Theta = \theta|E)$. Ovu gustoću nazivamo aposteriorna gustoća.

3. $\dfrac{P(E|\Theta = \theta)}{P(E)}$ je izraz koji pokazuje koliko će novi skup informacija E izmijeniti naša vjerovanja o Θ, odnosno koliko će promijeniti apriornu distribuciju. Primijetimo i da je $P(E|\Theta = \theta)$ vjerodostojnost za fiksnu vrijednost θ.

3 Apriorna i aposteriorna distribucija

Neka je Φ skup svih vrijednosti koje parametar Θ može poprimiti. Nadalje, s π označimo apriornu distribuciju parameta Θ. Primijetimo da je $\pi : \Phi \longrightarrow \mathbb{R}$. Kao i dosad, jednu realizaciju parametra Θ označavat ćemo s θ. Primijetimo da je $\theta \in \Phi$. Neka je X slučajna varijabla ili vektor iz kojeg dolaze podaci, a x njegova realizacija. Uvjetnu funkciju gustoće od X, uz danu vrijednost $\Theta = \theta$, označimo s $f(x|\theta)$. Uvodimo sljedeće distribucije i oznake:

1. zajedničku distribuciju slučajnog vektora (Θ, X) s funkcijom gustoće:

$$\varphi(\theta, x) = f(x|\theta)\pi(\theta),$$

2. marginalnu distribuciju od X s funkcijom gustoće:

$$m(x) = \int \varphi(\theta, x) d\theta = \int f(x|\theta)\pi(\theta) d\theta,$$

3. aposteriornu distribuciju parametra Θ, odnosno uvjetnu distribuciju od Θ uvjetovano na realizaciju $X = x$ s funkcijom gustoće:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\xi)\pi(\xi) d\xi} = \frac{f(x|\theta)\pi(\theta)}{m(x)}.$$

Iz načina na koji je definirana aposteriorna distribucija parametra, zaključujemo da, ukoliko znamo funkciju gustoće $f(x|\theta)$ i apriornu funkciju gustoće parametra $\pi(\theta)$, možemo izračunati aposteriornu distribuciju parametra Θ (odnosno imamo formulu za to, sam izračun je često problematičan), odnosno ažurirati naša vjerovanja o parametru Θ uzimajući u obzir nove informacije $X = x$. Sada ćemo definirati Bayesovski statistički model.

Definicija 3.1. *Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor i neka je $\mathcal{P}$ neka familija vjerojatnosti na $(\Omega, \mathcal{F})$. Uređena trojka $(\Omega, \mathcal{F}, \mathcal{P})$ naziva se statistička struktura. Familija $\mathcal{P}$ je često parametrizirana i zapisuje se u obliku:*

$$\mathcal{P} = \{P_\theta : \theta \in \Phi\}.$$

Definicija 3.2. *Statistika na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je svaka slučajna varijabla (ili slučajni vektor za $k \geq 2$) $T : \Omega \to \mathbb{R}^k$ takva da za neki $n \in \mathbb{N}$ postoji n-dimenzionalni slučajni vektor $(X_1, X_2, ..., X_n)$ na $(\Omega, \mathcal{F}, \mathcal{P})$, te izmjeriva funkcija $t : R^n \to R^k$ takva da je $T = t(X_1, X_2, ..., X_n)$.*

Definicija 3.3. *Bayesovski statistički model slučajne varijable (ili slučajnog vektora) $X : \Omega \to R^k$ je familija $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Phi\}$ iz prethodne definicije, pri čemu je parametru Θ pridružena njegova apriorna distribucija s gustoćom $\pi(\theta)$, $\theta \in \Phi$.*

3.1 Izbor apriorne distribucije

Sada ćemo promotriti neke različite vrste apriornih distribucija, kao i metode odabira apriorne distribucije. Vidjeli smo na koji se način pomoću Bayesovog teorema može od apriorne distribucije doći do aposteriorne distribucije. Napomenimo ovdje i mogućnost jednog neobičnog pristupa, moguće je odabrati apriornu distribuciju koja nije određena vjerojatnosnom mjerom, nego mjerom (σ-konačnom na skupu Φ) takvom da vrijedi:

$$\int_{\Phi} \pi(\theta) d\theta = +\infty.$$

Definicija 3.4. *Ako za apriornu funkciju gustoće π vrijedi*

$$\int_{\Phi} \pi(\theta) d\theta = +\infty,$$

apriorna distribucija naziva se generalizirana ili nepravilna.

Dok god je aposteriorna distribucija konačna, moguće je koristiti i nepravilnu apriornu distribuciju, odnosno nepravilna apriorna distribucija ne dovodi nužno do nepravilne aposteriorne distribucije ([3]). U ostalim slučajevima, odnosno u slučaju nepravilne aposteriorne distribucije, statističko zaključivanje neće biti valjano i bayesovske metode se ne mogu koristiti.

Općenito, ključni korak u bayesovskoj statistici je dobar odabir apriorne distribucije. Dobar odabir apriorne distribucije nije trivijalan, a za cilj ima da apriorna distribucija sadrži što više informacija koje znamo o promatranom parametru Θ prije nego što su nam dani podaci. Zaključci u daljnjoj analizi mogu značajno varirati za različite apriorne distribucije. Hiperparametrima ćemo nazivati parametre apriorne distribucije.

Ovdje ćemo definirati i promotriti dvije metode izbora apriorne distribucije:

1. metodu konjugiranih apriornih distribucija,

2. metodu neinformativnih apriornih distribucija.

3.1.1 Konjugirane apriorne distribucije

Definicija 3.5. *Za familiju $\mathcal{F}$ vjerojatnosnih distribucija na Φ kažemo da je konjugirana za statističku strukturu $\mathcal{P}$ ako za svaki $\pi \in \mathcal{F}$ također vrijedi $\pi(\theta|x) \in \mathcal{F}$, za svaki x. Tada apriornu i aposteriornu distribuciju nazivamo konjugiranim distribucijama, a apriornu funkciju gustoće nazivamo konjugiranim apriorom za funkciju vjerodostojnosti danu pomoću modela $\mathcal{P}$.*

Primijetimo, familija $\mathcal{F}_0$ svih distribucija na Φ je trivijalna konjugirana familija na Φ. Naravno, ovaj izbor konjugirane familije nije koristan za izbor apriorne distribucije. Najčešće se pokušava pronaći što manja parametrizirana familija $\mathcal{F}$. Pretpostavimo da imamo neki slučajni uzorak X s funkcijom gustoće $f(x|\theta)$, pri čemu je θ neka realizacija parametra Θ. Kao što je rečeno, tada se apriorna gustoća $\pi(\theta)$ ažurira čime dobivamo novu distribuciju parametra opisanu funkcijom gustoće $\pi(\theta|x)$. Želimo da ovo ažuriranje ne promijeni cijelu strukturu $\pi(\theta)$, odnosno ne promijeni klasu distribucije, nego samo njezine parametre. To je razlog zbog kojeg tražimo parametrizirane konjugirane familije. Pokažimo dva primjera konjugirane apriorne distribucije.

Primjer 3.1. *Pretpostavimo da imamo jednu realizaciju x_1 slučajne varijable X_1 s normalnom distribucijom $\mathcal{N}(\mu, \sigma^2)$, funkcijom gustoće f i neka je σ^2 poznat. Neka je apriorna distribucija parametra $\mu = \Theta$ normalna distribucija s parametrima μ_0 i σ_0^2. Funkciju gustoće apriorne distribucije ćemo, kao i dosad, označiti s π. Izračunajmo aposteriornu distribuciju parametra Θ, uz uvjet da se realizira $X_1 = x_1$, i pokažimo da se radi o normalnoj distribuciji:*

$$\pi(\theta|x_1) = \frac{f(x_1|\theta)\pi(\theta)}{f(x_1)} = \frac{\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x_1-\theta)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_0}e^{\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}}}{\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x_1-\epsilon)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_0}e^{\frac{-(\epsilon-\mu_0)^2}{2\sigma_0^2}}\, d\epsilon} = \frac{e^{\frac{-(x_1-\theta)^2}{2\sigma^2}} e^{\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}}}{\int\limits_{-\infty}^{\infty} e^{\frac{-(x_1-\epsilon)^2}{2\sigma^2}} e^{\frac{-(\epsilon-\mu_0)^2}{2\sigma_0^2}}\, d\epsilon}.$$

Funkciju pod integralom u nazivniku možemo zapisati kao e^{-j} gdje je

$$\begin{aligned}
j &= \frac{\sigma_0^2 x_1^2 - 2x_1\epsilon\sigma_0^2 + \sigma_0^2\epsilon^2 + \sigma^2\epsilon^2 - 2\sigma^2\mu_0\epsilon + \sigma^2\mu_0^2}{2\sigma^2\sigma_0^2} \\
&= \frac{(\sigma^2+\sigma_0^2)\epsilon^2 - 2\epsilon(\sigma_0^2 x_1 + \sigma^2\mu_0)}{2\sigma^2\sigma_0^2} + \frac{\sigma_0^2 x_1^2 + \sigma^2\mu_0^2}{2\sigma^2\sigma_0^2} \\
&= \left(\frac{\epsilon - \frac{\sigma_0^2 x_1 + \sigma^2\mu_0}{\sigma^2+\sigma_0^2}}{\sqrt{2}\frac{\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}}}\right)^2 - \frac{\frac{(\sigma_0^2 x_1+\sigma^2\mu_0)^2}{\sigma^2+\sigma_0^2}}{2\frac{\sigma^2\sigma_0^2}{\sigma^2\sigma_0^2}} + \frac{\sigma_0^2 x_1^2 + \sigma^2\mu_0^2}{2\sigma^2\sigma_0^2} \\
&= \left(\frac{\epsilon - \frac{\sigma_0^2 x_1 + \sigma^2\mu_0}{\sigma^2+\sigma_0^2}}{\sqrt{2}\frac{\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}}}\right)^2 + \frac{-\sigma_0^4 x_1 - 2\sigma_0^2\sigma^2 x_1\mu_0 - \sigma^4\mu_0^2 + \sigma_0^2\sigma^2 x_1^2 + \sigma^4\mu_0^2 + \sigma_0^4 x_1^2 + \sigma^2\sigma_0^2\mu_0^2}{2\sigma^2\sigma_0^2(\sigma^2+\sigma_0^2)} \\
&= \left(\frac{\epsilon - \frac{\sigma_0^2 x_1 + \sigma^2\mu_0}{\sigma^2+\sigma_0^2}}{\sqrt{2}\frac{\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}}}\right)^2 + \frac{(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}.
\end{aligned}$$

Uvodimo supstituciju $u = \frac{\epsilon - \frac{\sigma_0^2 x_1 + \sigma^2\mu_0}{\sigma^2+\sigma_0^2}}{\sqrt{2}\frac{\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}}}$, $d\epsilon = \frac{\sqrt{2}\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}}du$ *i dobivamo*

$$\int\limits_{-\infty}^{\infty} e^{\frac{-(x_1-\epsilon)^2}{2\sigma^2}} e^{\frac{-(\epsilon-\mu_0)^2}{2\sigma_0^2}}\, d\epsilon = \frac{\sqrt{2}\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}} \cdot e^{-\frac{(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}} \int\limits_{\infty}^{\infty} e^{-u^2} du.$$

Pomnožimo i podijelimo s $\sqrt{\pi}$:

$$\int\limits_{-\infty}^{\infty} e^{\frac{-(x_1-\epsilon)^2}{2\sigma^2}} e^{\frac{-(\epsilon-\mu_0)^2}{2\sigma_0^2}}\, d\epsilon = \frac{\sqrt{2\pi}\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}} \cdot e^{-\frac{(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}} \int\limits_{\infty}^{\infty} \frac{e^{-u^2}}{\sqrt{\pi}} du.$$

Vrijedi da je $\int\limits_{\infty}^{\infty} \frac{e^{-u^2}}{\sqrt{\pi}} du = 1.$ [1] *Ovime smo rješili integral iz nazivnika. Dakle, vrijedi:*

$$\pi(\theta|x_1) = \frac{e^{\frac{-(x_1-\theta)^2}{2\sigma^2}} e^{\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}}}{\frac{\sqrt{2\pi}\sigma\sigma_0}{\sqrt{\sigma^2+\sigma_0^2}} e^{\frac{-(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}}} = \frac{\sqrt{\sigma^2+\sigma_0^2}}{\sqrt{2\pi}\sigma\sigma_0} \cdot e^{-k},$$

[1] $\lim\limits_{u\to\infty} erf(u) = 1$, gdje je erf funkcija greške (error function); vidjeti K.B. Oldham, J.C. Myland , J. Spanier, poglavlje The Error Function $erf(x)$ and Its Complement $erfc(x)$ iz knjige An Atlas of Functions. Springer, New York, NY,2008.

gdje je

$$k = \frac{(x_1-\theta)^2}{2\sigma^2} + \frac{(\theta-\mu_0)^2}{2\sigma_0^2} - \frac{(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)} = \frac{\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}(x_1-\theta^2) + \frac{\sigma^2}{\sigma^2+\sigma_0^2}(\theta-\mu_0)^2 - \frac{\sigma^2\sigma_0^2}{(\sigma^2+\sigma_0^2)^2}(x_1-\mu_0)^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}$$

$$= \frac{\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}x_1^2 - 2\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}x_1\theta + \frac{\sigma_0^2}{\sigma^2+\sigma_0^2}\theta^2 + \frac{\sigma^2}{\sigma^2+\sigma_0^2}\theta^2\frac{\sigma^2}{\sigma^2+\sigma_0^2}\theta\mu_0 + \frac{\sigma^2}{\sigma^2+\sigma_0^2}\mu_0^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}$$

$$+ \frac{\frac{\sigma^2\sigma_0^2}{(\sigma^2+\sigma_0^2)^2}x_1^2 - 2\frac{\sigma^2\sigma_0^2}{(\sigma^2+\sigma_0^2)^2}x_1\mu_0 + \frac{\sigma^2\sigma_0^2}{(\sigma^2+\sigma_0^2)^2}\mu_0^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} = \ldots$$

$$= \frac{\theta^2 - 2\theta\left(\frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}\right) + \left(\frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}\right)^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}$$

$$= \frac{\left(\theta - \left(\frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}\right)\right)^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}.$$

Sada imamo:

$$\pi(\theta|x_1) = \frac{e^{\frac{-(x_1-\theta)^2}{2\sigma^2}}e^{\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}}}{e^{\frac{-(x_1-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}}} = \frac{\sqrt{\sigma^2+\sigma_0^2}}{\sqrt{2\pi}\sigma\sigma_0}\cdot e^{-k} = \frac{\sqrt{\sigma^2+\sigma_0^2}}{\sqrt{2\pi}\sigma\sigma_0}e^{-\frac{\left(\theta-\left(\frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2}+\frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}\right)\right)^2}{2\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}},$$

što je funkcija gustoće normalne slučajne varijable s očekivanjem $\mu_1 := \frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}$ *i varijancom* $\sigma_1^2 := \frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}$.

Dodajmo i drugi podatak x_2, *istim postupkom za aposteriornu distribuciju parametra* $\pi(\theta|x_1,x_2)$ *dobivamo normalnu distribuciju s očekivanjem*

$$\mu_2 = \frac{\sigma_1^2x_2}{\sigma^2+\sigma_1^2} + \frac{\sigma^2\mu_1}{\sigma^2+\sigma_1^2} = \frac{\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}x_2}{\sigma^2+\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} + \frac{\sigma^2\left(\frac{\sigma_0^2x_1}{\sigma^2+\sigma_0^2}+\frac{\sigma^2\mu_0}{\sigma^2+\sigma_0^2}\right)}{\sigma^2+\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} = \frac{\frac{\sigma^2\sigma_0^2x_1+\sigma^2\sigma_0^2x_2}{\sigma^2+\sigma_0^2}}{\frac{\sigma^2\sigma_0^2+\sigma^4+\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} + \frac{\frac{\sigma^4\mu_0}{\sigma^2+\sigma_0^2}}{\frac{\sigma^2\sigma_0^2+\sigma^4+\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}$$

$$= \frac{\sigma^2\sigma_0^2(x_1+x_2)}{\sigma^2(\sigma^2+2\sigma_0^2)} + \frac{\sigma^4\mu_0}{\sigma^2(\sigma^2+2\sigma_0^2)} = \frac{\sigma_0^2(x_1+x_2)}{\sigma^2+2\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+2\sigma_0^2}$$

i varijancom

$$\sigma_2^2 = \frac{\sigma^2\sigma_1^2}{\sigma^2+\sigma_1^2} = \frac{\frac{\sigma^2\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}{\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}+\sigma^2} = \frac{\frac{\sigma^4\sigma_0^2}{\sigma^2+\sigma_0^2}}{\frac{\sigma^2\sigma_0^2+\sigma^4+\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} = \frac{\sigma^4\sigma_0^2}{\sigma^2(\sigma^2+2\sigma_0^2)} = \frac{\sigma^2\sigma_0^2}{\sigma^2+2\sigma_0^2}.$$

Ukoliko imamo realizacije $x_1,\ldots,x_n$ *jednostavnog slučajnog uzorka iz normalne distribucije s poznatom varijancom aposteriornu distribuciju parametra dobijemo induktivno kao normalnu distribuciju s očekivanjem* $\frac{\sigma_0^2\sum\limits_{i=1}^{n}x_i}{\sigma^2+n\sigma_0^2} + \frac{\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}$ *i varijancom* $\frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2}$ *izraženim u terminima parametara apriorne distribucije i poznate varijance populacije.*

Napomena 3.1. *Gama funkcija* $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ *definira se kao*

$$\gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

Za gama funkciju vrijedi:

1. $\gamma(z) = (z-1)\gamma(z-1)$,
2. $\gamma(z+1) = a^{z+1}\int_0^\infty t^z e^{-at}dt,\ a > 0.$

Dokaz.

1.

$$\begin{aligned}\gamma(z) &= \int_0^\infty t^{z-1}e^{-t}dt = t^{z-1}(-e^{-x})\Big|_0^\infty - \int_0^\infty (z-1)t^{z-2}(-e^{-t})dt \\ &= 0 + (z-1)\int_0^\infty t^{z-2}e^{-t}dt = (z-1)\gamma(z-1).\end{aligned}$$

2.

$$\gamma(z+1) = \int_0^\infty x^z e^{-x}dx = \left| x = at, dx = adt \right| = \int_0^\infty (at)^z e^{-at}adt = a^{z+1}\int_0^\infty t^z e^{-at}dt.$$

□

Gama funkciju označavat ćemo s γ, *a gama distribuciju s* Γ.

Primjer 3.2. *Pretpostavimo da imamo jednu realizaciju* x *slučajne varijable* X *s Poissonovom distribucijom. Neka je apriorna distribucija parametra* Θ *gama distribucija s parametrima* $\alpha > 0$ *i* $\beta > 0$. *Funkciju gustoće apriorne distribucije ćemo, kao i dosad, označiti s* π. *Izračunajmo aposteriornu distribuciju parametra* Θ, *uz uvjet da se realizira* $X = x$, *i pokažimo da se radi o gama distribuciji:*

$$\pi(\theta|x) = \frac{P(X = x|\theta)\pi(\theta)}{P(X = x)} = \frac{\frac{\theta^x e^{-\theta}}{x!}\cdot\frac{\beta^\alpha\theta^{\alpha-1}e^{-\beta\theta}}{\gamma(\alpha)}}{\int_0^\infty \frac{\epsilon^x e^{-\epsilon}\beta^\alpha\epsilon^{\alpha-1}e^{-\beta\epsilon}}{x!\gamma(\alpha)}d\epsilon} = \frac{\theta^x e^{-\theta}\theta^{\alpha-1}e^{-\beta\theta}}{\int_0^\infty \epsilon^x e^{-\epsilon}\epsilon^{\alpha-1}e^{-\beta\epsilon}d\epsilon}.$$

Riješimo integral iz nazivnika (koristimo napomenu 3.1):

$$\int_0^\infty \epsilon^x e^{-\epsilon}\epsilon^{\alpha-1}e^{-\beta\epsilon}d\epsilon = \int_0^\infty \epsilon^{\alpha+x-1}e^{-(\beta+1)\epsilon} = \frac{\gamma(\alpha+x)}{(\beta+1)^{\alpha+x}}.$$

Nakon što smo riješili integral iz nazivnika dobivamo:

$$\pi(\theta|x) = \frac{\theta^x e^{-\theta}\theta^{\alpha-1}e^{-\beta\theta}}{(\beta+1)^{-(\alpha+x)}\gamma(\alpha+x)}$$

$$= \frac{\theta^x e^{-\theta}\theta^{\alpha-1}e^{-\beta\theta}(\beta+1)^{(\alpha+x)}}{\gamma(\alpha+x)}$$
$$= \frac{\theta^{\alpha+x-1}(\beta+1)^{\alpha+x}e^{-\theta(\beta+1)}}{\gamma(\alpha+x)},$$

a ovo je funkcija gustoće gama distribucije s parametrima $\alpha + x$ i $\beta + 1$.
Ukoliko imamo realizacije $x_1, ..., x_n$ jednostavnog slučajnog uzorka iz Poissonove distribucije aposteriornu distribuciju parametra dobijemo induktivno kao gama distribuciju s parametrima $\alpha + \sum_{i=1}^{n} x_i$ i $\beta + n$.

Dodatno, postoji rezultat koji govori kako se za distribucije iz eksponencijalne klase mogu pronaći konjugirane apriorne distribucije.

Napomena 3.2. *Neka je $f(x|\theta) = h(x)e^{\langle\theta|x\rangle-\phi(\theta)}$ funkcija gustoće iz eksponencijalne klase distribucija* [2]. *Apriorna funkcija gustoće njene konjugirane familije tada je dana s:*

$$\pi(\theta|\mu,\lambda) = K(\mu,\lambda)e^{\langle\theta|\mu\rangle-\lambda\phi(\theta)},$$

gdje su μ i λ parametri apriorne distribucije (hiperparametri). Pri tome je $\lambda > 0$ i $\frac{\mu}{\lambda} \in Int(N)$, gdje je $Int(N)$ interior prirodnog prostora parametara $N = \{\theta; \int e^{\langle\theta|x\rangle}h(x)dx < +\infty\}$, a $K(\mu,\lambda)$ konstanta. Neka je $x = (x_1, ..., x_n)$ vektor podataka. Tada je aposteriorna funkcija gustoće dana s $\pi(\theta|\mu + \sum_{i=1}^{n} x_i, \lambda + n)$.

Više o ovom rezultatu i njegovim primjenama može se naći u [11].

3.1.2 Neinformativne apriorne distribucije

Dakle, pokazali smo na koji način konjugirane familije distribucija mogu biti korisne za izbor apriorne distribucije. Primijetimo da, ako nemamo dovoljno podataka možda nećemo moći odrediti iz koje distribucije iz eksponencijalne klase dolaze podaci što će onemogućiti pronalazak apriorne distribucije iz eksponencijalne familije. Zato, kao alternativu možemo koristiti neinformativne apriorne distribucije. Česte neinformativne apriorne distribucije su uniformna distribucija ili Jeffreysova apriorna distribucija koja se temelji na informaciji Fishera.

Definicija 3.6. *Neka je $\pi(\theta)$ odabrana apriorna funkcija gustoće neprekidne slučajne varijable (ili vektora) kojom modeliramo parametar Θ. Kažemo da je π neinformativna apriorna funkcija gustoće, ako je invarijantna na reparametrizaciju, tj. ako izbor parametara (odnosno realizacije parametara) ne utječe na izbor apriorne funkcije gustoće. Ako parametar θ reparametriziramo u $\eta = g(\theta)$, gdje je g bijekcija, tada mora vrijediti:*

$$\pi(\eta) = \pi(g^{-1}(\eta))\left|\frac{\partial(g^{-1}(\eta))}{\partial(\eta)}\right|,$$

Za podatke iz binomnog modela, Laplace (francuski matematičar Pierre-Simon Laplace) je među prvima koristio neinformativne apriorne distribucije. Ovdje ćemo pokazati jedan njegov primjer.

---
[2] S $\langle\cdot|\cdot\rangle$ je označen skalarni produkt.

Primjer 3.3. *Pretpostavimo da imamo kutiju koja sadrži n kuglica, od kojih su neke bijele, a neke crne boje. Postotak bijelih kuglica modeliramo slučajnom varijablom A. Pretpostavimo da je u prvom izvlačenju izvučena bijela kuglica (ovaj događaj označimo kao $X = x_B$). Zanima nas s kolikom vjerojatnošću se A realizira s p_0? Kako bi riješio problem, Laplace je pretpostavio da su sve moguće realizacije slučajne varijeble A jednako vjerojatne, odnosno da je A uniformno distribuirana na $\{0, \frac{1}{n}, \frac{2}{n}..., 1\}$. Aposteriorna distribucija od A se tada izvodi pomoću Bayesovog teorema i dobijemo*

$$P(A = p_0|X = x_B) = \frac{P(X = x_B|A = p_0)P(A = p_0)}{P(X = x_B)}$$

Kako je apriorna distribucija uniformna, a skup mogućih realizacija od A je skup od $n + 1$ elemenata, vrijedi $P(A = p_0) = \frac{1}{n+1}$. Vjerojatnost da je iz kutije izvučena bijela kugla uz uvjet da je proporcija bijelih kugli u kutiji jednaka p_0 je upravo p_0. Vjerojatnost $P(X = x_B)$ ćemo raspisati po formuli potpune vjerojatnosti (vidjeti [1]). Dakle, imamo:

$$P(A = p_0|X = x_B) = \frac{p_0 \cdot \frac{1}{n+1}}{\sum_{k=0}^{n} \frac{k}{n} \cdot \frac{1}{n+1}} = \frac{p_0 \cdot \frac{1}{n+1}}{\frac{1}{n(n+1)} \sum_{k=1}^{n} k} = \frac{p_0 \cdot \frac{1}{n+1}}{\frac{n(n+1)}{2n(n+1)}} = \frac{p_0 \cdot \frac{1}{n+1}}{\frac{1}{2}} = \frac{2p_0}{n+1}.$$

Primjerice za $n = 3$ apriorna distribucija slučajne varijable A je

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Nakon što je izvučena bijela kuglica, uz korištenje Bayesove formule dobivamo sljedeću aposteriornu distribuciju:

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ \\ 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}.$$

Britanski matematičar Sir Harold Jeffreys je predložio pristup izbora apriorne distribucije koji se temelji na informaciji Fishera.

Definicija 3.7. *Neka je X slučajni uzorak iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Phi\}$. Fisherovu informaciju s obzirom na uzorak X definiramo kao:*

$$I(\theta) = E_\theta\left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta}\right)^2\right].$$

Definicija 3.8. *Jeffreysova neinformativna apriorna distribucija dana je funkcijom gustoće:*

$$\pi_j(\theta) \propto \sqrt{I(\theta)}.$$

Više o ovoj apriornoj distribuciji se može naći u [7].

4 Bayesovska procjena parametara

4.1 Funkcija gubitka i procjena parametra

Nakon što odaberemo apriornu distribuciju i izračunamo aposteriornu distribuciju, željeli bismo na dobar način procijeniti parametar. Kako bismo pronašli dobrog procjenitelja određujemo kriterij na temelju kojeg ćemo uspoređivati procjenitelje. Taj kriterij ćemo nazivati gubitak. S $\mathcal{D}$ označimo skup svih mogućih odluka i nazovimo ga prostorom odluka.

Definicija 4.1. *Funkcija gubitka je bilo koja funkcija l iz skupa $\Phi \times \mathcal{D}$ u skup $[0, +\infty]$.*

Za funkciju gubitka možemo reći da daje informacije o tome koliko je dobra odluka δ ako se parametar Θ realizira s θ. Odluka δ je ovdje zapravo ona vrijednost koju uzimamo za procjenu (najčešće uzimamo $\mathcal{D} = \Phi$). Funkcija gubitka se uglavnom određuje na subjektivan način, ona ovisi o konkretnom problemu i analitičaru koji ga rješava. Često se kao funkcije gubitka koriste neke jednostavne funkcije kao što su kvadratna vrijednost razlike i apsolutna vrijednost razlike. Bayesovski pristup u statističkom zaključivanju ovisi o inicijalnom izboru tri faktora:

1. distribucije slučajnog uzorka, odnosno funkcije gustoće $f(x|\theta)$,
2. apriorne distribucije parametra $\pi(\theta)$ i
3. funkcije gubitka $l(\theta, \delta)$.

Ponekad se pristup donošenju odluke, odnosno procjeni parametra, gleda iz konteksta maksimizacije nagrade umjesto iz konteksta minimizacije gubitka. Funkciju koja "mjeri" nagradu nazivamo funkcija korisnosti i označavamo s U. Za isti problem, ako bismo izabrali funkciju l kao funkciju gubitka, analognim načinom razmišljanja za funkciju korisnosti bismo izabrali

$$U(\theta, \delta) = -l(\theta, \delta).$$

Općenito, za nepoznat θ, nije moguće uvijek uniformno minimizirati funkciju gubitka $l(\theta, \delta)$. Stoga, kao efikasan kriterij za uspoređivanje različitih odluka uvodi se funkcija rizika

$$R(\theta, \delta) = E_\theta\big[l(\theta, \delta(X))\big] = \int_{\mathbb{R}} l(\theta, \delta(x)) f(x|\theta) dx.$$

gdje je $\delta(x) \in \mathcal{D}$ zapravo odluka, odnosno procjena za θ, temeljena na podacima x koji su jedna realizacija slučajnog vektora X.

Za apriornu distribuciju π definiramo i funkciju integriranog rizika kao

$$r(\pi, \delta) = E_\pi\big[R(\Theta, \delta)\big] = \int_{\Phi} R(\theta, \delta)\pi(\theta) d\theta = \int_{\Phi} \pi(\theta) \int_{\mathbb{R}} l(\theta, \delta(x)) f(x|\theta) dx d\theta.$$

Definicija 4.2. *Bayesovski procjenitelj $\hat{\delta}^{\pi}$ definira se kao:*

$$\hat{\delta}_{\pi} := \hat{\delta}_{\pi}(x) = arg \min_{\delta} \zeta(x, \delta|x), x \in \mathbb{R},$$

gdje je

$$\zeta(\pi, \delta|x) = E_{\pi}\big[l(\Theta, \delta)|X = x\big] = \int\limits_{\Phi} l(\theta, \delta)\pi(\theta|x)d\theta.$$

Funkcija $\zeta(x, \delta|x)$ naziva se aposteriorni očekivani gubitak. Vrijednost $r(\pi) = r(\pi, \hat{\delta}_{\pi})$ naziva se Bayesov rizik.

Sljedeći teorem daje kriterij za bayesovskog procjenitelja na temelju aposteriorne distribucije.

Teorem 4.1. *Neka procjenitelj $\delta(x)$ za svaki $x \in \mathbb{R}$ minimizira $\zeta(\pi, \delta|x)$, odnosno minimizira očekivani gubitak. Tada $\delta(x)$ minimizira i funkciju inegriranog rizika $r(\pi, \delta)$ i vrijedi:*

$$r(\pi, \delta) = \int\limits_{\mathbb{R}} \zeta(\pi, \delta(x)|x)m(x)dx,$$

pri čemu je $m(x)$ funkcija gustoće marginalne distribucije od X.

Dokaz. Vrijedi $l(\theta, \delta) \geq 0$. Jednakost iz teorema dobivamo koristeći definicije apriorne distribucije, aposteriorne distribucije te Fubinijevog teorema. Dobivamo:

$$r(\pi, \theta) = \int\limits_{\Phi} \pi(\theta) \int\limits_{\mathbb{R}} l(\theta, \delta(x))f(x|\theta)dxd\theta$$

$$= \int\limits_{\mathbb{R}} \int\limits_{\Phi} l(\theta, \delta(x))f(x|\theta)\pi(\theta)d\theta dx$$

$$= \int\limits_{\mathbb{R}} \Bigg(\int\limits_{\Phi} l(\theta, \delta(x))\pi(\theta|x)d\theta \Bigg) m(x)dx$$

$$= \int\limits_{\mathbb{R}} \zeta(\pi, \delta(x)|x)m(x)dx.$$

Kako je marginalna funkcija gustoće nenegativna i integral ima svojstvo linearnosti, vidimo da $\delta = \delta(x)$ koji minimizira aposteriorni očekivani gubitak ζ minimizira i funkciju integralnog rizika r. $\square$

Sada ćemo definirati nekoliko klasičnih funkcija gubitka.
Kvadratna funkcija gubitka definirana je sljedećim izrazom:

$$l(\theta, \delta) = (\theta - \delta)^2.$$

Propozicija 4.1. *Za kvadratnu funkciju gubitka bayesovski procjenitelj $\hat{\delta}_\pi$ je aposteriorno očekivanje od Θ za koje vrijedi:*

$$\delta_\pi(x) = E_\pi\big[\Theta|X = x\big] = \frac{\int\limits_\Phi \theta f(x|\theta)\pi(\theta)d\theta}{\int\limits_\Phi f(x|\theta)\pi(\theta)d\theta}, \; x \in \mathbb{R}.$$

Dokaz. Vrijedi

$$E_\pi\big[(\Theta - \delta)^2|X = x\big] = E_\pi\big[\Theta^2|X = x\big] - 2\delta E_\pi\big[\Theta|X = x\big] + \delta^2.$$

Očigledno aposteriorni gubitak postiže minimum u $\hat{\delta}_\pi(x) = E_\pi[\Theta|X = x]$. □

Funkcija apsolutnog gubitka definira se kao:

$$l(\theta, \delta) = |\theta - \delta|,$$

ili općenitije:

$$l_{k1,k2}(\theta, \delta) = \begin{cases} k_2(\theta - \delta), & \theta > \delta \\ k_1(\delta - \theta), & \text{inače.} \end{cases}$$

Kao svojevrsne kombinacije funkcije apsolutnog i funkcije kvadratnog gubitka mogu se promatrati i funkcije sljedećeg oblika:

$$\tilde{l}(\theta, \delta) = \begin{cases} (\delta - \theta)^2, & |\delta - \theta| < k \\ 2k(\delta - \theta) - k^2, & \text{inače.} \end{cases}$$

Ovakve funkcije također usporavaju rast gubitka za velike greške.

Gubitak "0-1" za sliku ima skup $\{0, 1\}$ i često se koristi kod klasičnog pristupa testiranju hipoteza. Ova funkcija gubitka definira se kao:

$$l(\theta, \delta) = \begin{cases} 1, & \theta \neq \delta \\ 0, & \text{inače.} \end{cases}$$

Kao što je rečeno, odabir funkcije gubitka je u principu subjektivan i u nekim slučajevima ovi jednostavni klasični gubici mogu biti loš izbor. Kao alternativa mogu se koristiti gubici koji uspoređuju $f(X|\theta)$ i $f(X|\delta)$. Takve funkcije gubitka su oblika

$$l(\theta, \delta) = h(f(\cdot|\theta),\, f(\cdot|\delta)).$$

Od funkcija ovog oblika uobičajene su entropijska i Hellingerova razlika koje se definiraju na sljedeći način:

1. Entropijska razlika

$$l_e(\theta, \delta) = E_\theta\left[\ln\left(\frac{f(X|\theta)}{f(X|\delta)}\right)\right]$$

2. Hellingerova razlika

$$l_H(\theta, \delta) = \frac{1}{2}E_\theta\left[\left(\sqrt{\frac{f(X|\theta)}{f(X|\delta)}} - 1\right)^2\right].$$

Osim ovog pristupa procjeni, možemo koristiti i procjenitelj maksimalne vjerodostojnosti. Način pristupa procjenitelju maksimalne vjerodostojnosti je analogan klasičnom pristupu, uz modifikacije koje uključuju specifičnosti bayesovskog pristupa statističkom zaključivanju. Drugim riječima, bayesovski procjenitelj metodom maksimalne vjerodostojnosti je maksimizator funkcije $L(\theta|x)\pi(\theta)$, gdje je L funkcija vjerodostojnosti. Bayesovski procjenitelj maksimalne vjerodostojnosti se nekad naziva i penalizirani procjenitelj maksimalne vjerodostojnosti.

Kako je poznata aposteriorna funkcija gustoće parametra, možemo za procjenitelja $\hat{\delta}_\pi(x)$ (koji procjenjuje $h(\theta)$) odrediti preciznost procjene. Da bismo to napravili možemo gledati aposteriornu kvadratnu grešku:

$$E_\pi\big[(\hat{\delta}_\pi(x) - h(\Theta))^2|X = x\big].$$

Primijetimo, za $\hat{\delta}_\pi(x) = E_\pi\big[h(\Theta)|X = x\big]$ ova greška je jednaka $Var_\pi(\Theta|X = x)$).

4.2 Pouzdana područja i statistički testovi

Slijedi kratak pregled teorije vezane za testiranje statističkih hipoteza i procjene pouzdanih intervala, odnosno pouzdanih područja. Primijetimo da je svrha bayesovskog pristupa pronaći aposteriornu distribuciju parametra. To znači da će, jednom kad imamo aposteriornu distribuciju, biti lako izračunati razne vjerojatnosti koje se tiču vrijednosti parametra, što dovodi do vrlo intuitivnog načina za testiranje hipoteza i pronalazak kritičnih područja.

Neka su dani podaci $x_1, x_2, ..., x_n$ koji su realizacija slučajnog uzorka $X = (X_1, X_2, ..., X_n)$. Na temelju ovoga kreiran je model $\mathcal{P} = \{f(x|\theta) : \theta \in \Phi\}$, pri čemu je θ parametar kojeg treba procijeniti. Pretpostavimo da testiramo hipoteze

$$H_0 : \theta \in \Phi_0$$

$$H_1 : \theta \in \Phi_1,$$

pri čemu je $\Phi_0 \cap \Phi_1 = \emptyset$ i $\Phi_0 \cup \Phi_1 = \Phi$.

U Newman-Pearsonovoj teoriji je testiranje formalizirano kroz prostor odluka $\mathcal{D}$ koji je dvočlani skup, odnosno ograničen na $\{0, 1\}$, gdje jedan element (1) znači odbacivanje hipoteze H_0, a drugi zadržavanje hipoteze H_0. Dakle, problem testiranja možemo promatrati kroz zaključivanje o indikator funkciji $\mathbb{I}_{\Phi_0}$. Neka je funkcija $\tau : \mathbb{R}^n \to [0, 1]$ test za parametar θ. Hipotezu H_0 odbacujemo u korist hipoteze H_1 ukoliko za x, koji je realizacija od X, vrijedi $\tau(x) = 1$. Hipotezu H_0 ne odbacujemo u korist hipoteze H_1 ukoliko za x, koji je realizacija od X, vrijedi $\tau(x) = 0$.

Jakost statističkog testa definiramo kao funkciju $\gamma : \Phi \to [0, 1]$ s:

$$\gamma(\theta) := P_\theta(\tau(X) = 1) = P_\theta(X \in C) = E_\theta\big[\tau(X)\big],$$

pri čemu je je $C = \tau^{-1}(\{1\}) = \{x \in \mathbb{R} : \tau(x) = 1\}$ kritično područje testa.

Bayesovski pristup testiranju je vrlo intuitivan i jednostavno shvatljiv i bez statističkih znanja. Jedna od najčešćih želja u zaključivanju o parametru je pronaći pouzdane intervale za vrijednost parametra. Kako je u bayesovskoj statistici parametar modeliran kao slučajna varijabla, pouzdani (1-α) interval ($[a, b]$) se lako dobije kao interval za koji vrijedi $P(\Theta \in [a, b]|X = x) = 1 - \alpha$. Ukoliko želimo testirati hipotezu $\mathcal{H}_0 : \theta \in \Phi_0 \subseteq \Phi$ protiv hipoteze $\mathcal{H}_1 : \theta \in \Phi_1 \subseteq \Phi$ uspoređujemo $P_\pi(\Theta \in \Phi_0|X = x)$ i $P_\pi(\Theta \in \Phi_1|X = x)$. Na ovaj način, moguće je testirati i više hipoteza istovremeno. Naravno, mogući su i drugi pristupi testiranju u bayesovskoj statistici poput promatranja Bayesovih faktora ([4]).

Definicija 4.3. *Neka je π apriorna distribucija. Za skup C_x kažemo da je α-vjerodostojan ako vrijedi*

$$P_\pi(\Theta \in C_x | X = x) \geq 1 - \alpha,$$

odnosno $P_{\Theta|X=x}(C_x) \geq 1-\alpha$, gdje je $P_{\Theta|X=x}$ aposteriorna distribucija parametra Θ. Dodatno, za C_x kažemo da je α-vjerodostojno područje najveće aposteriorne gustoće (označavamo s HPD - highest posterior density) ako vrijedi:

$$\{\theta : \pi(\theta|x) > k_\alpha\} \subseteq C_x \subseteq \{\theta : \pi(\theta|x) \geq k_\alpha\},$$

pri čemu je k_α najveći broj za koji vrijedi

$$P_\pi(\Theta \in C_x | X = x) \geq 1 - \alpha.$$

Primjer 4.1. *Neka imamo n nezavisnih realizacija Bernoullijeve slučajne varijable s vjerojatnosti uspjeha p, gdje je p realizacija parametra A i neka je $A \sim Beta(\alpha, \beta)$. Neka se u n ponavljanja realiziralo z uspjeha. Odredimo aposteriornu distribuciju parametra A.*

$$\pi(p|z, N) = \frac{\pi(p) \cdot P(z, n|p)}{P(z, n)} = \frac{\frac{p^{\alpha-1}(1-p)^{\beta-1}}{\beta(\alpha,\beta)} \cdot \binom{n}{z} p^z (1-p)^{n-z}}{\binom{n}{z}\left(\frac{\alpha}{\alpha+\beta}\right)^z \left(\frac{\beta}{\alpha+\beta}\right)^{n-z}} = \frac{p^{\alpha+z-1}(1-p)^{\beta+n-z-1}}{\frac{\gamma(\alpha)\gamma(\beta)}{\gamma(\alpha+\beta)} \cdot \frac{\alpha^z \beta^{n-z}}{(\alpha+\beta)^n}}.$$

Općenito, znamo da za gama funkciju vrijedi $\gamma(x+1) = x\gamma(x)$. Iz ovoga dobivamo da je $\gamma(\alpha+z) = \alpha^z\gamma(\alpha)$. Identično, $\gamma(\alpha+n-z) = \beta^{n-z}\gamma(\beta)$ i $\gamma(\alpha+\beta+n) = (\alpha+\beta)^n\gamma(\alpha+\beta)$. Odavde slijedi:

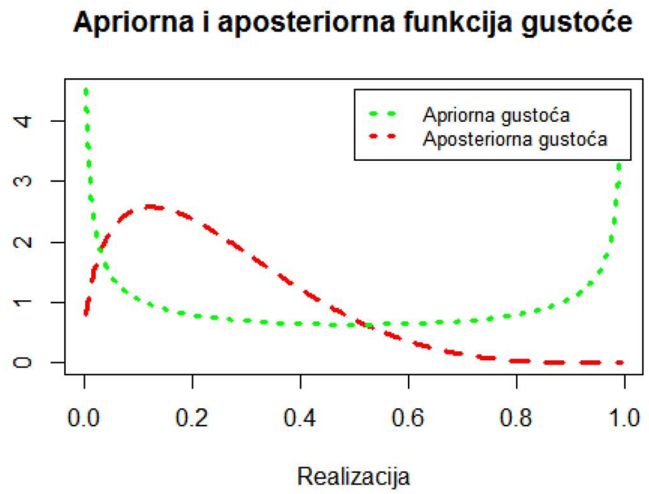
$$f(p|z, n) = \frac{p^{\alpha+z-1}(1-p)^{\beta+n-z-1}}{\frac{\gamma(\alpha)\gamma(\beta)}{\gamma(\alpha+\beta)} \cdot \frac{\alpha^z \beta^{n-z}}{(\alpha+\beta)^n}} = \frac{p^{\alpha+z-1}(1-p)^{\beta+n-z-1}}{\frac{\gamma(\alpha+z)\gamma(\beta+n-z)}{\gamma(\alpha+\beta+n)}} = \frac{p^{\alpha+z-1}(1-p)^{\beta+n-z-1}}{\beta(\alpha+z, \beta+n-z)}.$$

Dakle, za aposteriornu distribuciju parametra dobivamo $A \sim Beta(\alpha + z, \beta + n - z)$. Promotrimo slučaj $\alpha = \beta = \frac{1}{2}$, $n = 5$, $z = 1$. Promotrimo grafički apriornu i aposteriornu distribuciju parametra A.

Stvarnu vrijednost parametra procjenjujemo kao očekivanje aposteriorne distribucije parametra $A \sim Beta(\alpha + z, \beta + n - z)$, odnosno

$$\hat{p}^\pi = \frac{\alpha + z}{\alpha + z + \beta + n - z} = \frac{\frac{3}{2}}{6} = \frac{1}{4}.$$

Nadalje, iz aposteriorne distribucije možemo lako izračunati razne vjerojatnosti. Primjerice, jedan 95%*-tni pouzdani interval za stvarnu vrijednost parametra p je* $[0.02251277, 0.6286264]$*, gdje je* 0.02251277 0.025*-kvantil, a* 0.6286264 0.975*-kvantil aposteriorne $Beta(\frac{3}{2}, \frac{9}{2})$ distribucije.*



Slika 1: Apriorna i aposteriorna funkcija gustoće

Primjer 4.2. *Promatrajući postotak rođene muške djece u Parizu, Laplace je želio testirati je li vjerojatnost rođenja muške djece, koju ćemo označiti s* Θ*, veća od* $\frac{1}{2}$*. Imao je podatke da je rođeno* 251 527 *muške djece i* 241 945 *ženske djece i pretpostavio da* Θ *ima uniformnu apriornu distribuciju na* $[0,1]$ *(odnosno* $Beta(1,1)$ *distribuciju). Već smo pokazali da u slučaju binomnih podataka (z uspjeha u n eksperimenata) i apriorne beta distribucije aposteriorna distribucija ima beta distribuciju* $Beta(\alpha+z, \beta+n-z)$*. Dakle, u ovom slučaju, aposteriorna distribucija od* Θ *je* $Beta(1+251527, 1+241945)$*, odnosno* $Beta(251528, 241946)$*.*

Sada možemo testirati hipotezu koja nas zanima, je li vjerojatnost rođenja muškog djeteta veća od 0.5*, odnosno zanima nas vjerojatnost da se parametar* Θ *realizira s* $\theta > 0.5$*. Ta vjerojatnost jednaka je*

$$1 - P(\Theta \leq \frac{1}{2} |\ (251\ 527,\ 241\ 945\ )) = 1 - P(Beta(251528, 241946) \leq \frac{1}{2}) = 1 - 1.15 \cdot 10^{-42}.$$

Ovo navodi na zaključak da se Θ *gotovo sigurno realizira s vrijednošću većom od* $\frac{1}{2}$*.*

Primjer 4.3. *Na sveučilištu Stetson (koje se nalazi u DeLandu, u Floridi) prikupljeni su podaci* [3] $x = (x_1, ..., x_{20})$ *o prosječnom dnevnom trajanju vožnje u dvadeset različitih gradova u saveznoj državi Floridi. Shapiro-Wilkovim testom nije odbačena normalnost podataka (*$p = 0.5157$*) pa, za potrebe ovog primjera, pretpostavimo da se radi o podacima koji su normalno distribuirani. Varijanca podataka iznosi* 16.47671*. Očekivanje populacije ćemo modelirati normalnom slučajnom varijablom. Neka apriorna distribucija očekivanja bude* $\Theta \sim \mathcal{N}(100, 9)$*. Iz primjera 3.1 znamo oblik aposteriorne distribucije i u ovom slučaju radi se o normalnoj* $\mathcal{N}(41.17939, 0.754748)$ *distribuciji. Sada želimo testirati nekoliko hipoteza. Neka je:*

$$\mathcal{H}_0 : \Theta < 40,$$

$$\mathcal{H}_1 : 42 < \Theta < 43,$$

$$\mathcal{H}_2 : 41.5 < \Theta < 42.$$

[3] Podaci su dostupni na stranicama sveučilišta i nalaze se u dokumentu BestCity.xls

Dobivamo

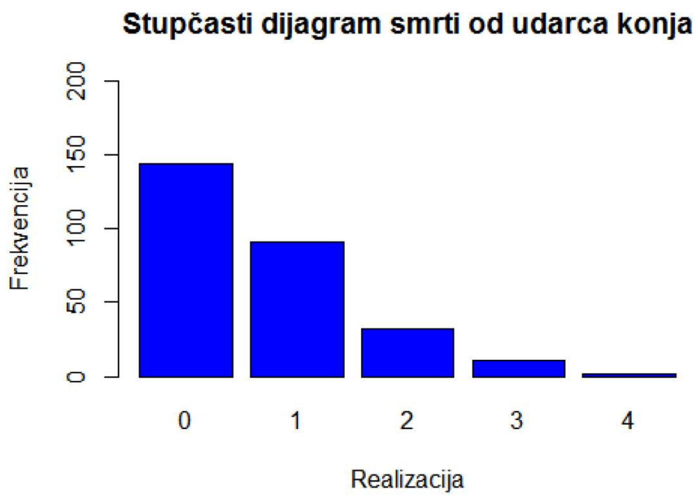
$$P(\Theta < 40 | X = (x_1, ..., x_{20})) = 0.05907065,$$

$$P(42 < \Theta < 43 | X = (x_1, ..., x_{20})) = 0.1305311,$$

$$P(41.5 < \Theta < 42 | X = (x_1, ..., x_{20})) = 0.1970338$$

i zaključujemo da je hipoteza H_2 najjača među ove tri hipoteze pri čemu imamo i vjerojatnost realizacije svake od ovih hipoteza.

Primjer 4.4. *Poznati su podaci $x = (x_1, ..., x_{280})$ o smrtima u pruskoj vojsci uzrokovani udarcem konja ([10]). Podaci su poznati za 20 različitih godina (od 1875. do 1894.) i za 14 različitih konjaničkih jedinica unutar pruske vojske.*



Slika 2: Stupčasti dijagram smrti od udarca konja

Na temelju stupčastog dijagrama podataka pretpostavimo da su podaci iz Poissonove distribucije. Za apriornu distribuciju parametra Poissonove distribucije uzmimo $\Theta \sim \Gamma(10, 1)$, tj. gama distribuciju s parametrima $\alpha = 10$ i $\beta = 1$. Iz primjera 3.2 znamo oblik aposteriorne distribucije i u ovom slučaju radi se o $\Gamma(10 + 196, 1 + 280)$ distribuciji. Sada želimo testirati nekoliko hipoteza. Neka je:

$$\mathcal{H}_0 : \Theta < 0.7,$$

$$\mathcal{H}_1 : 0.8 < \Theta,$$

$$\mathcal{H}_2 : 0.7 < \Theta < 0.75.$$

Dobivamo

$$P(\Theta < 0.7 | X = (x_1, ..., x_{280})) = 0.2628244,$$

$$P(0.8 < \Theta | X = (x_1, ..., x_{280})) = 0.09764941,$$

$$P(0.7 < \Theta < 0.75 | X = (x_1, ..., x_{280})) = 0.3746143$$

i zaključujemo da je hipoteza H_2 najjača među ove tri hipoteze pri čemu imamo i vjerojatnost realizacije svake od ovih hipoteza.

Primjer 4.5. *Radi usporedbe klasičnog i bayesovskog pristupa simulirajmo $n = 100$ realizacija normalne slučajne varijable s očekivanjem* 0 *i varijancom $\sigma^2 = 25$. Procjena očekivanja klasičnim pristupom jednaka je aritmetičkoj sredini simuliranih podataka i iznosi $\bar{x}_n = -0.2940414$. Tada znamo da je $[\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z, \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z]$, gdje je z 0.975-kvantil normalne $\mathcal{N}(0,1)$ distribucije, 95%-tni pouzdani interval za vrijednost parametra (ovaj rezultat dobivamo koristeći centralni granični teorem, vidjeti [1]). Drugim riječima, 95%-tni pouzdani interval za vrijednost parametra dobiven klasičnom statistikom je $[-1.274023, 0.6859406]$. Procijenimo sada očekivanje bayesovskim pristupom. Za apriornu distribuciju parametra uzmimo $\mathcal{N}(3,4)$. Korištenjem tvrdnje iz primjera 3.1 dobivamo da je $\mathcal{N}(-0.1002742, 0.2352941)$ aposteriorna distribucija parametra. Parametar procjenjujemo kao očekivanje aposteriorne distribucije, odnosno kao -0.1002742. Najmanji* 95%*-tni pouzdani interval za stvarnu vrijednost parametra je $[-1.050996, 0.8504479]$, gdje je -1.050996 0.025-kvantil, a 0.8504479* 0.975*-kvantil aposteriorne $\mathcal{N}(-0.1002742, 0.2352941)$ distribucije.*

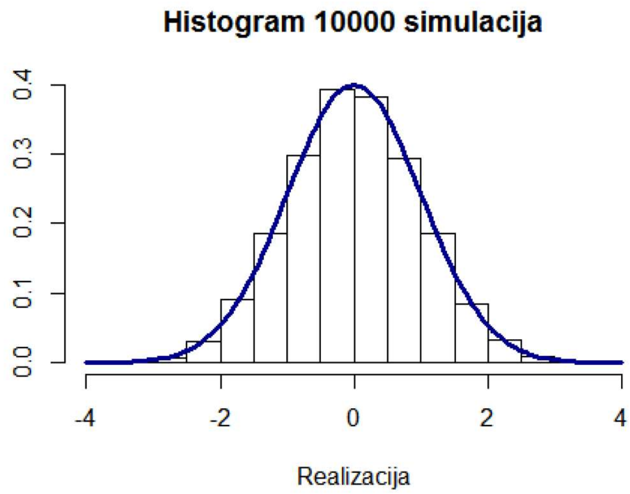
5 Metropolis-Hastings algoritam

Monte Carlo je općeniti naziv za tehnike računanja koje koriste slučajne brojeve, odnosno slučajne uzorke. Dosad smo promotrili nekoliko primjera u kojima smo aposteriornu distribuciju parametra egzaktno izračunali. Naravno, tada smo mogli iskoristiti egzaktno znanje o aposteriornoj distribuciji parametra kako bismo mogli reći više o parametru: kolika je vjerojatnost da se stvarna vrijednost parametra nalazi u nekom intervalu i slično. Međutim, primjeri koje smo pokazali su bili izrazito jednostavno (primjer s kuglicama u kutiji) ili pomno odabrani kako bi bilo moguće aposteriornu distribuciju dobiti egzaktno (primjer s apriornom beta distribucijom parametra). Često nije moguće na taj način dobiti dobiti aposteriornu distribuciju. Naime, primjenom Bayesove formule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

dobivamo aposteriornu distribuciju parametra. Međutim, eksplicitno izračunavanje integrala aposteriorne funkcije gustoće, potrebno kako bismo mogli zaključivati o aposteriornoj distribuciji parametra (primjerice, izračunati očekivanje i varijancu) često nije moguće, ili je vrlo teško. Dakle, potrebno je rješiti ovaj problem kako bismo i u takvim slučajevima mogli koristiti bayesovski pristup za zaključivanje o parametru. Jedna razumna ideja je pokušati rješiti integral numeričkim algoritmima. Druga ideja je sljedeća: ako bismo mogli nekako simulirati velik broj realizacija iz aposteriorne distribucije, mogli bismo, uz vrlo male pogreške, procijeniti sve ono što bismo mogli izračunati iz distribucije da znamo egzaktno integrirati funkciju gustoće aposteriorne distribucije (očekivanje, standardnu devijaciju, medijan, kvantile, vjerojatnost da se slučajna varijabla realizira u nekom intervalu i slično). U ovom pristupu postavljaju se dva pitanja: kako simulirati realizacije aposteriorne distribucije i na koji način izvršiti željene procjene.

Primjer 5.1. *Pretpostavimo da promatramo normalnu slučajnu varijablu X s očekivanjem 0 i varijancom 1 ($\mathcal{N}(0,1)$) i neka imamo mogućnost simulacije 10000 nezavisnih realizacija ove slučajne varijable. U R-u se ovo lako napravi naredbom "rnorm(10000,0,1)". Ukoliko promotrimo histogram i funkciju gustoće ovih simulacija (slika 3), vidimo da gustoća simulacija vrlo dobro aproksimira teorijsku funkciju gustoće normalne slučajne varijable. Ukoliko izračunamo prosjek simuliranih vrijednosti (što koristimo kao procjenu očekivanja), dobivamo -0.00659359. Medijan simulacija je -0.00463824. Procjena za standardnu devijaciju je 0.9935902, a za varijancu 0.9872215. Na temelju ovih simulacija vrijednost $P(X < 0)$ (čiju stvarnu vrijednost znamo iz teorije, ona iznosi $\frac{1}{2}$) procjenjujemo s 0.5025. Jednako lako možemo procijeniti i ostale vjerojatnosti. Naravno, veći broj simulacija donosi veću vjerojatnost boljih procjena.*



Slika 3: Usporedba histograma 10000 simulacija s teorijskom funkcijom gustoće

Postavlja se pitanje: kako simulirati realizacije iz aposteriorne distribucije? Jedan od načina na koji se ovo može napraviti su razni Monte Carlo Markovljev lanac (engleski: Monte Carlo Markov chain, označavat ćemo s MCMC) algoritmi.

Monte Carlo metode je naziv za široku klasu algoritama koji se temelje na slučajnom uzorkovanju kako bi se dobio numerički rezultat. Originalni Monte Carlo pristup je metoda korištena za računanje integrala. Pretpostavimo da želimo izračunati integral

$$\int_a^b h(x)dx.$$

Ako možemo napraviti dekompoziciju funkcije h kao produkt funkcija g i f pri čemu je f funkcija gustoće neke slučajne varijable X, tada dobivamo

$$\int_a^b h(x)dx = \int_a^b g(x)f(x)dx = E\big[g(X)\big].$$

Ovime smo integral pokazali kao očekivanje slučajne varijable $g(X)$. Dakle, simulacijom velikog broja realizacija $x_1, ..., x_n$ slučajne varijable X vrijednost integrala procjenjujemo kao

$$\int_a^b h(x)dx = E\big[g(X)\big] \approx \frac{1}{n}\sum_{i=1}^{n} g(x_i).$$

MCMC algoritmi su Monte Carlo algoritmi koji se temelje na teoriji Markovljevih lanaca. Više o Markovljevim lancima može se naći u [8] i [9].

Sami MCMC algoritmi počinju se razvijati u 40-tim godinama dvadesetog stoljeća. S razvojem MCMC algoritama raste i privlačnost i upotreba bayesovskog pristupa. Ovdje ćemo proučiti Metropolis-Hastings algoritam koji je jedan od osnovnih MCMC algoritama. Koraci u Metropolis-Hastings algoritmu su sljedeći:

1. Startamo s inicijalnom vrijednosti $\theta_0 \in \Phi$ koju parametar može poprimiti. Inicijalnu vrijednost θ_0 biramo proizvoljno među onima za koje vrijedi $\pi(\theta_0) > 0$.

2. U k-tom koraku, $k \in \{1, ..., n\}$, $n \in \mathbb{N}$, iz predložene distribucije Q simuliramo jednu realizaciju θ. Ova distribucija naziva se predložena distribucija (eng. proposal, nekad i candidate-generating distribution). Za višedimenzionalne parametre za predloženu distribuciju mora vrijediti simetričnost ($f(\theta^{(1)}, \theta^{(2)}) = f(\theta^{(2)}, \theta^{(1)})$).

3. Za kandidata θ_k izračunamo vjerojatnost prihvaćanja (eng. acceptance rate ili acceptance probability)

$$\alpha(\theta_k) = \frac{\pi(\theta_k) \cdot L(\theta_k|x)}{\pi(\theta_{k-1}) \cdot L(\theta_{k-1}|x)},$$

   gdje su x podaci, L funkcija vjerodostojnosti, a π apriorna distribucija parametra.

4. Ako je $\alpha \geq 1$, prihvaćamo kandidata θ_k.
   Ako je $\alpha < 1$, kandidata θ_k prihvaćamo s vjerojatnošću α. Ovo možemo napraviti tako da simuliramo jednu realizaciju u iz uniformne $\mathcal{U}(0, 1)$ slučajne varijable - kandidata θ_k prihvaćamo ako je $\alpha \geq u$, a odbijamo (odnosno stavljamo $\theta_k = \theta_{k-1}$) ako $\alpha < u$.

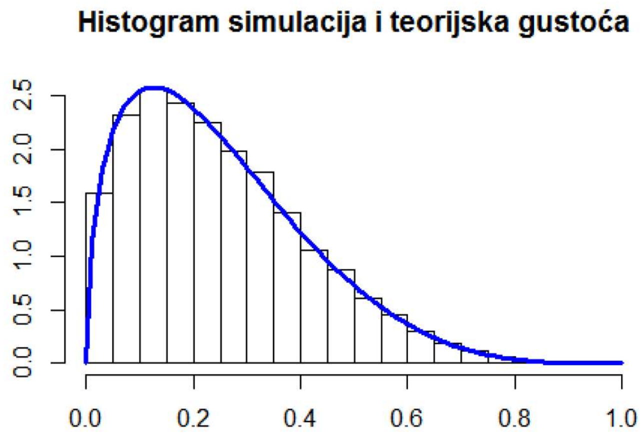
5. Ponovimo drugi korak ukoliko nismo postigli zaustavni kriterij. U suprotnom, algoritam staje.

Ovako kreiran Markovljev lanac je ireducibilan i aperiodičan i za njega vrijedi da je njegova granična distribucija ujedno i njegova stacionarna distribucija ([8], [9]). Dakle, potrebno je simulirati dovoljno velik uzorak kako bi ovako kreiran Markovljev lanac postigao svoju stacinarnu distribuciju. Stacionarna distribucija ovako kreiranog Markovljevog lanca odgovara aposteriornoj distribuciji parametra (za dokaz ove tvrdnje vidjeti [5] i [13]). Neki od načina na koje se mogu dobiti informacije o tome je li lanac postigao svoju stacionarnu distribuciju su Geweke test (ili dijagnostika) i Raftery-Lewis test ([13]).

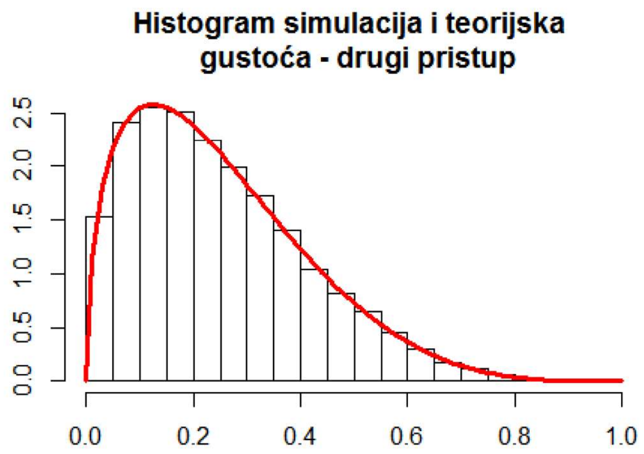
Što se tiče predložene distribucije, algoritam funkcionira za širok spektar distribucija. Broj koraka algoritma ovisi o izboru predložene distribucije. Apriorna distribucija parametra i uniformna distribucija na skupu svih mogućih vrijednosti koje parametar može poprimiti su neki česti izbori za predloženu distribuciju.

Dodatno, važno je napomenuti da predložena distribucija Q ne mora biti ista u svakom koraku algoritma. Često se za predloženu distribuciju u k-tom koraku algoritma uzima slučajna varijabla koja odgovara k-tom koraku slučajne šetnje $\theta_0 + \sum_{i=1}^{k} q_k$ gdje su $q_1, ..., q_k$ prethodne realizacije slučajne varijable Q (vidjeti [13]). Drugim riječima, θ_1 simuliramo kao $\theta_0 + Q$, θ_2 simuliramo kao $\theta_1 + Q$ itd. U ovom slučaju se za Q najčešće uzima normalna distribucija s očekivanjem 0.

Primjer 5.2. *Neka imamo binomne podatke i apriornu beta distribuciju parametra. Za $n = 5$ (broj eksperimenata), $z = 1$ (broj uspjeha) i apriornu $Beta(\frac{1}{2}, \frac{1}{2})$ distribuciju parametra egzaktno smo izračunali aposteriornu distribuciju parametra ($Beta(\frac{3}{2}, \frac{9}{2})$). Pokušajmo ovdje provesti Metropolis-Hastings algoritam i usporedimo dobiveno s teorijskom distribucijom.*
Primjer ćemo provesti u dva slučaja: u prvom ćemo kao predloženu distribuciju uzeti apriornu distribuciju $Beta(\frac{1}{2}, \frac{1}{2})$, a u drugom ćemo θ_i simulirati kao realizacije slučajne šetnje, odnosno svaki θ_k (osim početne vrijednosti) ćemo simulirati iz $\theta_{k-1}+X$, gdje ćemo u primjeru za X uzeti $\mathcal{N}(0, 0.04)$. U oba slučaja za početnu vrijednost ćemo odabrati $\theta_0 = \frac{1}{2}$. Napravit



Slika 4: Usporedba histograma simulacija i teorijske gustoće



Slika 5: Usporedba histograma simulacija i teorijske gustoće - drugi pristup

*ćemo pedeset tisuća (*50000*) simulacija iz predložene distribucije. Grafički prikažimo histogram simulacija u oba slučaja i usporedimo s teorijskom funkcijom gustoće aposteriorne distribucije.*

Dodatno, u prvom slučaju vrijednost K-S statistike je 0.0049571*, a* 0.0066962 *u drugom slučaju, što potvrđuje da simulacije odlično aproksimiraju teorijsku distribuciju.*

Primjer 5.3. *Neka je dana jedna realizacija jednostavnog slučajnog vektora* $x = (x_1, ..., x_{100})$, *gdje je* $x_i, i \in \{1, ..., 100\}$ *jedna simulacija iz Weibullove slučajne varijable s parametrima* $k = 2$ *i* $\theta = 8$. *Funkcija gustoće Weibullove slučajne varijable je*

$$f(x) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} e^{-(\frac{x}{\theta})^k}.$$

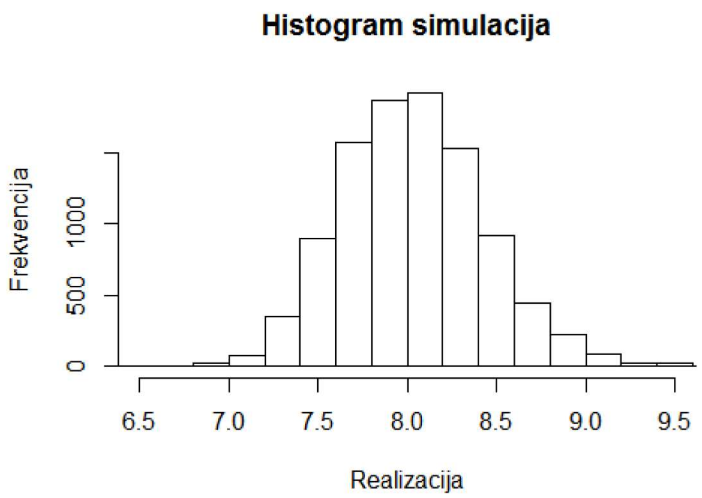
Na temelju tog slučajnog vektora Metropolis-Hastings algoritmom procijenimo parametar θ *(za parametar* $k = 2$ *uzimamo da je poznat). Za apriornu distribuciju parametra uzmimo* $\Theta \sim \mathcal{E}(0.1)$. *Funkciju gustoće apriorne distribucije parametra označimo s* π. *Primijetimo da za aposteriornu distribuciju parametra vrijedi:*

$$\pi(\theta|x) = \frac{\frac{k^{100}}{\theta^{100}}\left(\frac{x_1}{\theta}\right)^{(k-1)} \cdot ... \cdot \left(\frac{x_{100}}{\theta}\right)^{(k-1)} e^{-(\frac{x_1}{\theta})^k} \cdot ... \cdot e^{-(\frac{x_{100}}{\theta})^k} 0.1e^{-0.1\theta}}{c},$$

gdje je c *konstanta. Primijetimo da je u* n*-tom koraku algoritma vjerojatnost prihvaćanja* $\alpha(\theta_n) = \frac{\pi(\theta_n|x)}{\pi(\theta_{n-1}|x)}$ *jednaka (nakon skraćivanja):*

$$\alpha(\theta_n) = \frac{\theta_{n-1}^{100k}}{\theta_n^{100k}} \cdot e^{-(\frac{x_1}{\theta_n})^k - ... - (\frac{x_{100}}{\theta_n})^k + (\frac{x_1}{\theta_{n-1}})^k + ... + (\frac{x_{100}}{\theta_{n-1}})^k} \cdot e^{-0.1(\theta_n - \theta_{n-1})}$$

Za početni korak uzmimo $\theta_0 = 2$, *a za predloženu distribuciju u* n*-tom koraku uzmimo* $\theta_n = \theta_{n-1} + X$, *gdje je* $X \sim \mathcal{N}(0, 0.25)$. *Algoritam provodimo u* 10000 *koraka. Grafički prikažimo gustoću dobivenih simulacija. Znamo da u Bayesovskoj statistici parametar procjenjujemo*



Slika 6: Histogram simulacija dobivenih Metropolis-Hastings algoritmom

kao očekivanje aposteriorne distribucije. S obzirom da aposteriornu distribuciju parametra nemamo egzaktno, njezino očekivanje procjenjujemo aritmetičkom sredinom simulacija koja iznosi 7.972093 *(prisjetimo se da je stvarna vrijednost parametra jednaka* 8*).*

Dodatak

U dodatku se nalaze R-kodovi i podaci korišteni u primjerima.

Primjer 4.3 U primjeru su korišteni sljedeći podaci.

Commute
49.2
45.3
39.8
46.8
39.9
49.5
44.4
44.8
44.9
44.7
40
38.7
41.1
42.8
37.8
48.4
44.5
41.4
53.5
45

Primjer 4.4 U primjeru su korišteni sljedeći podaci.

1875.	0	0	0	0	0	0	0	1	1	0	0	0	1	0
1876.	2	0	0	0	1	0	0	0	0	0	0	0	1	1
1877.	2	0	0	0	0	0	1	1	0	0	1	0	2	0
1878.	1	2	2	1	1	0	0	0	0	0	1	0	1	0
1879.	0	0	0	1	1	2	2	0	1	0	0	2	1	0
1880.	0	3	2	1	1	1	0	0	0	2	1	4	3	0
1881.	1	0	0	2	1	0	0	1	0	1	0	0	0	0
1882.	1	2	0	0	0	0	1	0	1	1	2	1	4	1
1883.	0	0	1	2	0	1	2	1	0	1	0	3	0	0
1884.	3	0	1	0	0	0	0	1	0	0	2	0	1	1
1885.	0	0	0	0	0	0	1	0	0	2	0	1	0	1
1886.	2	1	0	0	1	1	1	0	0	1	0	1	3	0
1887.	1	1	2	1	0	0	3	2	1	1	0	1	2	0
1888.	0	1	1	0	0	1	1	0	0	0	0	1	1	0
1889.	0	0	1	1	0	1	1	0	0	1	2	2	0	2
1890.	1	2	0	2	0	1	1	2	0	2	1	1	2	2
1891.	0	0	0	1	1	1	0	1	1	0	3	3	1	0
1892.	1	3	2	0	1	1	3	0	1	1	0	1	1	0
1893.	0	1	0	0	0	1	0	2	0	0	1	3	0	0
1894.	1	0	0	0	0	0	0	0	1	0	1	1	0	0

Primjer 5.1 Sve opisano u ovom primjeru dobiveno je sljedećim R kodom:

```
a=rnorm(10000,0,1)
hist(a, main="Histogram 10000 simulacija", xlab="Frekvencija", ylab="Realizacija")
plot(density(a), col="red", xlab="",ylab="",main="Gustoća simulacija")
mean(a)
median(a)
sd(a)
var(a)
c=a[a<0]
length(c)/10000
```

Primjer 5.2 Za prvu simulaciju koristimo sljedeći R kod:

```
theta=rep(0, 50000)
theta[1]=0.5
a=runif(49999, 0,1)
for (i in 2:50000){
theta[i]=a[i-1]
alfa_1=theta[i]^{-0.5}*(1-theta[i])^{-0.5}*theta[i]^1*(1-theta[i]) ^{5-1}
alfa_2=(theta[i-1]^{-0.5}*(1-theta[i-1])^{-0.5}*theta[i-1]^1 * (1-theta[i-1])^{5-1})
alfa=alfa_1/alfa_2
un=runif(1,0,1)
un
if(alfa < un){
theta[i]=theta[i-1]
}
}
mean(theta)
median(theta)
length(theta)
plot(theta)
hist(theta,xlim=c(0,1),xlab="",ylab="",main="Histogram simulacija
i teorijska gustoća", prob=TRUE)
curve(dbeta(x,1.5,4.5),col="blue",lwd=3,add=TRUE,yaxt="n")
```

Za drugu varijantu Metropolis-Hastings algoritma korišten je sljedeći R kod:

```
theta=rep(0, 50000)
theta[1]=0.5
for (i in 2:50000){
a=rnorm(1,0,0.2)
theta[i]=theta[i-1]+a
if(theta[i]<0 ){theta[i]=theta[i-1]}
if(theta[i]>1){theta[i]=theta[i-1]}
alfa_1=theta[i]^{-0.5}*(1-theta[i])^{-0.5}*theta[i]^1*(1-theta[i])  ^{5-1}
alfa_2=(theta[i-1]^{-0.5}*(1-theta[i-1])^{-0.5}*theta[i-1]^1 * (1-theta[i-1])^{5-1})
alfa=alfa_1/alfa_2
un=runif(1,0,1)
un
if(alfa < un){
theta[i]=theta[i-1]
}
}
mean(theta)
median(theta)
length(theta)
plot(theta)
hist(theta,xlim=c(0,1),xlab="",ylab="",main="Histogram simulacija
 i teorijska \n gustoća - drugi pristup",prob=TRUE)
curve(dbeta(x,1.5,4.5),col="red",lwd=3,add=TRUE,yaxt="n")
```

Literatura

[1] M. BENŠIĆ, N. ŠUVAK, *Uvod u Vjerojatnost i Statistiku*, Sveučilište J.J. Strossmayera, Odjel za matematiku, Osijek, 2014.

[2] J.M. BERNARDO, A.F.M. SMITH *Bayesian Theory*, Wiley, New York, 2014.

[3] I. BOGOJEVIĆ, *Bayesovska Regresijska Analiza*, Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Matematički odsjek, Zagreb, 2015.

[4] P. CONGDON *Bayesian Statistical Modeling*, 2nd edition, Wiley, Chichester, 2006.

[5] M. HAUGH, *MCMC and Bayesian Modeling*, Monte-Carlo Simulation, Columbia University, 2017.

[6] P. D. HOFF, *A First Course in Bayesian Statistical Methods*, Springer, New York, 2009.

[7] H. JEFFREYS, *An Invariant Form for the Prior Probability in Estimation Problems*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences. Vol. 186 (Sep. 24, 1946): 453–461

[8] A. KOVAČEVIĆ, *Monte Carlo Simulacije*, Diplomski rad, Sveučilište J.J. Strossmayera, Odjel za Matematiku, Osijek, 2015.

[9] J. R. NORRIS *Markov Chains*, Cambridge University Press, 1998.

[10] M.P. QUINE, E. SENETA, *Bortkiewicz's Data and the Law of Small Numbers*, International Statistical Review/Revue Internationale de Statistique, Vol. 55 (Aug. 1987): 173–181

[11] C.P. ROBERT, *The Bayesian Choice, A Decision-Theoretic Motivation*, Springer, New York 1994.

[12] T. TOPIĆ, *Bayesova Statistika i Procjena Vrijednosti Ulaganja*, Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Matematički odsjek, Zagreb, 2016.

[13] B. WALSH, *Markov Chain Monte Carlo and Gibbs Sampling*,Issues in Biostatistical Analysis, University of Arizona, 2002.

Sažetak

Tema rada je bayesovsko statističko zaključivanje. Nakon definicije osnovnih pojmova iz vjerojatnosti objašnjene su razlike u Bayesovskom pristupu u odnosu na klasični pristup statističkom zaključivanju. Objašnjeni su pojmovi apriorne i aposteriorne distribucije i diskutirana je važnost funkcije vjerodostojnosti i Bayesove formule. Potom je navedeno nekoliko različitih pristupa izboru apriorne distribucije i dana je teorija koja se tiče procjene parametra, testiranja hipoteza i procjene pouzdanih intervala.
Dodatno, nakon primjedbe da se pri zaključivanju o aposteriornoj distribuciji pojavljuju integrali koje je teško ili nemoguće egzaktno izračunati (primjerice, pri računanju očekivanja ili varijance aposteriorne distribucije) dan je Metropolis-Hastings algoritam za simulaciju realizacija iz aposteriorne distribucije. Na nekoliko primjera pokazano je kako Metropolis-Hastings algoritam funkcionira.

Ključne riječi:

Bayesovsko statističko zaključivanje, apriorna i aposteriorna distribucija, funkcija vjerodostojnosti, Bayesova formula, funkcija gubitka, procjena parametra, pouzdani interval, statistički test, Metropolis-Hastings algoritam.

Summary

Topic of the work is bayesian statistical inference. After the definition of basic terms from the probability theory, differences between Bayesian and classical approach to statistical inference were explained. Also, it was explained what prior and posterior distributions are, and the importance of likelihood function and Bayes' formula were discussed. Then, several methods of choosing prior distribution were listed and theory regarding parameter estimation, hypothesis testing and confidence interval estimation was given.
Furthermore, it was noted that integrals which are difficult or impossible to calculate often appear when making an inference about posterior distribution (for example, when calculating expectation or variance of the posterior distribution). To solve this, Metropolis-Hastings algorithm, which is used to sample from posterior distribution, was given. Several examples were used to demonstrate how Metropolis-Hastings algorithm works.

Key words:

Bayesian statistical inference, prior and posterior distribution, likelihood function, Bayes' formula, loss, parameter estimation, confidence interval, statistical test, Metropolis-Hastings algorithm.

Životopis

Roden sam 4. srpnja 1994. u Stuttgartu, u Njemačkoj. Pohađao sam Osnovnu školu Ljudevita Gaja u Osijeku, a nakon toga Prvu gimnaziju u Osijeku. Gimnaziju završavam 2013. godine te tada upisujem Sveučilišni preddiplomski studij matematike na Odjelu za matematiku u Osijeku. Studij završavam 2016. godine sa završnim radom na temu Bernoullijevi i Eulerovi polinomi te time stječem akademski stupanj prvostupnika matematike. Iste godine upisujem diplomski studij na Odjelu za matematiku, smjer Financijska matematika i statistika. U proljeće 2018. odrađujem praksu u Zavodu za javno zdravstvo Osječko-baranjske županije.