Bayesovska regresijska analiza

Bogojević, Iva

Master's thesis / Diplomski rad

2015

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

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:529822

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-08-07**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Iva Bogojević

BAYESOVSKA REGRESIJSKA ANALIZA

Diplomski rad

Voditelj rada:
prof. dr. sc. Miljenko Huzak

Zagreb, srpanj 2015.

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

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Zahvaljujem svom mentoru prof. dr. sc. Miljenku Huzaku na strpljenju, pomoći i vodstvu pri izradi ovog diplomskog rada.
Hvala mojoj obitelji na razumijevanju i podršci tokom studiranja.

Sadržaj

Uvod

Statistička analiza se može provoditi iz dva kuta gledišta: frekvencionističog i bayesovskog. Bitna razlika između ta dva pristupa leži u interpretaciji pojma vjerojatnosti. U frekvencionističkom pristupu vjerojatnost događaja je limes relativnih frekvencija. Stoga primjena tog pristupa nije uvijek moguća u praksi. Kod proučavanja rijetkih događaja možda nećemo imati dostupne velike uzorke podataka i u takvim slučajevima zagovornici frekvencionističkog pristupa će pribjeći teorijskim rezultatima. Bayesovski pristup temelji se na subjektivnoj interpretaciji vjerojatnosti - stupanj uvjerenja koji se ažurira uslijed dobivanja novih informacija.

Osnovna ideja bayesovske statistike je da parametri statističkih modela moraju i sami biti modelirani kao slučajni. Pri tome, mi i prije skupljanja podataka imamo neku inicijalnu (više ili manje subjektivnu) ideju o razdiobi parametra. Inicijalna, tzv. apriorna razdioba omogućuje nam da u model ugradimo npr. prethodno ili ekspertno znanje, no ona je izbor subjektivnosti u zaključivanju što je tipično i osnovni prigovor ovoj teoriji. Bayesovski pristup ipak dozvoljava da izračunamo vjerojatnost oblika $\mathbb{P}(\theta \in A)$, gdje je A bilo koji skup u prostoru parametara, dok je u klasičnoj statistici θ nepoznat, no fiksan, pa ovakav izraz čak ni nema smisla računati.

Pretpostavimo da su dane opservacije $X_1, X_2, \ldots X_n$, te da X_i pripada parametarskoj vjerojatnosnoj distribuciji s funkcijom gustoće $f_i(x_i|\theta)$, gdje je θ nepoznati parametar, a funkcija f_i poznata. Kada postavimo statistički model, treba provesti zaključke o parametru θ, odnosno o njegovoj distribuciji. Pri tome parametar θ postaje slučajna varijabla s određenom distribucijom, a označavat ćemo ga s Θ.

U prvom poglavlju ćemo se baviti bayesovskom statistikom. Navesti ćemo Bayesov teorem te definirati funkciju vjerodostojnosti, apriornu i aposteriornu distribuciju. Nakon definicija tih distribucija, bavit ćemo se odabirom apriorne distribucije te proučiti dvije metode odabira: konjugirane apriorne distribucije te neinformativne apriorne distribucije. Također ćemo se baviti Bayesovom procjenom parametara i spomenuti razne funkcije gubitka te normalne modele. Kratko ćemo spomenuti i što su Bayesovi pouzdani intervali te procjena pouzdanih intervala za parametar Θ iz bayesovske perspektive. U drugom poglavlju ćemo se upoznati s bayesovskim linearnim regresijskim modelima, prvo univarijatnim za koji ćemo navesti primjer, a zatim i multivarijatnim.

Regresijska analiza je središnji alat u primijenjenoj statistici čiji je cilj odgovoriti na pitanje kako pojedine varijable (nezavisne varijable) utječu na određeni ishod (zavisnu varijablu). Pretpostavimo da imamo standardni regresijski problem, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{K-1} X_{K-1} + \epsilon$, gdje je Y zavisna varijabla, X_k nezavisna varijabla, $k = 1, \ldots K - 1$, β_0 je slobodni član, β_k regresijski koeficijenti, $k = 1, \ldots K - 1$, te $\epsilon \sim N(0, \sigma^2)$ regresijski šum (greška). U klasičnom smislu regresijske parametre β i σ^2 ćemo procjeniti metodom najmanjih kvadrata. U bayesovskom pristupu ćemo regresijski parametarski vektor (β, σ^2) označiti kao (B, Σ^2) te ga promatrati kao slučajnu veličinu. Spomeniti ćemo slučaj nepravilne apriorne distribucije te informativne konjugirane apriorne distribucije i kakve aposteriorne distribucije dobijemo.

U dodatku A ćemo navesti uobičajne vjerojatnosne distribucije sa pripadajućim funkcijama gustoće te njihovim prvim i drugim momentima, a u dodatku B programski kod primjera linearnog regresijskog modela u R-u.

Poglavlje 1

Bayesovska statistika

1.1 Funkcija vjerodostojnosti i Bayesov teorem

Uvod

Pretpostavimo da imamo jednako distribuirane slučajne varijable $X_1, X_2, \ldots, X_n$, ali još ne znamo koja je to distribucija. Umjesto toga kreiramo skup distribucija koji se naziva *parametarska familija* i označena je sa $\mathcal{P}_0$. Na primjer, $\mathcal{P}_0$ može sadržavati sve normalne distribucije ili samo one sa varijancom 1, ili sve binomne distribucije, ili sve Poissonove distribucije itd. Svaki od ovih slučajeva ima svojstvo da skup distribucija može biti indeksiran kao konačno dimenzionalna realna veličina koja se obično naziva **parametar**. Na primjer, ako parametarska familija sadržava sve normalne distribucije, tada parametar možemo označiti kao $\Theta = (M, \Sigma)$, gdje M označava sredinu, a Σ je oznaka za standardnu devijaciju. Skup svih mogućih vrijednosti parametra se naziva *parametarski prostor* i često je označen da Ω. Kada je $\Theta = \theta$, distribucija opservacija je označena sa $\mathbb{P}_\theta$. Očekivane vrijednosti su označene sa $\mathbb{E}_\theta(\cdot)$.

Isto ćemo označavati promatrane podatke X. X može biti vektor opservacija koje su međusobno nezavisne i jednako distribuirane ili X može biti općenita slučajna veličina. Skup mogućih vrijednosti za X je *prostor uzorka* i obično se označava sa $\mathcal{X}$. Članovi parametarske familije $\mathcal{P}_0$ bit će vjerojatnosti na $\mathcal{X}$. Ako je X neprekidna ili diskretna veličina, tada funkcija gustoće postoji (u odnosu na Lebesgueovu ili brojeću mjeru). Gustoću ćemo označiti kao $f_{X|\Theta}(\cdot|\theta)$. Na primjer, ako je X jedna slučajna varijabla sa neprekidnom (ili diskretnom) distribucijom, tada

$$P_\theta(a < X \leq b) = \int_a^b f_{X|\Theta}(x|\theta)dx = \left(\sum_{x\in[a,b]} f_{X|\Theta}(x|\theta)\right).$$

Ako je $X = (X_1, X_2, \ldots, X_n)$, gdje su X_i nezavisne jednako distribuirane slučajne varijable sa funkcijom gustoće $f_{X_i|\Theta}(\cdot|\theta)$ kada je $\Theta = \theta$, tada

$$f_{X|\Theta}(x|\theta) = \prod_{i=1}^{n} f_{X_i|\Theta}(x_i|\theta), \tag{1.1}$$

gdje je $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. Nakon opažanja podataka $X_1 = x_1, \dots, X_n = x_n$, funkcija (1.1), kao funkcija od θ za fiksni x, se naziva *funkcija vjerodostojnosti*, a označava sa $L(\theta)$.

Funkcija vjerodostojnosti

Definicija 1.1.1. *Neka je X slučajna varijabla (ili slučajni vektor) i x njezina opažena vrijednost, te $f(x|\theta)$ njezina funkcija gustoće. Funkciju po parametru θ:*

$$L(\theta|x) := f(x|\theta), \theta \in \Phi \tag{1.2}$$

zovemo funkcijom vjerodostojnosti na osnovi opažene vrijednosti x statistike X.

Funkcija vjerodostojnosti nam zapravo pokazuje koliko je vjerodostojna svaka vrijednost parametra θ, ako znamo da je $X = x$.

Za ilustraciju koncepta funkcije vjerodostojnosti, navest ćemo dva primjera. Jedan je baziran na Poissonovoj distribuciji (diskretna distribucija), a drugi je baziran na normalnoj distribuciji (neprekidna distribucija).

Primjer 1.1.2. ***Funkcija vjerodostojnosti za Poissonovu distribuciju***
Poissonova distribucija je diskretna distribucija kojom se izražava vjerojatnost pojave događaja unutar fiksiranog vremenskog intervala ako je poznato prosječno vrijeme između pojave dva uzastopna događaja i ako su događaji medusobno nezavisni. Ona je određena parametrom $\theta \geq 0$, čije značenje je očekivani broj događaja tijekom zadanog vremenskog intervala. Vjerojatnosna distribucija Poissonove slučajne varijable X je opisana sljedećim izrazom

$$\mathbb{P}_\theta(X = k) = \frac{\theta^k}{k!} e^{-\theta}, k = 0, 1, 2, \dots$$

Pretpostavimo da imamo uzorak od n opažanja $(x_1, x_2, \dots, x_n)$ koji dolazi iz Poissonove distribucije. Tada funkcija vjerodostojnosti za parametar θ glasi:

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X = x_i) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\theta} \propto \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}.$$

Primjer 1.1.3. ***Funkcija vjerodostojnosti za normalnu distribuciju***
Normalna distribucija (Gaussova distribucija) je jedna od najvažnijih distribucija za vjerojatnosne i statističke modele zbog njezine relativne jednostavnosti i dostupnosti atraktivnih

teorijskih rezultata koji se na nju odnose. Dva parametra ju opisuju: parametar lokacije, μ, što je također i očekivanje te distribucije, te parametar disperzije, σ, kojeg nazivamo standardnom devijacijom. Funkcija gustoće vjerojatnosti za normalno distribuiranu slučajnu varijablu Y se definira sa:

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

te označava kao $Y \sim N(\mu,\sigma)$.
Pretpostavimo da imamo uzorak od n opažanja $(y_1, y_2, \ldots, y_n)$ za koji pretpostavljamo da dolazi iz normalne distribucije. Tada funkcija vjerodostojnosti za parametre μ i σ glasi:

$$L(\mu,\sigma|y_1,y_2,\ldots,y_n) = \sum_{i=1}^{n} f(y_i|\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}} \propto \sigma^{-n} e^{-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}}.$$

Bayesov teorem

Često želimo nešto zaključiti o parametru razdiobe ako imamo neke podatke iz te razdiobe. U klasičnom pristupu pretpostavljamo da je parametar nepoznata konstanta, a da podaci predstavljaju realizaciju slučajnog uzorka. Kod bayesovskog pristupa, uzimamo da je parametar koji još ne znamo slučajan, a podaci koje smo dobili konstantni. Razdiobu parametara možemo odabrati na temelju poznavanja načina kako se provodi slučajni pokus.

Diskretni iskaz Bayesovog teorema

Bayesov teorem je okosnica bayesovskog okvira. Formalno, to je osnovni rezultat iz teorije vjerojatnosti, povezujući bezuvjetnu distribuciju slučajne varijable s njezinom uvjetnom distribucijom. Filozofija bayesovskog pristupa temelji se na tumačenju da je vjerojatnost mjera stupnja vjerovanja koje imamo o neizvjesnom događaju. Bayesov teorem je pravilo koje se može koristiti za ažuriranje uvjerenja koje netko ima u svjetlu novih informacija (na primjer, opaženih podataka).
Prvo razmatramo diskretnu verziju Bayesovog teorema.

Teorem 1.1.4. ***(Bayesov teorem)***
Neka su A i B događaji na skupu Ω i neka je $\mathbb{P}(B) > 0$. Tada vrijedi:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{1.3}$$

Iz gornje jednakosti očito slijedi:

$$\mathbb{P}(A|B) \propto \mathbb{P}(B|A)\mathbb{P}(A). \tag{1.4}$$

Neka je parametar Θ diskretna slučajna varijabla. Ako pretpostavimo da E označava skup informacija o Θ, zanima nas vjerojatnost da slučajna varijabla, odnosno parametar Θ, poprimi vrijednost θ. Tada iz Bayesovog teorema slijedi:

$$P(\Theta = \theta | E) = P(\Theta = \theta) \frac{P(E|\Theta = \theta)}{P(E)}. \tag{1.5}$$

Iz ovog vidimo da vrijedi:

1. $P(\Theta = \theta | E)$ je zapravo funkcija gustoće parametra Θ nakon što smo uzeli u obzir dane informacije E o parametru, odnosno ono što ćemo kasnije zvati *aposteriorna funkcija gustoće*.

2. $P(\Theta = \theta)$ je funkcija gustoće od Θ prije nego što smo uzeli u obzir informacije E, odnosno takozvana *apriorna funkcija gustoće*.

3. $\frac{P(E|\Theta=\theta)}{P(E)}$ je član koji pokazuje koliki utjecaj ima informacija E na inferenciju o parametru Θ, dok je $P(E|\Theta = \theta)$ upravo *vjerodostojnost* vrijednosti θ.

1.2 Apriorna i aposteriorna distribucija

Promatramo vjerojatnosnu funkciju distribucije parametra θ. Uvodi se funkcija gustoće π *apriorne distribucije* na Φ, gdje je skup Φ skup svih vrijednosti koje može parametar θ poprimiti.

Od sada nadalje parametru θ pridružujemo vjerojatnosnu distribuciju, te on postaje slučajna varijabla, odnosno slučajni vektor u općenitom smislu. Označavat ćemo ga s Θ. Jedna njegova realizacija biti će $\theta \in \Theta$. Sad ćemo pomoću funkcije gustoće $f(x|\theta)$ slučajnog uzorka X i apriorne funkcije gustoće $\pi(\theta)$ parametra Θ definirati sljedeće distribucije:

1. *zajedničku distribuciju* slučajnog vektora (Θ, X) s gustoćom :

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

2. *marginalnu distribuciju* od X s gustoćom:

$$m(x) = \int \varphi(\theta, x) d\theta = \int f(x|\theta)\pi(\theta) d\theta; \tag{1.6}$$

3. *aposteriornu distribuciju* od Θ, kao uvjetnu distribuciju od Θ uz dano $X = x$ s gustoćom:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\xi)\pi(\xi)d\xi} = \frac{f(x|\theta)\pi(\theta)}{m(x)}. \tag{1.7}$$

Iz (1.7) možemo zaključiti da ako nam je poznata gustoća $f(x|\theta)$ i funkcija gustoće apriorne distribucije parametra Θ, $\pi(\theta)$, možemo lako doći i do funkcije gustoće aposteriorne distribucije parametara Θ uz uvjet $X = x$.

Dakle, aposteriorna distribucija sadrži informacije i od slučajnog uzorka i od apriorne distribucije. Ona izražava zaključak statističara o vjerojatnosnoj distribuciji parametra θ nakon što je dobivena realizacija slučajnog uzorka.

Budući da smo definirali sve potrebne komponente, sad možemo definirati što je bayesovski statistički model, ali prvo ćemo se prisjetiti definicije statističke strukture te statistike.

Definicija 1.2.1. *Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor i neka je $\mathcal{P}$ familija vjerojatnosti na $(\Omega, \mathcal{F})$. Uređena trojka $(\Omega, \mathcal{F}, \mathcal{P})$ se naziva statistička struktura.*

Familija $\mathcal{P}$ je često parametrizirana i piše se u obliku:

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

Definicija 1.2.2. *Statistika na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je svaka slučajna varijabla (ili slučajni vektor) $T : \Omega \to \mathbb{R}^k$ takva da postoji $n \in N$ i n-dimenzionalni slučajni vektor $(X_1, X_2, \ldots, X_n)$ na $(\Omega, \mathcal{F}, \mathcal{P})$, te izmjerivo preslikavanje $t : \mathbb{R}^n \to \mathbb{R}^k$ takvo da je $T = t(X_1, X_2, \ldots, X_n)$.*

Definicija 1.2.3. *Bayesovski statistički model slučajne varijable (ili slučajnog vektora) $X : \Omega \to \mathbb{R}^p$ je parametarski statistički model $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Phi\}$, iz definicije 1.2.1, s tim da je parametru Θ pridružena njegova apriorna distribucija s gustoćom $\pi(\theta)$, $\theta \in \Phi$.*

1.3 Nepravilna apriorna distribucija

Kada se nepoznati parametar promatra kao slučajna varijabla Θ sa poznatom funkcijom distribucije π, pomoću Bayesovog teorema lako je doći do aposteriorne distribucije i zaključka o parametru Θ. Međutim, u većini slučajeva, umjesto vjerojatnosnom mjerom, apriorna distribucija u praksi je određena sa mjerom za koju vrijedi:

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

i koja je σ-konačna na prostoru parametara Φ.

Definicija 1.3.1. *Kada za apriornu funkciju gustoće π vrijedi*

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

apriorna distribucija naziva se ***nepravilna ili generalizirana****.*

Nepravilna apriorna distribucija može dovesti do nepravilne aposteriorne distribucije. Kada je aposteriorna distribucija nepravilna, zaključci neće biti valjani, nije integrabilna i bayesovske metode se ne mogu koristiti.

Iako se nepravilne apriorne distribucije mogu koristiti dokle god je aposteriorna distribucija konačna, dobra je praksa izbjeći ih.

Uobičajni dogovor je uzeti aposteriornu distribuciju $\pi(\theta|x)$ povezanu sa nepravilnom apriornom distribucijom $\pi(\theta)$ danom sa sljedećom Bayesovom formulom:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Phi} f(x|\xi)\pi(\xi) d\xi},$$

kada je marginalna distribucija $\int_{\Phi} f(x|\xi)\pi(\xi) d\xi$ dobro definirana.

1.4 Odabir apriorne funkcije distribucije

Cijela bayesovska statistika se zasniva na apriornoj funkciji distribucije i njenom odabiru. Kako apriornu funkciju distribucije biramo upravo na temelju danih informacija o parametru Θ, očito je da odabir dovoljno precizne apriorne funkcije distribucije neće biti često trivijalan.

Cilj odabira apriorne distribucije je da ona sadrži sve ili većinu informacija koje su nam dane za parametar Θ. Zbog toga je potrebno promatrati aproksimacije apriorne funkcije distribucije na temelju danih informacija. Upravo na odabiru te aproksimacije ovisi značajnost daljnje analize. Parametre apriorne funkcije distribucije nazivat ćemo *hiperparametrima*.

Obradit ćemo dvije metode odabira apriornih funkcija distribucije:

- metodu konjugiranih apriornih distribucija i
- metodu neinformativnih apriornih distribucija.

Konjugirane apriorne distribucije

Promatramo klasični parametarski pristup kod kojeg je subjektivnost ograničena, te u čijoj je pozadini empirijska bayesovska tehnika.

Definicija 1.4.1. *Familija $\mathcal{F}$ vjerojatnosnih distribucija na Φ je konjugirana za statističku strukturu $\mathcal{P}$ ako za svaki $\pi \in \mathcal{F}$ također vrijedi $\pi(\theta|x) \in \mathcal{F}$, za svako x. Apriornu i aposteriornu distribuciju tada nazivamo konjugiranim distribucijama, a apriornu funkciju gustoće konjugiranim apriorom za funkciju vjerodostojnosti danu pomoću modela $\mathcal{P}$.*

Trivijalni primjer konjugirane familije na Φ je familija $\mathcal{F}_0$ koja sadrži sve distribucije na Φ, koji je naravno beskoristan za izbor apriorne distribucije. Cilj metode je pronaći najmanju moguću *parametriziranu* familiju $\mathcal{F}$ iz definicije 1.4.1. Primjetimo da najmanju familiju $\mathcal{F}$ nećemo tražiti kao presjek svih konjugiranih familija jer je taj presjek često prazan skup.

Neka je dan slučajni uzorak $X \sim f(x|\theta)$, gdje je θ realizacija parametra Θ. Tada funkcija $f(x|\theta)$ funkciju gustoće apriorne distribucije $\pi(\theta)$ modificira u funkciju gustoće aposteriorne distribucije $\pi(\theta|x)$. No, ta modifikacija ne bi smjela mijenjati cijelu strukturu od $\pi(\theta)$ već samo njezine parametre. Radikalnija promjena od π je stoga neprihvatljiva. Zbog toga se promatraju parametrizirane konjugirane familije. Takve konjugirane apriorne distribucije se često nazvaju i *objektivnim* zato što ovise o distribuciji slučajnog uzorka, odnosno o funkciji gustoće $f(x|\theta)$.

Konjugirane distribucije eksponencijalnih familija

Konjugirane apriorne distribucije se obično povezuju sa određenim tipom uzoračkih distribucija. Takve distribucije čine *eksponencijalne familije*.

Definicija 1.4.2. *Neka je X slučajni uzorak na Ω iz modela $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Phi)\}$. Model $\mathcal{P}$ je k-parametarska eksponencijalna familija ako je gustoća $f(\cdot|\theta)$ dana u obliku:*

$$f(x|\theta) = C(\theta)h(x)e^{(R(\theta)|T(x))}.$$

Pri tome su $C : \Phi \to \mathbb{R}_+$ i $h : \Omega \to \mathbb{R}_+$, te $R : \Phi \to \mathbb{R}^k$ i $T : \Omega \to \mathbb{R}^k$ izmjerive funkcije, te je sa $(\cdot|\cdot)$ od sada nadalje označen skalarni produkt vektora.
Ako je $\Phi \subset \mathbb{R}^k$ i $\Omega \subset \mathbb{R}^k$, te vrijedi:

$$f(x|\theta) = C(\theta)h(x)e^{(\theta|x)},$$

kažemo da je familija $\mathcal{P}$ prirodna.

Definicija 1.4.3. *Neka je $f(x|\theta) = C(\theta)h(x)e^{(\theta|x)}$ funkcija gustoće prirodne eksponencijalne familije . Prirodni prostor parametara definira se kao:*

$$N = \left\{\theta; \int e^{(\theta|x)}h(x)d\mu(x) < +\infty\right\}, \tag{1.8}$$

pri čemu je μ ona σ-konačna mjera u odnosu na koju je $f(x|\theta), \theta \in \Phi$, funkcija gustoće. Kažemo da je eksponencijalna familija regularna ako je N otvoren skup i ako je $dim(N) = dim(K)$, gdje je K zatvarač konveksne ljuske nosača[1] *od μ.*

Propozicija 1.4.4. [2] *Neka je*

$$f(x|\theta) = h(x)e^{(\theta|x)-\phi(\theta)} \tag{1.9}$$

funkcija gustoće eksponencijalne familije. Tada je gustoća njene konjugirane familije dana sa:

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda)e^{(\theta|\mu)-\lambda\cdot\phi(\theta)}. \tag{1.10}$$

Pri tom je $\lambda > 0$ i $\frac{\mu}{\lambda} \in N^0$, gdje je N^0 interior prirodnog prostora parametara iz definicije 1.4.3, a $K(\mu, \lambda)$ konstanta. Aposteriorna funkcija gustoće time je dana sa $\pi(\theta \mid \mu + x, \lambda + 1)$, a μ i λ su hiperparametri.

Prethodna propozicija daje nam važan rezultat koji nam govori kako možemo za eksponencijalnu familiju pronaći njenu konjugiranu familiju. Sljedeća tablica (Tablica 1.1) pokazuje konjugirane distribucije za najpoznatije distribucije eksponencijalnih familija.

Neinformativne apriorne distribucije

U prethodnom poglavlju smo pokazali da su konjugirane distribucije korisne za zadavanje prave apriorne distribucije. No, ukoliko nemamo dovoljno informacija za neki model, ne možemo formirati apriornu funkciju gustoće. Tada je hiperparametre konjugirane distribucije nemoguće odrediti. U tom slučaju umjesto klasičnog pristupa (npr. metoda maksimalne vjerodostojnosti), i dalje koristimo bayesovski pristup. Koristimo takozvane *neinformativne apriorne distribucije*. Dva problema se javljaju kod ove metode. Prvo, rezultirajuća distribucija nije vjerojatnosna kada je parametarski prostor nekompaktan. Drugi problem koji se javlja je *invarijantnost na reparametrizaciju*. Ako umjesto realizacije θ parametra Θ odlučimo uzeti transformaciju $\eta = g(\theta)$, želimo biti sigurni da to neće promjeniti apriornu distribuciju.

[1]Nosač mjere μ definira se kao: $\sup\mu := \left\{x \in \mathbb{R}^k;\ \textit{za svaku otvorenu okolinu } U_x \textit{ od } x \textit{ je } \mu(U_x) > 0\right\}$.

[2]Navedena je u knjizi C.P. Robert, *The Bayesian Choice, A Decision-Theoretic Motivation*, Springer, 1994.

$f(x\|\theta)$	$\pi(\theta)$	$\pi(\theta\|x)$
Normalna	Normalna	
$N(\theta, \sigma^2)$	$N(\mu, \tau^2)$	$N\left(\frac{\sigma^2\mu+\tau^2 x}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$ $\varrho^{-1} = \sigma^2 + \tau^2.$
Poissonova	Gama	
$\mathcal{P}(\theta)$	$\Gamma(\alpha,\beta)$	$\Gamma(\alpha + x, \beta + 1)$
Gama	Gama	
$\Gamma(\nu, \theta)$	$\Gamma(\alpha,\beta)$	$\Gamma(\alpha + \nu, \beta + x)$
Binomna	Beta	
$\mathcal{B}(n, \theta)$	$Beta(\alpha,\beta)$	$Beta(\alpha + x, \beta + n - x)$
Negativna binomna	Beta	
$Neg\mathcal{B}(m, \Theta)$	$Beta(\alpha,\beta)$	$Beta(\alpha + m, \beta + x)$
Multinomna	Dirichlet	
$\mathcal{M}_k(\theta_1, ..., \theta_k)$	$\mathcal{D}(\alpha_1, ..., \alpha_k)$	$\mathcal{D}(\alpha_1 + x_1, ..., \alpha_k + x_k)$

Tablica 1.1: Konjugirane apriorne i aposteriorne gustoće određenih distribucija iz eksponencijalne familije

Definicija 1.4.5. *Neka je $\pi(\theta)$ apriorna funkcija gustoće odabrana nekom metodom. Kažemo da je π neinformativna, ako je invarijantna na reparametrizaciju, odnosno ako izbor realizacije parametara ne utječe na izbor apriorne gustoće. Ako parametar θ reparametriziramo u $\eta = g(\theta)$, tada mora vrijediti:*

$$\pi(\eta) = \pi(g^{-1}(\eta)) \left[\frac{\partial(\theta)}{\partial(\eta)}\right].$$

Laplace je bio prvi koji je koristio neinformativne metode i odredio neinformativnu apriornu distribuciju. Navesti ćemo njegova dva primjera.

Primjer 1.4.6. *Neka kutija sadrži n bijelih i crnih kuglica. Ako je prva izvučena kuglica iz kutije bijela, zanima nas kolika je vjerojatnost da je postotak bijelih kuglica u kutiji P jednak p_0? Pri rješavanju ovog problema, Laplace je pretpostavio da su svi mogući postoci bijelih kuglica jednako vjerojatni, odnosno da je slučajna varijabla P uniformno distribuirana na $\{0, \frac{1}{n}, \ldots, 1\}$. Aposteriorna distribucija od P se može izvesti pomoći Bayesovog teorema i:*

$$\mathbb{P}(P = p_0 \mid uz\ dani\ podatak) = \frac{2p_0}{n+1}.$$

Primjer 1.4.7. *Uzevši u obzir postotak rođene muške djece u Parizu, Laplace je želio testirati da je vjerojatnost rođenja muške djece, u oznaci x, veća od $\frac{1}{2}$. Za* 251 527 *rođene* *muške djece i* 241 945 *rođene ženske djece, pretpostavljajući da x ima uniformnu apriornu distribuciju na* $[0, 1]$*, Laplace je dobio:*

$$\mathbb{P}(x \leq 1/2 \mid (251\,527; 241\,945)) = 1.15 \cdot 10^{-42}.$$

Tada je zaključio da je vjerojatnost x gotovo sigurno veća od 50%*. Pod istim pretpostavkama, usporedio je rođenja muške djece u Londonu i Parizu i došao da zaključka da je veća vjerojatnost rođenja muške djece u Engleskoj.*

Jeffrey je predložio pristup koji ne ovisi o invarijantim strukturama, tzv. *Jeffreyeve neinformativne apriorne distribucije* koje se temelje na *Fisherovoj informaciji*.

Definicija 1.4.8. *Neka je X slučajni uzorak iz modela* $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Phi\}$*. Definiramo Fisherovu informaciju s obzirom na uzorak X kao:*

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial log f(X|\theta)}{\partial \theta} \right)^2 \right]. \tag{1.11}$$

Fisherova informacija mjeri količinu informacija koju uzorak X nosi o parametru θ.

Ako je model $\mathcal{P}$ regularan i funkcija f dva puta derivabilna, za Fisherovu informaciju vrijedi:

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta \left[\left(\frac{\partial log f(X|\theta)}{\partial \theta} \right)^2 \right] \\ &= -\mathbb{E}_\theta \left[\left(\frac{\partial^2 log f(X|\theta)}{\partial^2 \theta} \right) \right]. \end{aligned}$$

Definicija 1.4.9. *Jeffreyjeva neinformativna apriorna distribucija ima funkciju gustoće danu sa:*

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

1.5 Bayesova procjena parametara

Funkcija gubitka i korisnosti

Trebamo procjeniti nepoznati parametar i naći najboljeg procjenitelja za njega. Kako možemo imati više procjenitelja moramo odrediti kriterij na osnovu kojeg ćemo ih uspoređivati. Taj se kriterij obično naziva **gubitak** (*loss*).
Neka je $\mathcal{D}$ skup mogućih odluka, a zovemo ga *prostor odluka*, najčešće je $\mathcal{D} = \Phi$.

Definicija 1.5.1. ***Funkcija gubitka*** *je svaka funkcija L iz skupa* $\Phi \times \mathcal{D}$ *u skup* $[0, +\infty\rangle$.

Funkcija gubitka ocjenjuje grešku $L(\theta, d)$ kod odluke d kada parametar Θ poprima vrijednost θ. Kada je $\mathcal{D} = \Phi$, d se interpretira kao ona vrijednost od Θ za koju smo se odlučili nakon procjene.

Odabir funkcije gubitka je u principu subjektivan, ovisi o postavljenom problemu i onom tko ga proučava. U praksi je često nespretno odrediti funkciju gubitka, pa se često koriste klasični gubitci zbog jednostavnosti i dobre matematičke primjenjivosti i potkrijepljenosti dokazima.

U bayesovskom pristupu statističko zaključivanje bi trebalo započeti strogim određivanjem triju faktora:

- familije distribucija slučajnog uzorka $f(x|\theta)$
- distribucije parametara $\pi(\theta)$ i
- gubitka $L(\theta, d)$.

Vratimo se sada na statistički model. Vidimo da se sastoji od prostora elementarnih događaja Ω, parametarskog skupa Φ te skupa odluka $\mathcal{D}$. Metoda se sastoji od izbora odluke $d \in \mathcal{D}$ za parametar Θ koji poprima vrijesnosti $\theta \in \Phi$, na temelju opservacije $x \in \mathbb{R}$ slučajne varijable X, gdje su X i Θ povezani pomoću funkcije distribucije $f(x|\theta)$. U većini slučajeva d se bira tako da se najbolje procjeni neka funkcija od θ, $h(\theta)$. Pretpostavlja se da svaka odluka $d \in \mathcal{D}$ može biti procjenjena pomoću nagrade r s korisnošću $U(r)$. Od sada na dalje ćemo funkciju korisnosti označavati sa $U(\theta, d)$, s naglaskom da ona ovisi samo o dva faktora, θ i d, uz uvjet $\Theta = \theta$.

Nakon konstrukcije funkcije korisnosti, možemo definirati funkciju gubitka na sljedeći način:

$$L(\theta, d) = -U(\theta, d).$$

Općenito, funkcija gubitka je nenegativna, pa to povlači da je $U(\theta, d) \leq 0$. Očito, općenito je nemoguće uniformno minimizirati funkciju gubitka $L(\theta, d)$ kada je θ nepoznat. Da bi se izveo efikasan kriterij uspoređivanja promatra se *funkcija rizika* kao srednji gubitak:

$$R(\theta, d) = \mathbb{E}_\theta[L(\theta, \delta(X))] = \int_{\mathbb{R}} L(\theta, \delta(x)) f(x|\theta) dx, \tag{1.12}$$

gdje je $d = \delta(x)$ procjena od θ, ovisna o realizaciji x slučajne varijable X. Takav pristup se naziva frekvencionistički.

Bayesovski pristup umjesto srednjeg gubitka danog sa (1.12) koristi *aposteriorni očekivani gubitak*:

$$\varrho(\pi, d|x) := \mathbb{E}^\pi[L(\Theta, d)|X = x] = \int_\Phi L(\theta, d)\pi(\theta|x) d\theta$$

koji promatra očekivanje gubitka na temelju aposteriorne distribucije parametra Θ uz danu opservaciju x slučajne varijable X. Uz dani $X = x$, srednji gubitak koji je rezultat odluke d je zapravo $\varrho(\pi, d|x)$. Aposteriorni očekivani gubitak je funkcija u ovisnosti od x što nam ne stvara probleme za razliku od funkcije definirane u (1.12) koja je u ovisnosti o θ koji je nepoznat.

Za danu apriornu distribuciju $\pi(\theta)$, moguće je definirati *integrirani rizik*:

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

Od posebnog je interesa što ova funkcija pridružuje realni broj svakom procjenitelju a ne funkciju od θ jer u postupku traženja optimalnog procjenitelja omogućuje njihovu direktnu usporedbu.

Slijedi teorem koji govori da su aposteriorni očekivani gubitak i integrirani rizik ekvivalentni, odnosno da dovode do iste odluke, tj. do istog optimalnog procjenitelja.

Teorem 1.5.2. *Ako procjenitelj $\delta(x)$ za svaki $x \in \mathbb{R}$ minimizira aposteriorni očekivani gubitak, $\varrho(\pi, \delta|x)$, tada minimizira i inegrirani rizik $r(\pi, \delta)$ i vrijedi:*

$$r(\pi, \delta) = \int_{\mathbb{R}} \varrho(\pi, \delta(x)|x)m(x)dx, \tag{1.13}$$

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

Dokaz. Kako je $L(\theta, \delta) \geq 0$, jednakost (1.13) slijedi direktno iz Fubinijevog teorema i definicija apriorne i aposteriorne distribucije. Vrijedi:

$$\begin{aligned} r(\pi, \theta) &= \int_{\Phi}\int_{\mathbb{R}} L(\theta, \delta(x))f(x|\theta)dx\,\pi(\theta)\,d\theta \\ &= \int_{\mathbb{R}}\int_{\Phi} L(\theta, \delta(x))f(x|\theta)\pi(\theta)\,d\theta\,dx \\ &= \int_{\mathbb{R}}\left(\int_{\Phi} L(\theta, \delta(x))\pi(\theta|x)\,d\theta\right)m(x)\,dx \\ &= \int_{\mathbb{R}} \varrho(\pi, \delta(x)|x)m(x)\,dx. \end{aligned} \tag{1.14}$$

Iz (1.13) direktno slijedi ako $\delta(x)$ minimizira $\varrho(\pi, \delta(x)|x)$ za svaki $x \in \mathbb{R}$, tada zbog nenegativnosti marginalne funkcije distribucije i linearnosti integrala, minimizira i integralni rizik $r(\pi, \theta)$. □

Ovaj rezultat nas dovodi do definicije Bayesovog procjenitelja.

Definicija 1.5.3. *Bayesov procjenitelj δ^π definiran je sa:*

$$\delta^\pi(x) = \operatorname*{Argmin}_d \varrho(\pi, d|x), x \in \mathbb{R}.$$

Vrijednost $r(\pi) = r(\pi, \delta^\pi)$ se naziva Bayesov rizik.

Sada promotrimo neke klasične funkcije gubitka.

Neki klasični gubitci

Kvadratni gubitak

Definicija 1.5.4. *Kvadratna funkcija gubitka definirana je pomoću sljedeće formule:*

$$L(\theta, d) = (\theta - d)^2. \tag{1.15}$$

Propozicija 1.5.5. *Bayesov procjenitelj δ^π za kvadratnu funkciju gubitka je aposteriorno očekivanje od Θ dano sa:*

$$\delta^\pi(x) = \mathbb{E}^\pi[\Theta|X = x] = \frac{\int_\Phi \theta f(x|\theta)\pi(\theta)\, d\theta}{\int_\Phi f(x|\theta)\pi(\theta)\, d\theta}, \ x \in \mathbb{R}.$$

Dokaz. Budući da vrijedi

$$\mathbb{E}^\pi[(\Theta - \delta)^2|X = x] = \mathbb{E}^\pi[\Theta^2|X = x] - 2\delta\mathbb{E}^\pi[\Theta|X = x] + \delta^2,$$

aposteriorni gubitak zapravo postiže svoj minimum u $\delta^\pi(x) = \mathbb{E}^\pi[\Theta|X = x]$. □

Apsolutni gubitak

Funkcija apsolutnog gubitka definirana je sa:

$$L(\theta, d) = |\theta - d|,$$

ili općenito:

$$L_{k_1,k_2}(\theta, d) = \begin{cases} k_2(\theta - d), & \text{ako je } \theta > d \\ k_1(d - \theta), & \text{inače.} \end{cases}$$

Ovakve funkcije rastu sporije od kvadratne funkcije gubitka. Kod njih je konveksnost zadržana, a ne kažnjavaju prejako velike, ali malo vjerojatne greške. Možemo promatrati i kombinaciju apsolutnog i kvadratnog gubitka:

$$\tilde{L}(\theta, d) = \begin{cases} (d - \theta)^2, & \text{ako je } |d - \theta| < k \\ 2k|d - \theta| - k^2, & \text{inače.} \end{cases}$$

Takva funkcija usporava rast kvadratnog gubitka za velike greške čime je uklonjena velika mana kvadratnog gubitka.

Gubitak "0-1"

Ovaj gubitak se najčešće koristi u klasičnom pristupu testiranja hipoteza. Kazna vezana uz procijenu δ je 0 ako je odgovor točan, inače je 1, odnosno:

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

Stvarni gubitci

U nekim slučajevima ne postoji dovoljno informacija pa je teško odrediti funkciju gubitka. To se događa kad nam je bitna distribucija $f(x|\theta)$. U takvim slučajevima prirodno je koristiti gubitke koji direktno uspoređuju $f(\cdot|\theta)$ i $f(\cdot|\delta)$ vezane uz parametar θ i procjenu δ.

Funkcije gubitka,

$$L(\theta, \delta) = d(f(\cdot|\theta), f(\cdot|\delta)),$$

su neparametarske.

Dvije su uobičajne mjere razlika:

- *Entropijska razlika*

$$L_e(\theta, \delta) = \mathbb{E}_\theta\left[\log\left(\frac{f(X|\theta)}{f(X|\delta)}\right)\right]$$

- *Hellingerova razlika*

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

Bayesova procjena parametara

Kada je zadana apriorna funkcija gustoće, lako je moguće odrediti aposteriornu funkciju gustoće pomoću funkcije gustoće uzoraka $f(x|\theta)$. Tada je aposteriorna distribucija detaljni sažetak informacija o distribuciji parametara Θ na temelju informacija dostupnim o parametru Θ iz aposteriorne distribucije i informacija koje donosi sam uzorak X, odnosno njegova distribucija. Bayesovska verzija metode vjerodostojnosti povlači da bi inferencija za parametar Θ trebala ovisiti o aposteriornoj distribuciji i opaženoj vrijednosti X.

Kao mogući izbor za procjenitelja za parametar Θ na temelju aposteriorne funkcije gustoće koristimo tzv. *bayesovski procjenitelj maksimalne vjerodostojnosti* (oznaka: MBLE), koji maksimizira $L(\theta|x)\pi(\theta)$. Ovaj procjenitelj se također može nazvati *penalizirani procjenitelj maksimalne vjerodostojnosti*.

Primjer 1.5.6. *Neka je $X \sim \mathcal{B}(n,p)$. Jeffrejeva apriorna distribucija u ovom slučaju bi bila beta distribucija s parametrima $(\frac{1}{2},\frac{1}{2})$, odnosno:*

$$\pi_j(p) = \frac{1}{Beta(\frac{1}{2},\frac{1}{2})} p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}.$$

Sljedeće dvije apriorne distribucije su predložili Laplace i Haldane 1931.godine:

$$\pi_1(p) = 1 \qquad i \qquad \pi_2(p) = p^{-1}(1-p)^{-1}.$$

Funkcija vjerodostojnosti je:

$$L(p,x) = \binom{n}{k} p^x(1-p)^{n-x}.$$

Pripadajući bayesovski procjenitelji maksimalne vjerodostojnosti, za $n > 2$, su tada jednaki:

$$\begin{aligned}
\delta_j(x) &= \operatorname*{Argmax}_p \left(\binom{n}{k} p^x(1-p)^{n-x} \cdot \frac{1}{B(\frac{1}{2},\frac{1}{2})} p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} \right) \\
&= \max\left(\frac{x-1/2}{n-1}, 0 \right), \\
\delta_1(x) &= \operatorname*{Argmax}_p \left(\binom{n}{k} p^x(1-p)^{n-x} \right) = \frac{x}{n}, \\
\delta_2(x) &= \operatorname*{Argmax}_p \left(\binom{n}{k} p^x(1-p)^{n-x} \cdot p^{-1}(1-p)^{-1} \right) \\
&= \max\left(\frac{x-1}{n-2}, 0 \right).
\end{aligned}$$

Kada je $n = 1$, δ_j i δ_2 su jednaki δ_1 (jer je X Bernoullijeva slučajna varijabla sa vjerojatnosti uspjeha p). Za $n = 2$ i $x = 1$, procjenitelj δ_2 je također jednak δ_1. Primjetimo, za veliki n, sva tri procjenitelja su jednaka.

Budući da je aposteriorna distribucija poznata, odnosno njena funkcija gustoće, moguće je za procjenitelja $\delta^\pi(x)$ parametra $h(\theta)$ odrediti i preciznost procjene pomoću, npr. *aposteriorne kvadratne greške*:

$$\mathbb{E}^\pi[(\delta^\pi(x) - h(\Theta))^2 | X = x],$$

što je jednako $Var^\pi(h(\Theta|X = x))$ kada je $\delta^\pi(x) = \mathbb{E}^\pi[h(\Theta)|X = x]$.
Prisjetimo se da smo do sada, uz danu funkciju gubitka $L(\theta,\delta)$, Bayesov procjenitelj $\delta^\pi(x)$, definirali kao:

$$\delta^\pi(x) = \operatorname*{Argmin}_\delta \mathbb{E}^\pi[L(\Theta,\delta)|X = x].$$

Premda postoje mnogi procjenitelji parametra Θ, pokazalo se da je najprecizniji Bayesov procjenitelj jer se temelji na funkciji greške. Najpoznatiji su Bayesovi procjenitelji koji su povezani sa kvadratnom funkcijom gubitka i najčešće su usko povezani sa distribucijom preko očekivanja, varijance, medijana, kvantila itd.

Navodimo tablicu (Tablica 1.2) Bayesovih procjenitelja za parametar Θ uz kvadratnu grešku, te pripadajuće konjugirane apriorne distribucije iz eksponencijalne familije.

Distribucija	Apriori (konjugirana) distribucija	Aposteriori očekivanje
Normalna $N(\theta, \sigma^2)$	Normalna $N(\mu, \tau^2)$	$\frac{\mu\sigma^2+\tau^2 x}{\sigma^2+\tau^2}$
Poissonova $\mathcal{P}(\theta)$	Gamma $\Gamma(\alpha,\beta)$	$\frac{\alpha+x}{\beta+1}$
Gamma $\Gamma(\nu, \theta)$	Gamma $\Gamma(\alpha,\beta)$	$\frac{\alpha+\nu}{\beta+x}$
Binomna $\mathcal{B}(n, \theta)$	Beta $Beta(\alpha,\beta)$	$\frac{\alpha+x}{\alpha+\beta+n}$
Negativna binomna $Neg\mathcal{B}(n, \theta)$	Beta $Beta(\alpha,\beta)$	$\frac{\alpha+n}{\alpha+\beta+x+n}$
Multinomna $\mathcal{M}_k(n; \theta_1, ..., \theta_k)$	Dirichletova $\mathcal{D}(\alpha_1, ..., \alpha_k)$	$\frac{\alpha_i+x_i}{(\sum_j \alpha_j)+n}$

Tablica 1.2: Bayesovi procjenitelji za parametar Θ uz kvadratni gubitak i konjugirani apriori distribucija eksponencijalne familije

Normalni modeli

Kao što je do sad već spomenuto, normalna distribucija je jedna od najvažnijih u statistici. Promotrimo opservaciju X iz višedimenzionalne distribucije $N_p(\theta, \Sigma^2)$, gdje je $\theta \in \mathbb{R}^p$ realizacija parametra Θ, a $\Sigma^2 \in M_{pxp}$ kovarijacijska matrica. Pretpostavimo da je kovarijacijska matrica poznata te da je potrebno odrediti očekivanje. Konjugirana apriorna distribucija je također normalna, $N_p(\mu, A)$, a aposteriorna distribucija od Θ uz $X = x$, je dana sa:

$$N_p(x - \Sigma^2(\Sigma^2 + A)^{-1}(x - \mu), (A^{-1} + \Sigma^{-2})^{-1}).$$

Tada je Bayesov procjenitelj parametra Θ uz kvadratni gubitak jednak:

$$\begin{aligned} \delta^\pi(x) &= \mathbb{E}^\pi[\Theta|X = x] \\ &= x - \Sigma^2(\Sigma^2 + A)^{-1}(x - \mu) \\ &= (A^{-1} + \Sigma^{-2})^{-1}(\Sigma^{-2}x + A^{-1}\mu). \end{aligned} \tag{1.16}$$

Primjetimo da $\delta^{\pi}(x)$ može biti zapisan kao konveksna kombinacija opservacija X i apriorne sredine μ. Što je veća točnost apriorne distribucije parametara Θ, to je Bayesov procjenitelj sve bliži μ.
Ako imamo n opservacija $X_1, X_2, \ldots, X_n$ iz gornjeg modela, koristimo dovoljnu statistiku:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N_p\left(\theta, \frac{1}{n}\Sigma^2\right)$$

i nastavljamo prethodnu analizu.

Procjena varijance

U većini slučajeva, varijanca modela je djelomično ili skroz nepoznata. Tada je nužno uzeti u obzir apriornu distribuciju za parametre (Θ, Σ^2). Ako su komponente $X_1, X_2, \ldots, X_n$ nezavisne normalno distribuirane slučajne varijable, kovarijacijska matrica Σ^2 će biti dijagonalna matrica. Ako su uz to $X_1, X_2 \ldots, X_n$ također jednako distribuirane s varijancom σ^2, kovarijacijska matrica će biti poznata do na multiplikativnu konstantu σ^2, pa je moguće problem svesti na jednodimenzionalni slučaj. Stoga ćemo promatrati jednodimenzionalne normalne modele.
U slučaju da nam je poznato očekivanje, potrebno je samo procjeniti σ^2. U slučaju da je očekivanje nepoznato, definiramo sljedeće statistike:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

i

$$S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

gdje su $X_1, X_2, \ldots, X_n$ opservacije iz modela $N_p(\theta, \sigma^2)$. Funkciju vjerodostojnosti možemo pokazati kao:

$$L(\theta, \sigma|\bar{x}, s^2) \propto \sigma^{-n} exp\left\{-\frac{1}{2\sigma^2}\left(s^2 + n(\bar{x}-\theta)^2\right)\right\},$$

gdje su s^2 realizacije od S^2, a $\bar{x}$ realizacije od $\bar{X}$, pa Bayesov procjenitelj ovisi samo o s^2 i $\bar{x}$.
Jeffrejeva distribucija za ovaj model je dana sa $\pi_j(\theta, \sigma^2) = \frac{1}{\sigma^2}$, no bolje je promatrati neinformativnu apriornu gustoću $\pi(\theta, \sigma^2) = \frac{1}{\sigma}$. U tom slučaju , vrijedi:

$$L(\theta, \sigma|\bar{x}, s^2)\pi(\theta, \sigma) \propto \sigma^{-n-1} exp\left\{-\frac{1}{2\sigma^2}\left(s^2 + n(\bar{x}-\theta)^2\right)\right\}, \tag{1.17}$$

Iz toga slijedi sljedeća propozicija. No, prvo napomenimo da budući da samo sveli problem na jednodimenzionalni slučaj, od sada nadalje Σ^2 neće označavati kovarijacijsku matricu već će biti slučajna varijabla sa realizacijom σ^2, gdje je σ^2 varijanca.

Propozicija 1.5.7. *Neka su $X_1, X_2, \ldots, X_n$ nezavisno jednako distribuirani slučajni vektori iz $N(\theta, \sigma^2)$. Tada je aposteriorna distribucija parametara (Θ, Σ^2) povezana sa apriornom funkcijom gustoće $\pi(\theta, \sigma^2) = \frac{1}{\sigma}$, jednaka:*

$$\begin{aligned} \Theta|\sigma^2, \bar{X} = \bar{x}, S^2 = s^2 &\sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) \\ \Sigma^2|\bar{X} = \bar{x}, S^2 = s^2 &\sim Inv\Gamma\left(\frac{n-1}{2}, \frac{s^2}{2}\right). \end{aligned} \tag{1.18}$$

Ovdje $Inv\Gamma(\alpha, \beta)$ označava neprekidnu funkciju distribucije sa gustoćom:

$$\pi(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha) y^{\alpha+1}} e^{\frac{-\beta}{x}} \mathbb{1}_{(0,+\infty)}(y),$$

a naziva se inverzna gama distribucija.

Jednadžba dana sa (1.18) zbilja definira aposteriornu distribuciju parametara (Θ, Σ^2), budući da su dane marginalna distribucija od Σ^2, te distribucija od Θ uz uvjet $\Sigma^2 = \sigma^2$. Dokaz propozicije je direktna posljedica (1.17) i

$$\pi(\theta, \sigma^2|\bar{x}, s^2) \propto \sigma^{-1} exp\left\{\frac{-n(\bar{x}-\theta)^2}{2\sigma^2}\right\} \sigma^{-n} exp\left\{\frac{-s^2}{2\sigma^2}\right\} \sigma^{-1}.$$

Prema tome je marginalna aposteriorna distribucija parametra Σ^2 istog tipa kao i kada je Θ poznat. Suprotno tome, marginalna aposteriorna distribucija od Θ je modificirana, budući da iz (1.18) slijedi:

$$\begin{aligned} &\pi(\theta|\bar{x}, s^2) \propto \{s^2 + n(\bar{x}-\theta)^2\}^{-n/2}, \\ &\Theta|\bar{x}, s^2 \sim t\left(n-1, \bar{x}, \frac{s^2}{n(n-1)}\right). \end{aligned} \tag{1.19}$$

Jednakost (1.19) je definirana za $n \geq 2$, dok je ekvivalent tog izraza uz Jeffreyevu distribuciju, $\pi^*(\theta, \sigma^2) = \frac{1}{\sigma^2}$, s n stupnjeva slobode, definiran sa svaki $n \geq 1$.

Konjugirane distribucije su oblika (1.18). Iznimka je što parametri Θ i Σ^2 ne moraju biti a priori nezavisni. Dakle, apriorna distribucija očekivanja Θ ovisi o parametru Σ i o preciznosti procjenitelja samog očekivanja.

Promotrimo dani izraz:

$$\pi(\theta, \sigma^2) = \pi_1(\theta|\sigma^2)\, \pi_2(\sigma^2),$$

gdje je π_1 funkcija gustoće normalne distribucije $N(\mu, \sigma^2/n_0)$ i π_2 je inverzna gama distribucija $Inv\Gamma(\nu/2, {s_0}^2/2)$. Tada je aposteriorna distribucija parametara (Θ, Σ^2):

$$\begin{aligned}\pi(\theta, \sigma^2|x) &\propto \sigma^{-n-\nu-3} exp\left\{-\frac{1}{2\sigma^2}\left[s^2 + {s_0}^2 + n_0(\theta-\mu)^2 + n(\bar{x}-\theta)^2\right]\right\}\\ &= \sigma^{-n-\nu-3} exp\left\{-\frac{1}{2\sigma^2}\left[{s_1}^2 + n_1(\theta-\theta_1)^2\right]\right\},\end{aligned} \tag{1.20}$$

gdje su

$$\begin{aligned} n_1 &= n + n_0, \qquad \theta_1 = \frac{1}{n_1}(n_0\theta_0 + n\bar{x}),\\ {s_1}^2 &= s^2 + {s_0}^2 + ({n_0}^{-1} + n^{-1})^{-1}(\theta_0 - \bar{x})^2. \end{aligned}$$

Te distribucije su zapravo konjugirane, budući da vrijedi:

$$\begin{aligned}\pi(\theta|\bar{x}, s^2, \sigma^2) &\propto \frac{1}{\sigma} exp\left\{-\frac{n_1(\theta-\theta_1)^2}{2\sigma^2}\right\},\\ \pi(\sigma^2|\bar{x}, s^2) &\propto \sigma^{-n-\nu-2} exp\left\{\frac{-{s_1}^2}{2\sigma^2}\right\}.\end{aligned}$$

Kao i u neinformativnom slučaju marginalna aposteriorna distribucija parametra Θ je studentova t-distribucija. Općenito, n_0 je manji od veličine uzorka n. Primjetimo, ako n_0/n ide u 0, dobijemo da vrijedi:

$$\Theta|\bar{x}, \sigma^2 \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right),$$

što odgovara aposteriornoj distribuciji povezanoj sa Jeffrejevim priorom. Ova činjenica je jedan od pokazatelja da neinformativna distribucija često odgovara graničnoj vrijednosti konjugiranih distribucija.

U višedimenzionalnom slučaju i dalje je moguće izvesti konjugiranu apriornu distribuciju. Neka je dano n opservacija, $X_1, X_2, \ldots, X_n$ iz modela $N_p(\Theta, \Sigma)$. Tada je dovoljna statistika dana sa:

$$\begin{aligned}\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i\\ S &= \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^t\end{aligned}$$

i funkcija vjerodostojnosti sa:

$$L(\theta, \Sigma|\bar{x}, s) \propto |\Sigma^2|^{-n/2} exp - \frac{1}{2}\left\{n(\bar{x}-\theta)^t\Sigma^{-1}(\bar{x}-\theta) + tr(\Sigma^{-1}s)\right\},$$

uz $S = s$ i $\bar{X} = \bar{x}$. Ta funkcija vjerodostojnosti navodi na sljedeće konjugirane distribucije:

$$\begin{aligned}\Theta|\Sigma &\sim N_p\left(\mu, \frac{\Sigma}{n_0}\right)\\ \Sigma^{-1} &\sim W_p(\alpha, W),\end{aligned}$$

gdje je W_p p-dimenzionalna Wishartova distribucija. Tada su aposteriorne distribucije dane sa:

$$\begin{aligned}\Theta|\Sigma, \bar{X} = \bar{x}, S = s &\sim N_p\left(\frac{n_0\mu + n\bar{x}}{n_0 + n}, \frac{\Sigma}{n_0 + n}\right)\\ \Sigma^{-1}|\bar{X} = \bar{x}, S = s &\sim W_p(\alpha + n, W_1(\bar{x}, s)),\end{aligned}$$

gdje je

$$W_1(\bar{x}, s)^{-1} = W^{-1} + s + \frac{nn_0}{n + n_0}(\bar{x}-\mu)(\bar{x}-\mu)^t.$$

Napomenimo da je višedimenzionalni slučaj generalizacija jednodimenzionalnog slučaja prikazanog gore, budući da je Wishartova distribucija W_p generalizacija χ^2-distribucije u dimenziji p.

1.6 Testovi i pouzdana područja

Postoji velika razlika između klasičnog i Bayesovog pristupa testiranju. Bayesov pristup je privlačan zbog dobre definiranosti vjerojatnosnih hipoteza. U neinformativnim slučajevima rezultati klasičnog pristupa testiranju mogu voditi sličnim rezultatima kao i u Bayesovom pristupu.

Testovi

Neka je $X = (X_1, X_2, \ldots, X_n)$ slučajni uzorak iz neprekidnog modela $\mathcal{P} = \{f(x|\theta) : \theta \in \Phi\}$, gdje je θ parametar koji želimo procijeniti. Pretpostavimo da želimo testirati sljedeće hipoteze:

$$\begin{aligned}H_0 &: \theta \in \Phi_0\\ H_1 &: \theta \in \Phi_1,\end{aligned}$$

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

U Newman-Pearsonovoj teoriji je problem testiranja formaliziran kroz prostor odluka $\mathcal{D}$ koji je ograničen na $\{da, ne\}$ ili ekvivalentno na $\{0, 1\}$, gdje 1 znači odbacivanje H_0, tj. nul-hipoteze, a 0 ne-odbacivanje. Stoga, probleme testiranja ima smisla promatrati kroz donošenje zaključaka o indikatoru $\mathbb{1}_{\Phi_0}$. No, nekad se zbog procjene grešaka i složenosti bolje odlučiti za testove koji postižu vrijednosti na intervalu $[0, 1]$.

Statistički test za parametar θ je funkcija $\tau : \mathbb{R}^n \to \{0, 1\}$. Ako je x realizacija od X za koju je $\tau(x) = 1$, odbacujemo H_0 u korist H_1, tj. alternativne hipoteze, a ako je $\tau(x) = 0$ ne odbacujemo H_0 u korist H_1. Definiramo *jakost testa* kao funkciju $\gamma : \Phi \to [0, 1]$ sa:

$$\gamma(\theta) := \mathbb{P}_\theta(\tau(X) = 1) = \mathbb{P}_\theta(X \in C) = \mathbb{E}_\theta[\tau(X)],$$

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

Definicija 1.6.1. *Kažemo da je τ^* uniformno najjači test za testiranje H_0 u odnosu na H_1 za parametar θ, ako za svaki drugi test τ vrijedi:*

$$\sup_{\theta \in \Phi_0} \gamma_\tau(\theta) \leq \sup_{\theta \in \Phi_0} \gamma_{\tau^*}(\theta) \Rightarrow \gamma_\tau(\theta) \leq \gamma_{\tau^*}(\theta),$$

za sve $\theta \in \Phi_1$.

Oznaka: UMP(*uniformly most powerfull*).

Definicija 1.6.2. *Kažemo da je test τ uniformno najjači nepristrani test na razini značajnosti α ako je uniformno najjači test na razini značajnosti α u skladu s definicijom 1.6.1 i ako vrijedi:*

$$\sup_{\theta \in \Phi_0} \gamma_\tau(\theta) \leq \inf_{\theta \in \Phi_1} \gamma_\tau(\theta).$$

Oznaka: UMPU(*uniformly most powerfull unbiased*).

Bayesova pouzdana područja

Osim procjene za vrijednost parametra modela, ponekad je potrebno odrediti interval ili područje unutar kojeg će se s velikom vjerojatnošću nalaziti prava vrijednost parametra. Od takvih područja očekujemo da sadrže istinsku vrijednost parametra te da budu mala.

U Newman-Pearsonovoj teoriji pouzdano područje je moguće izvesti iz kritičnog područja UMP, a time specijalno i UMPU testa. Ako je:

$$C_{\theta_0} = \{x : \tau_{\theta_0}(x) = 1\}$$

kritično područje testa

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0,$$

gdje je τ_{θ_0} UMPU test na razini značajnosti α, a tada je odgovarajuće pouzdano područje za θ dano sa:

$$\begin{aligned} C_x &= \{\theta : \ x \in C_\theta\} \\ &= \{\theta : \ \tau_\theta(x) = 1\} \end{aligned}$$

i $\mathbb{P}(\theta \in C_X|\Theta = \theta) = 1 - \alpha$. Općenito, skup C_x se naziva *pouzdano područje* na razini značajnosti α ako za svaki $\theta \in \Phi$, gdje je θ realizacija parametra Θ, vrijedi:

$$\mathbb{P}(\theta \in C_X|\Theta = \theta) \geq (1 - \alpha).$$

Primjer 1.6.3. *Neka je $X \sim N(\theta, \sigma^2)$. 95% UMPU test je dan s:*

$$\tau_\theta(x) = \mathbb{I}_{[0,1.96]}\left(\frac{|x - \theta|}{\sigma}\right),$$

a odgovarajuće pouzdano područje za θ, kad je σ poznat, dano je sa:

$$C_x = [x - 1.96\sigma, x + 1.96\sigma].$$

Primjer 1.6.4. *Neka je $X \sim t_p(N, \theta, I_p)$ p-dimenzionalna t-distribucija sa N stupnjeva slobode i funkcijom gustoće:*

$$f(x|\theta) \propto \left(1 + \frac{1}{N}||x - \theta||^2\right)^{-\frac{N+p}{2}}.$$

Budući da $\frac{||X-\theta||^2}{p}$ ima Fisherovu distribuciju $F(p, N)$, $(1 - \alpha)$% pouzdano područje biti će:

$$C_x = \left\{\theta; \ ||x - \theta||^2 \leq p f_\alpha(p, N)\right\},$$

gdje je $f_\alpha(p, N)$ α-kvantil Fisherove distribucije $F(p, N)$.

U okvirima bayesovske statistike, procjena pouzdanih intervala je problem obrnut od računanja aposteriorne i/ili prediktivne vjerojatnosti. Pouzdani intervali su zapravo intervali sa zadanom aposteriornom vjerojatnosti. Za $\alpha \in <0, 1>$, $100(1 - \alpha)$% dvostrani bayesovski pouzdani interval odgovara $(\alpha/2)$ (u oznaci: $\theta_{\alpha/2}$) i $(1 - \alpha/2)$ (u oznaci: $\theta_{1-\alpha/2}$)

kvantilu aposteriorne distribucije parametra Θ. Takav interval za jednodimenzionalni parametar Θ zapisujemo kao

$$[\theta_{\alpha/2}, \theta_{1-\alpha/2}].$$

Ako je pak potrebno procijeniti više parametara odjednom, tada govorimo o *pouzdanim područjima*. U Newman-Pearsonovoj teoriji pouzdana područja se mogu izvesti iz UMPU testova (preko argumenta dualnosti). Ova metoda se često koristi u praksi, na primjer, u slučaju linearne regresije, no svejedno ima nekoliko mana. Newman-Pearsonov pristup nije bez nedostataka i optimalnost UMPU testa ne mora vrijediti. Dakle, ni pouzdano područje izvedeno iz takvog testa ne mora biti optimalno.

Definicija 1.6.5. *Neka je C_x neko pouzdano područje na razini značajnosti α. Kažemo da je skup $A \in \Omega$ negativno pristran značajan podskup za pouzdano područje C_x ako postoji $\epsilon > 0$ takav da vrijedi*

$$\mathbb{P}_\theta(\theta \in C_X | X \in A) \leq 1 - \alpha - \epsilon,$$

za svaki $\theta \in \Phi$, realizaciju parametra Θ.

Slično možemo definirati i *pozitivno pristran značajan podskup* za pouzdano područje C_x.

Definicija 1.6.6. *Neka je C_x neko pouzdano područje na razini značajnosti α. Kažemo da je skup $A \in \Omega$ pozitivno pristran značajan podskup za pouzdano područje C_x ako postoji $\epsilon > 0$ takav da vrijedi*

$$\mathbb{P}_\theta(\theta \in C_X | X \in A) \geq 1 - \alpha + \epsilon,$$

za svaki $\theta \in \Phi$, realizaciju parametra Θ.

Definicija 1.6.7. *Za apriornu distibuciju π, skup C_x naziva se α-vjerodostojan ako vrijedi*

$$\mathbb{P}_\pi(\Theta \in C_x | X = x) \geq 1 - \alpha, \tag{1.21}$$

odnosno $\mathbb{P}_{\Theta|X=x}(C_x) \geq 1 - \alpha$, gdje je $\mathbb{P}_{\Theta|X=x}$ aposteriorna distribucija parametra Θ.
To područje se naziva α-vjerodostojno područje najveće aposteriorne gustoće (u oznaci HPD[3]) ako vrijedi:

$$\{\theta;\ \pi(\theta|x) > k_\alpha\} \subset C_x \subset \{\theta;\ \pi(\theta|x) \geq k_\alpha\}, \tag{1.22}$$

gdje je k_α najveći takav broj tako da vrijedi (1.21).

Napomena 1.6.8.

- *Primjetimo kako* (1.22) *uključuje i poseban slučaj kada je $\{\theta;\ \pi(\theta|x) = k_\alpha\}$ neprazan skup.*

[3] HPD označava *highest posterior density*

- *Ako postoji HPD područje, tada je ono i minimalno α-pouzdano područje.*

Primjer 1.6.9. *Neka je X iz binomne distribucije, tj. $X \sim \mathcal{B}(n, p)$ i neka je p realizacija parametra P tako da vrijedi $P \sim Beta(1/2, 1/2)$. Tada je aposteriorna distribucija od P uz dano $X = x$, tj. $P|X = x \sim Beta(x + 1/2, n - x + 1/2)$. Pouzdani intervali za P mogu se izvesti pomoću kumulativne funkcije beta distribucije s gore navedenim parametrima. U sljedećoj tablici ćemo pokazati pouzdane intervale za $n = 5, \alpha = 5\%$ i $\alpha = 10\%$.*

x	0	1	2
$\alpha = 5\%$	[0.000, 0.38]	[0.022, 0.621]	[0.094, 0.791]
$\alpha = 10\%$	[0.000, 0.308]	[0.036, 0.523]	[0.128, 0.74]

Tablica 1.3: Pozdani intervali za binomnu distribuciju

Primjetimo da kod diskretnih distribucija značajnu prednost ima bayesovski pristup u odnosu na klasični pristup. U stvari, pouzdani intervali uključuju randomizacijski korak za postizanje nominalne razine značajnosti. Apriorno modeliranje izbjegava randomizaciju i koristi prednost poznate apriorne informacije. Primjetimo također da se može pojaviti nepravilna apriorna distribucija i tu se ne susrećemo s istim poteškoćama kao kod testiranja nulte hipoteze. Zapravo, aposterirorna vjerodostojna područja mogu se izvesti dok god je aposteriorna distribucija definirana. Neka klasična vjerodostojna područja se mogu izraziti pomoću vjerodostojnih područja povezanih sa generaliziranom distribucijom.

Postoje rijetki slučajevi kada je pouzdano područje nepovezano sa HPD α-vjerodostojnim područjem, što može ukazivati na odabir neodgovarajuće apriorne distribucije. Tada je uobičajno rješenje zamjena HPD α-vjerodostojnog područja sa drugim vjerodostojnim područjem $[C_1(x), C_2(x)]$ tako da vrijedi

$$\mathbb{P}_\pi(\Theta < C_1(x)|X = x) = \mathbb{P}_\pi(\Theta > C_2(x)|X = x) = \frac{\alpha}{2}.$$

Iako je ovo teorijski jednostavna metoda, njena primjena u praksi je često komplicirana, pogotovo kad je skup Φ velik ili ako aposteriorna funkcija distribucije nije dana u zatvorenoj formi. Prvo rješenje problema bi bilo primjena nekih numeričkih metoda kojima se nećemo baviti. Drugo rješenje je konstrukcija aproksimativnog pouzdanog područja pomoću normalne distribucije. Pretpostavimo da je aposteriorna distribucija od Θ otprilike normalna

$$N_p\left(\mathbb{E}^\pi(\Theta|X = x), Var^\pi(\Theta|X = x)\right).$$

Tada je aproksimacija pouzdanog područja dana s:

$$C_\alpha = \{\theta;\ (\theta - \mathbb{E}^\pi[\Theta|X = x])^t Var^\pi(\Theta|X = x)^{-1}(\theta - \mathbb{E}^\pi[\Theta|X = x]) \leq {k_\alpha}^2\},$$

gdje je ${k_\alpha}^2$ α-kvantil χ^2-distibucije s p stupnjeva slobode. Ovakva aproksimacija je opravdana samo za velike uzorke ali se ipak koristi zbog brzog i efikasnog načina za dobiti pouzdana područja.

Predloženi su i drugi načini izvođenja pouzdanih područja. Izvođenje HPD područja povezanog uz minimizaciju volumena prostora tako da vrijedi

$$\mathbb{P}(\Theta \in C_\alpha | \Theta = \theta) \geq 1 - \alpha.$$

Izvodimo ih pomoću funkcije gubitka koja uzima u obzir volumen pouzdanog područja (u oznaci: vol) i njegovu vjerojatnost prekrivanja. Na primjer, jednostavna verzija toga je linearna kombinacija:

$$L(C, \theta) = vol(C) + c\mathbb{1}_{[\theta \notin C]},$$

s pripadnom funkcijom rizika

$$R(C, \theta) = \mathbb{E}_\theta(vol(C_X)) + c\mathbb{P}(\Theta \notin C_X | \Theta = \theta).$$

Konstanta c može biti povezana uz određeni nivo značajnosti. Glavni nedostatak linearnog gubitka je nejednako vrednovanje volumena područja i vjerojatnosti prekrivanja. Indikatorska funckija, $\mathbb{1}_{[\theta \notin C]}$, varira između 0 i 1, dok volumen područja može rasti i do beskonačnosti pa ta asimetrija vodi pristranosti u korist malih područja povjerenja.

Važan čimbenik, koji je često zanemaren, u izvođenju intervala pouzdanosti je način na koji će interval biti korišten. Također je to vrlo bitno za konstrukciju funkcije gubitka. Donositelj odluke bi trebao birati funkciju gubitka prema svojim potrebama. Neke potrebe su, na primjer:

- određivanje pouzdanog područja parametara kao pomoć u daljnjem određivanju procjenitelja (na primjer, definiranje apriorne distribucije sa nosačem jednakim određenom pouzdanom području)

- oslanjanje na dobiveno pozdano područje za donošenjenje odluke o *testiranoj* hipotezi (odbacivanje nulte hipoteze ako pouzdano područje ne sadrži određenu vrijednost parametara)

- definiranje indikatora o učinkovitosti procjenitelja pomoću veličine (volumena) pouzdanog područja.

Poglavlje 2

Bayesovski linearni regresijski model

Regresijska analiza je središnji alat u primijenjenoj statistici čiji je cilj odgovoriti na pitanje kako pojedine varijable (nezavisne varijable) utječu na određeni ishod (zavisnu varijablu).

2.1 Univarijatni linearni regresijski model

Univarijatni linearni regresijski model pokušava objasniti varijabilnost jedne varijable, tzv. *zavisne varijable* uz pomoć druge varijable, tzv. *nezavisne* ili *eksplanatorne varijable*, stavljajući linearnu vezu među njima. Jednadžba modela:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{K-1} X_{K-1} + \epsilon, \tag{2.1}$$

gdje su:

- Y = zavisna varijabla
- X_k = nezavisna varijabla, $k = 1, \ldots, K-1$
- β_0 = slobodni član
- β_k = regresijski koeficijent, $k = 1, \ldots, K-1$, predstavlja utjecaj za koliko se jedinica mijenja $X_k, k = 1, \ldots, K-1$, obzirom na Y, zadržavajući preostalu nezavisnu varijablu X_j, $j \neq k$ fiksnom
- ϵ = regresijski šum ili greška.

Regresijski šum je neobjašnjeni dio linearne veze između zavisnih i nezavisnih varijabli. Dok $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{K-1} X_{K-1}$ predstavlja varijabilnost od Y objašnjenu sa $X_k, k = 1, 2, \ldots, K-1$, ϵ predstavlja varijabilnost od Y koja nije objašnjena.

Pretpostavimo da imamo n opservacija zavisnih i nezavisnih varijabli, i neka su (x_i, y_i) opažene vrijednosti od (X, Y), za $i = 1, 2, \dots, n$. Podaci su opisani sa:

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_{K-1} x_{K-1,i} + \epsilon_i \quad i = 1, \dots, n, \tag{2.2}$$

gdje je sa i označena i-ta opservacija slučajne varijable. Za opisati slučajnost, treba napraviti pretpostavku o distribuciji. Zbog jednostavnosti, pretpostavimo da su ϵ_i, $i = 1, ..., n$, nezavisno i jednako distribuirani (oznaka: n.j.d.) sa normalnom distribucijom, $\epsilon_i \sim N(0, \sigma^2)$. Tada zavisna varijabla, Y, također ima normalnu distribuciju, uvjetnu na $X = x_i$

$$Y|X = x_i \sim N(\mu_i, \sigma^2), \tag{2.3}$$

gdje je $\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_{k-1} x_{k-1,i}$, a $Y_i = \mathbb{E}[Y|X = x_i] + \epsilon_i = \mu_i + \epsilon_i$. Primijetimo da pretpostavka o jednakosti varijance u (2.3) je vrlo restriktivna. Poslje ćemo se vratiti na ovaj problem.

Izraz (2.2) se često piše kao:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{2.4}$$

gdje je $\mathbf{Y}$ $(n \times 1)$-dimenzionalni vektor,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},$$

β je $(K \times 1)$-dimenzionalni vektor,

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{K-1} \end{bmatrix},$$

a $\mathbf{X}$ je $(n \times K)$ matrica čija se prvi stupac sastoji samo od jedinica,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{K-1,1} \\ 1 & x_{1,2} & \dots & x_{K-1,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & \dots & x_{K-1,n} \end{bmatrix},$$

i ϵ je $(n \times 1)$-dimenzionalni vektor,

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Za regresijski šum pretpostavku o normalnoj distribuciji zapisujemo kao

$$\epsilon \sim N(0, \sigma^2 I_n),$$

gdje je I_n $(n \times n)$ matrica identiteta. Iz (2.4) trebamo procijeniti parametre β i σ^2. Pod pretpostavkom da je šum normalno distribuiran, zapisujemo funkciju vjerodostojnosti kao

$$L(\beta_0, \beta_1, \beta_2, \sigma | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \times exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_1 - \beta_0 - \beta_1 x_{1,i} - \cdots - \beta_{K-1} x_{K-1,i})^2 \right\},$$

gdje je

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

vektor opaženih vrijednosti od $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$. Ili, u vektorskom zapisu imamo funkciju vjerodostojnosti za parametre multivarijatne normalne distribucije

$$L(\beta, \sigma | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^t (\mathbf{y} - \mathbf{X}\beta) \right\}. \tag{2.5}$$

Bayesovska procjena univarijatnog regresijskog modela

U klasičnom smislu, regresijski parametri su obično procijenjeni metodom maksimalne vjerodostojnosti u odnosu na β i σ^2, na primjer, funkcija vjerodostojnosti je jednaka kao u (2.5) ako se pretpostavlja normalna distribucija. Kad pretpostavimo da je šum normalno distribuiran, metoda maksimalne vjerodostojnosti i metoda najmanjih kvadrata (u oznaci: MNK) dovode do istih procijenjenih parametara. Može se pokazati da je procjenitelj regresijskog koeficijenta, $\hat{\beta}$, metodom najmanjih kvadrata, dan sa

$$\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X} \mathbf{Y}, \tag{2.6}$$

gdje je sa simbolom $(\cdot)^t$ označeno matrično transponiranje. Procjenitelj za σ^2 je

$$\hat{\sigma}^2 = \frac{1}{n-K} (\mathbf{Y} - \mathbf{X}\hat{\beta})^t (\mathbf{Y} - \mathbf{X}\hat{\beta}). \tag{2.7}$$

Pretpostavit ćemo dva slučaja za apriornu distribuciju: nepravilnu apriornu distribuciju i informativnu konjugiranu apriornu distribuciju za regresijski parametarski vektor (β, σ^2). Označimo sa (B, Σ^2) te parametre kao slučajne veličine.

Analogno spomenutoj neinformativnoj apriornoj distribuciji ćemo opisati što je informativna apriorna distribucija. Dakle, ako imamo poznate informacije za neki model, tada možemo formirati apriornu funkciju gustoće i odrediti hiperparametre konjugirane distribucije. Tu formiranu distribuciju nazivamo *informativna konjugirana apriorna distribucija*. Navest ćemo kao primjer apriornu distribuciju temperature sutradan u podne. Bilo bi razumno da se za apriornu distribuciju uzme normalna razdioba s očekivanom vrijednosti jednakoj današnjoj popodnevnoj temperaturi, s varijancom koja je jednaka dnevnoj varijanci atmosferske temperature, ili alternatvno, distribucija temperature za taj dan u godini. Vidimo da aposteriorna distribucija (današnje temperature) postaje apriorna distribucija (sutrašnje temperature).

Nepravilna apriorna distribucija

Zajednička nepravilna apriorna distribucija za parametre (B, Σ^2) je dana sa gustoćom:

$$\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}, \tag{2.8}$$

gdje regresijski koeficijenti mogu poprimiti biti bilo koju realnu vrijednost, $-\infty < \beta_k < \infty$, za $k = 1, 2, \ldots, K$, a varijanca je pozitivna, $\sigma^2 > 0$.

Kombinirajući funkciju vjerodostojnosti iz (2.5) i apriornu distribuciju danu gore, dobivamo aposteriornu distribuciju parametara modela na sljedeći način:

- Aposteriorna distribucija parametra B uvjetno na $\Sigma^2 = \sigma^2$ je (multivarijatna) normalna: (vidjeti dodatak za definiciju distribucije)

  $$N(\hat{\beta}, (\mathbf{X}^t\mathbf{X})^{-1}\sigma^2), \tag{2.9}$$

  gdje je $\hat{\beta}$ procjena metodom najmanjih kvadrata iz (2.6) i $(\mathbf{X}^t\mathbf{X})^{-1}\sigma^2$ je kovarijacijska matrica od $\hat{\beta}$. Gustoću te distribucije označimo sa $\pi(\beta|\mathbf{y}, \mathbf{X}, \sigma^2)$.

- Aposteriorna distribucija od Σ^2 je inverzna χ^2- distribucija s gustoćom:

  $$\pi(\sigma^2|\mathbf{y}, \mathbf{X}) \sim Inv\chi^2(n - K, \hat{\sigma}^2), \tag{2.10}$$

  gdje je $\hat{\sigma}^2$ procjena od σ^2 iz (2.7).

Integrirajući zajedničku aposteriornu distribuciju, može se pokazati:

$$\pi(\beta, \sigma^2|\mathbf{y}, \mathbf{X}) = \pi(\beta|\mathbf{y}, \mathbf{X}, \sigma^2)\pi(\sigma^2|\mathbf{y}, \mathbf{X})$$

uzimajući u obzir σ^2 i da je bezuvjetna aposteriorna distribucija parametra B multivarijatna Studentova t-distribucija sa funkcijom gustoće danom sa:

$$\pi(\beta|\mathbf{y}, \mathbf{X}) \propto \left((n-K) + (\beta - \hat{\beta})^t \frac{\mathbf{X}^t\mathbf{X}}{\hat{\sigma}^2}(\beta - \hat{\beta})\right)^{-n/2}. \tag{2.11}$$

Primjetimo da je Studentova distribucija od B teškog repa. Premda je vektor sredina od B nepromjenjen, njegova varijanca je narasla (u prosjeku) za vrijednost $\nu/(\nu-2)$:

$$\Sigma_B = \hat{\sigma}^2(\mathbf{X}^t\mathbf{X})^{-1}\frac{\nu}{\nu-2},$$

gdje $\nu = n - K$ označava stupnjeve slobode parametara multivarijatne Studentove t-distribucije.

Da bi zaključili našu raspravu o aposteriornoj distribuciji u slučaju nepravilne apriorne distribucije, pretpostavimo da smo posebno zainteresirani za jedan regresijski koeficijent, neka je to B_k. Može se pokazati da standardizirani B_k ima Studentovu t-distribuciju sa $n-K$ stupnjeva slobode:

$$\frac{B_k - \hat{\beta}_k}{(h_{k,k})^{1/2}} \mid \mathbf{y}, \mathbf{X} \sim t_{n-K}, \tag{2.12}$$

gdje je $h_{k,k}$ k-ti dijagonalni element od $\hat{\sigma}^2(\mathbf{X}^t\mathbf{X})^{-1}$ i $\hat{\beta}_k$ je procjena od β_k metodom najmanjih kvadrata (odgovarajuća komponenta od $\hat{\beta}$). Bayesovski intervali za β_k se mogu analitički konstruirati.

Informativna apriorna distribucija

Pretpostavimo da regresijski koeficijent, vektor B, ima normalnu apriornu distribuciju (uvjetna na $\Sigma^2 = \sigma^2$), a Σ^2 inverznu χ^2 apriornu distribuciju:

$$B \mid \Sigma^2 = \sigma^2 \sim N(\beta_a, \sigma^2\mathbf{A})$$

i

$$\Sigma^2 \sim Inv\chi^2(\nu_0, {c_0}^2).$$

Četri parametra se moraju odrediti apriorno: $\beta_a, \mathbf{A}, \nu_0$ i ${c_0}^2$. Matrica $\mathbf{A}$ je često oblika $\tau^{-1}(\mathbf{X}^t\mathbf{X})^{-1}$ kako bi se dobila apriorna kovarijanca koja je jednaka kovarijacijskoj matrici procjenitelja od β dobivenog metodom najmanjih kvadrata. Variranjem (skalarnog) parametra τ dozvoljavamo prilagodbu stupnja pouzdanosti kada je sredina od B jednaka β_a: što je manja vrijednost od τ, to je veći stupanj nesigurnosti za B.

Najlakši način za izvesti apriornu sredinu β_a je fiksirati je na neku zadanu vrijednost (kao na primjer na 0, ovisno o konceptu procjene), osim ako je preciznija apriorna informacija poznata, ili je izjednačiti sa procjeniteljem dobivenim MNK, $\hat{\beta}$, dobivenog iz regresije (2.1) provedene na apriornom uzorku.

Parametri inverzne χ^2 distribucije mogu se dobiti pomoću apriornih podataka na sljedeći način:

$$\begin{aligned} \nu_0 &= n_0 - K \\ {c_0}^2 &= \frac{1}{\nu_0}(\mathbf{y}_0 - \mathbf{X}_0\hat{\beta_a})^t(\mathbf{y}_0 - \mathbf{X}_0\hat{\beta_a}), \end{aligned} \tag{2.13}$$

gdje se podaci označeni s indeksom 0 odnose na apriorne podatke. Ako nemamo dane apriorne podatke, inverzni χ^2 hiperparametri se mogu izraziti pomoću pretpostavljanja apriorne sredine i varijance od σ^2, pomoću funkcije gustoće inverzne χ^2 distribucije:

$$f(x|\nu, c) = \frac{1}{\Gamma(\nu/2)}\left(\frac{\nu}{2}\right)^{\nu/2} c^{\nu} x^{-(\nu/2+1)} exp\left(-\frac{\nu c}{2x}\right),$$

gdje su $\nu > 0$, $c > 0$ i $x > 0$.

Aposteriorna distribucija za model s parametrima B i Σ^2 ima isti oblik kao i apriorna distribucija. Međutim, njihovi parametri se ažuriraju u skladu sa podacima zajedno s apriornim pretpostavkama.

- Aposteriorna distribucija za B uvjetno na $\Sigma^2 = \sigma^2$ dana je sa:

  $$\pi(\beta|\mathbf{y}, \mathbf{X}, \sigma^2) \sim N(\beta^*, \Sigma_\beta),$$

  gdje je aposteriorna sredina i kovarijacijska matrica od β dana sa:

  $$\beta^* = (\mathbf{A}^{-1} + \mathbf{X}^t\mathbf{X})^{-1}(\mathbf{A}^{-1}\beta_a + \mathbf{X}^t\mathbf{X}\hat{\beta}) \tag{2.14}$$

  i

  $$\Sigma_\beta = \sigma^2(\mathbf{A}^{-1} + \mathbf{X}^t\mathbf{X})^{-1}.$$

  Možemo smatrati da je aposteriorna sredina prosječna težina apriornih sredina i procjenitelja od β dobivenog MNK, kao što je ranije u poglavlju već spomenuto.

- Inverzna χ^2 aposteriorna distribucija od Σ^2 je dana sa:

  $$\pi(\sigma^2|\mathbf{y}, \mathbf{X}) \sim Inv\chi^2(\nu^*, c^{2*}). \tag{2.15}$$

  Parametri aposteriorne distribucije od Σ^2 su dani sa:

  $$\begin{aligned} \nu^* &= \nu_0 + n \\ \nu^* c^{*2} &= (n-K)\hat{\sigma}^2 + (\beta_a - \hat{\beta})^t H(\beta_a - \hat{\beta}) + \nu_0 {c_0}^2), \end{aligned} \tag{2.16}$$

gdje je $H = ((\mathbf{X}^t\mathbf{X})^{-1} + \mathbf{A})^{-1}$.

Kako je spomenuto ranije, možemo derivirati marginalnu aposteriornu distribuciju od B tako da integriramo po σ^2 iz zajedničke aposteriorne distribucije. Ponovno dobijemo multivarijatnu Studentovu t-distibuciju, $t(t(\nu^*, \beta^*, \mathbf{Q}))$,

$$\pi(\beta|\mathbf{y}, \mathbf{X}) \propto (\nu^* + (\beta - \beta^*)^t\mathbf{Q}(\beta - \beta^*))^{-\nu^*/2},$$

gdje je $\mathbf{Q} = (\mathbf{A}^{-1} + \mathbf{X}^t\mathbf{X})/c^{*2}$.

Sredina od B ostaje ista, β^* (nezavisna je od σ^2), dok se bezuvjetna kovarijacijska matrica može izračunati koristeći funkciju gustoće multivarijatne normalne distibucije (vidjeti dodatak). Može se pokazati da je marginala aposteriorna distribucija za jedan koeficijent, B_k:

$$\frac{B_k - \beta_k{}^*}{(q_{k,k})^{1/2}} \mid \mathbf{y}, \mathbf{X} \sim t_{\nu_0+n-K}, \tag{2.17}$$

gdje je $q_{k,k}$ k-ti dijagonalni element matrice $\mathbf{Q}^{-1}$, a $\beta_k{}^*$ je k-ti element od β^*.

Predikcije

Pretpostavimo da želimo predvidjeti zavisnu varijablu Y p koraka unaprijed i označimo te varijable sa $p \times 1$ vektorom $\tilde{\mathbf{Y}} = (Y_{\tau+1}, Y_{\tau+2}, \ldots, Y_{\tau+p})$. Pretpostavimo da su buduće opservacije nezavisne varijable poznate i dane sa $\tilde{\mathbf{X}}$. Neka je $\tilde{\mathbf{y}} = (y_1, y_2, \ldots, y_p)$. Koristimo izraz

$$f(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{X}}, \mathbf{X})$$

za izraziti uvjetnu funkciju gustoće od $\tilde{\mathbf{Y}}$ uvjetno na $\mathbf{Y} = \mathbf{y}, \Sigma^2 = \sigma^2, B = \beta, \tilde{\mathbf{X}}$ i $\mathbf{X}$ odnosno za prediktivnu vjerojatnost u kontekstu linearne regresije:

$$f(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{X}}, \mathbf{X}) = \int \int f(\tilde{\mathbf{y}}|\beta, \sigma^2, \tilde{\mathbf{X}})\pi(\beta, \sigma^2|\mathbf{y}, \mathbf{X})d\beta \, d\sigma^2,$$

gdje je sa $\pi(\beta, \sigma^2|\mathbf{y}, \mathbf{X})$ označena zajednička aposteriorna distribucija od β i σ^2, a sa $f(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{X}}, \mathbf{X})$ uvjetna funkcija gustoće od $\mathbf{Y}$ uz dano $B = \beta, \Sigma^2 = \sigma^2, \tilde{\mathbf{X}}$.

Može se pokazati da je prediktivna distribucija multivarijatna Studentova t-distribucija. Pod pretpostavkom da smo u slučaju nepravilne apriorne distribucije, tada je prediktivna distribucija dana sa

$$f(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{X}}, \mathbf{X}) \sim t(n - K, \tilde{\mathbf{X}}\hat{\beta}, S), \tag{2.18}$$

gdje je $S = \sigma^2(I_p + \tilde{\mathbf{X}}(\mathbf{X}^t\mathbf{X})^{-1}\tilde{\mathbf{X}}^t)$, a $\hat{\beta}$ je aposteriorna sredina od B pod pretpostavkom da smo u slučaju nepravilne apriorne distribucije. U slučaju informativne apriorne distribucije, prediktivna distribucija od $\tilde{\mathbf{Y}}$ je

$$f(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{X}}, \mathbf{X}) \sim t(\nu_0 + n, \tilde{\mathbf{X}}\beta^*, \mathbf{V}), \tag{2.19}$$

gdje je $\mathbf{V} = c^{*2}(I_p + \tilde{\mathbf{X}}(A^{-1} + \mathbf{X}^t\mathbf{X})^{-1}\tilde{\mathbf{X}}^t)$, a β^* je aposteriorna sredina od B u (2.14).

Zasigurno, ponovno je moguće derivirati distribuciju za predikcijsku distribuciju jedne komponente od $\tilde{\mathbf{Y}}$ - univarijatna Studentova t-distibucija - u oba scenarija:

$$\frac{\tilde{Y}_k - \tilde{\mathbf{X}}^k\hat{\beta}_k}{{s_{k,k}}^{1/2}} \sim t_{n-K},$$

gdje je $\tilde{\mathbf{X}}^k$ k-ti redak od $\tilde{\mathbf{X}}$ (opservacije nezavisne varijable koje se odnose na k-ti budući period), a $s_{k,k}$ je k-ti dijagonalni element od sklarne matrice S, u (2.18), i vrijedi

$$\frac{\tilde{Y}_k - \tilde{\mathbf{X}}^k{\beta_k}^*}{{v_{k,k}}^{1/2}} \sim t_{v_0+n-K},$$

gdje $v_{k,k}$ označava $k-$ti dijagonalni element skalarne matrice $\mathbf{V}$ iz (2.19).

Slučaj nejednakih varijanci

Već smo ranije u poglavlju spomenuli da pretpostavka o jednakosti varijaci u (2.3) može biti restriktivna. Pojasnit ćemo to kroz dva primjera. Prvo, pretpostavimo da imamo n opservacija varijable Y dobivenih kroz određeni vremenski period. Uobičajna praksa u statističkoj procjeni je da se koristi skup podataka koji se duže vremena prikuplja, najčešće godinama. Promjene u ekonomiji i financijama mogu uzrokovati promjene u varijanci (kao i u sredinama). Pretpostavka o jednakosti varijanci bi tad vodila do precjenjivanja varijance u razdobljima gdje varijanca ima nisku vrijednost i isto tako podcjenjivanje varijance u razdobljima s visokom vrijednosti varijance.

Drugo, ako je naš problem procjene zasnovan na određenom vremenskom razdoblju, pretpostavka o jednakosti varijance može ponovno biti narušena. Sve veličine iz našeg uzorka mogu potencijalno imati različite varijance, stoga $var(Y_i) = {\sigma_i}^2$, umjesto $var(Y_i) = \sigma^2$ kao u (2.3). Tada bi procjena bila složenija zato što bi ovo značilo veći broj nepoznatih parametara (varijance i regresijskih koeficijenata) od danih podataka.

U praksi će vjerojatno biti moguće prepoznati homogene grupe uzoraka za koje pretpostavljamo da imaju jednake varijance. Može se pretpostaviti nekakava povezanost između nepoznatih varijanci, a to može voditi smanjenju broja nepoznatih parametara za procjenu. Sad ćemo pokazat jedan način kako se rješava slučaj nejednakih varijanci kad je uzorak podjeljen na dvije homogene grupe.

Označimo opservacije iz dvije grupe kao $\mathbf{Y}_1 = (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n_1})$ i $\mathbf{Y}_2 = (Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n_2})$ tako da je $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2})$ i $n_1 + n_2 = n$. Univarijatni regresijski izraz iz (2.1) je zapisan kao:

$$\begin{aligned} \mathbf{Y_1} &= \mathbf{X_1}\beta + \epsilon_1 \\ \mathbf{Y_2} &= \mathbf{X_2}\beta + \epsilon_2, \end{aligned} \tag{2.20}$$

gdje su $\mathbf{X_1}$ i $\mathbf{X_2}$ $(n_1 \times K)$ i $(n_2 \times K)$ matrice opservacija nezavisne varijable. Pretpostavimo da je šum nezavisan i distribuiran na sljedeći način:

$$\epsilon_1 \sim N(0, {\sigma_1}^2 I_{n_1})$$
$$\epsilon_2 \sim N(0, {\sigma_2}^2 I_{n_2}),$$

gdje ${\sigma_1}^2 \neq {\sigma_2}^2$. Funkcija vjerodostojnosti za parametre modela, $\beta, {\sigma_1}^2$ i ${\sigma_2}^2$ je dana sa

$$\begin{aligned} L(\beta, {\sigma_1}^2, {\sigma_2}^2 | \mathbf{y}, \mathbf{X_1}, \mathbf{X_2}) &\propto ({\sigma_1}^2)^{-n_1/2} ({\sigma_2}^2)^{-n_2/2} \\ &\times exp\left(-\frac{1}{2{\sigma_1}^2}(\mathbf{y_1} - \mathbf{X_1}\beta)^t(\mathbf{y_1} - \mathbf{X_1}\beta) - \frac{1}{2{\sigma_2}^2}(\mathbf{y_2} - \mathbf{X_2}\beta)^t(\mathbf{y_2} - \mathbf{X_2}\beta)\right), \end{aligned} \tag{2.21}$$

gdje je $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ vektor opaženih vrijednost od $\mathbf{Y}$, a $\mathbf{y_1}, \mathbf{y_2}$ vektori opaženih vrijednosti od $\mathbf{Y_1}, \mathbf{Y_2}$.

Neinformativna apriorna distribucija se može utvrditi, pod pretpostavkom da su parametri nezavisni. Tada je apriorna distribucija zapisana kao:

$$\pi(\beta, \sigma_1, \sigma_2) \propto \frac{1}{\sigma_1 \sigma_2}.$$

Jednostavno je zapisati zajedničku aposteriornu gustoću od $\beta, {\sigma_1}^2$ i ${\sigma_2}^2$, koja se onda može integrirati da bi dobili marginalnu aposteriornu distribuciju vektora regresijskih koeficijenata.

Pokazano je da je marginalna aposteriorna distribucija od B produkt dvije multivarijatne Studentove t-gustoće:

$$\pi(\beta | \mathbf{y}, \mathbf{X}_1, \mathbf{X}_2) \sim t(\nu_1, \hat{\beta}_1, S_1) \times t(\nu_2, \hat{\beta}_2, S_2),$$

gdje je za $i = 1, 2, \hat{\beta}_i$ procjena metodom najmanjih kvadrata od β iz (2.20),

$$\nu_i = n_i - K \quad S_i = \hat{s}_i^2(\mathbf{X}_i{}^t \mathbf{X}_i),$$

i

$$\hat{s}_i^2 = \frac{1}{n_i - K}(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_i)^t(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_i).$$

Također je pokazano da se marginalna aposteriorna distribucija od B može aproksimirati sa normalnom distribucijom.

Primjer

Navest ćemo jedan od poznatijih primjera linearne regresije. Pierce (1948.g) je mehanički izmjerio frekvenciju (broj pomaka krila po sekundi) cvrčanja (glasanja cvrčka) cvrčaka pri različitim temperaturama. Budući da su cvrčci hladnokrvne životinje, na njihove fiziološke procese i ukupni metabolizam utječe temperatura. Prema tome, postoji razlog za vjerovanje da temperatura ima značajan utjecaj na njihovo ponašanje kao što je frekvencija cvrčanja.

Mi želimo procjeniti kolika je temperatura na temelju cvrčanja cvrčaka.

Temperaturu ćemo izraziti u Fahrenheitima, a navest ćemo formulu pomoću koje ju možemo preračunati u Celzijeve stupnjeve:

$$^{\circ}C = (^{\circ}F - 32) \times 5/9.$$

Cvrcanje/sekundi	Temperatura (°F)
20.0	88.6
16.0	71.6
19.8	93.3
18.4	84.3
17.1	80.6
15.5	75.2
14.7	69.7
15.7	71.6
15.4	69.4
16.3	83.3
15.0	79.6
17.2	82.6
16.0	80.6
17.0	83.5
14.4	76.3

Tablica 2.1: Podaci o frekvenciji cvčanja cvrčaka i temperature izražene u Fahrenheitovim stupnjevima

Postavljamo linearni model

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

gdje su slučajne varijable **Y** zavisne i sadržavaju podatke o temperaturi, a **X** je spomenuta matrica dizajna čiji se prvi stupac sastoji od jedinica a drugi sadržava podatke o frekvenciji cvrčanja cvrčaka.

Zadat ćemo apriorne i izračunati aposterione parametre distribucije parametara B i Σ^2. Pretpostavit ćemo slučaj informativne konjugirane apriorne distribucije, dakle, regresijski koeficijent, vektor B, ima normalnu apriornu distribuciju (uvjetna na $\Sigma^2 = \sigma^2$), a Σ^2

inverznu χ^2 apriornu distribuciju:

$$B \mid \Sigma^2 = \sigma^2 \sim N(\beta_a, \sigma^2 \mathbf{A})$$

i

$$\Sigma^2 \sim Inv\chi^2(\nu_0, {c_0}^2).$$

Četri parametra se moraju odrediti apriorno: β_a, $\mathbf{A}$, ν_0 i ${c_0}^2$. Matrica $\mathbf{A}$ je oblika $\tau^{-1}(\mathbf{X}^t\mathbf{X})^{-1}$ kako bi se dobila apriorna kovarijanca koja je jednaka kovarijacijskoj matrici procjenitelja od β dobivenog metodom najmanjih kvadrata. Za primjer, pretpostavit ćemo da je apriorna sredina β_a jednaka procijenitelju dobivenom MNK, $\hat{\beta}$, dobivenog iz regresije na apriornom uzorku.

```
> cvrcanje<-c(20.0,16.0,19.8,18.4,17.1,15.5,14.7,15.7,
15.4,16.3,15.0,17.2,16.0,17.0,14.4)
>
> temp<-c(88.6,71.6,93.3,84.3,80.6,75.2,69.7,71.6,
69.4,83.3,79.6,82.6,80.6,83.5,76.3)
>
> y1<-temp
> x1<-cvrcanje
>#kreiranje matrice dizajna
> X2 <- cbind(matrix(1,15,1),x1)
```

Koristeći formulu (2.6) dobijem apriornu sredinu:

```
> b_kapica
[1] 22.848982  3.410323
```

odnosno, sredina slobodnog člana je $\beta_0 = 22.848982$, a sredina regresijskog koeficijenta $\beta_1 = 3.410323$.

Procjenjujemo kovarijacijsku matricu za B po formuli već spomenutoj u primjeru ($\sigma^2\mathbf{A}$):

```
>xxi<-solve(t(X2)%*%X2)
> aprior_var<-var(x1)*xxi
> aprior_var
                        x1
   19.799476 -1.18333333
x1 -1.183333  0.07142857
```

Dakle, dobili smo parametre apriorne distribucije od B.

Odredimo parametre za $Inv\chi^2$ apriornu distribuciju prema formuli (2.13):

```
> ni_0
[1] 13
> c2_0
[1] 16.33246
```

Znači, $\Sigma^2 \sim Inv\chi^2(13, 16.33)$.

Znamo da je aposteriorna distribucija od B normalna, sad odredimo njene parametre po formuli (2.14). Dobijemo redom aposteriornu sredinu te kovarijacijsku matricu od β:

```
> aposterior_sredina
        [,1]
   22.848982
x1  3.410323

> aposter_kovari
                           x1
    9.8997381 -0.59166667
x1 -0.5916667  0.03571429
```

Vidimo da je aposteriorna sredina $\beta_0 = 22.848982$, a $\beta_1 = 3.410323$, što se poklapa sa apriornim sredinama. Aposteriorna distribucija od Σ^2 je oblika (2.15) i njezini parametri su izračunati po formulama (2.16) redom:

```
> ni
[1] 28

>K<-2
>pom_c1<-(n-K)*sigma2_kapica
>pom_c2<-0
>pom_c3<-ni_0*c2_0
>c2<-(pom_c1+pom_c3)/ni
> c2
[1] 15.16586
```

Dobili smo da je aposteriorna distribucija od Σ^2 $Inv\chi^2(28, 15.16586)$.

Točkovni procjenitelji od Σ^2 se dobiju integrirajući funkciju gustoće $Inv\chi^2$-distribucije (vidjeti dodatak A za funkciju gustoće).

```
>mn.scIX2.sqrt<-function(x,n,t2)
{...}
```

Kako smo već spomenuli, marginalna distribucija od vektora B je multivarijatna Studentova t-distribucija sa parametrom β_a kojeg smo dobili MNK i $\sigma^2 A$. Za izračunati procjenitelje za svaku komponentu vektora B trebamo integirati zajedničku aposterironu distribuciju. Osnovno svojstvo multivarijatne Studentove t-distribucije je da su joj marginalne distribucije univarijatne Studentove t-distribucije, pa za svaki parametar vrijedi formula (2.17). U sljedećoj funkciji računam točkovne procjenitelje i kvantile aposteriorih distribucija, odnosno 95% pozdane intervale. Vrijedi da za $\alpha \in <0, 1>$, $100(1-\alpha)\%$ dvostrani bayesovski pouzdani interval odgovara $(\alpha/2)$ (u oznaci: $\theta_{\alpha/2}$) i $(1-\alpha/2)$ (u oznaci: $\theta_{1-\alpha/2}$) kvantilu aposteriorne distribucije parametra Θ.

```
bayesfitAnal<-function(lmfit){...}
```

Dobijem:

```
> lmfit<-lm(y1~x1)
> bayesfitAnal(lmfit)
                 coef         se        t    median  CrI.2.5. CrI.97.5.
(Intercept) 22.848982 11.4133221 2.001957 22.848982 -1.808001 47.505966
x1           3.410323  0.6855217 4.974785  3.410323  1.929343  4.891303
sigma        4.294814  0.9255154 4.640456  4.148490  2.930164  6.509538
```

Dakle, $\hat{\beta_0} = 22.848982, \hat{\beta_1} = 3.410323$ i $\hat{\sigma}^2 = 4.294814$, a 95% pouzdani interval su:

- Za β_0: $[-1.808001, 47.505966]$

- Za β_1: $[1.929343, 4.891303]$

- Za σ^2: $[2.930164, 6.509538]$

Cijeli kod se nalazi u dodatku B.

2.2 Multivarijatni linearni model

Vrlo često se susrećemo, posebno u financijama, s modeliranjem podataka koji se sastoje od imovine čiji povrati ili neki drugi atributi nisu nezavisni. Glavni problem multivarijatnog okvira je modeliranje zavisnosti među imovinama. U ovom ćemo poglavlju navesti osnovne multivarijatne regresijske procjene.

Pretpostavimo da imamo N zavisnih varijabli i poznato nam je T opservacija. To zapisujemo u $T \times N$ matricu koju ćemo označiti sa $\mathbf{Y}$,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_t \\ \vdots \\ \mathbf{Y}_T \end{bmatrix} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,N} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{t,1} & Y_{t,2} & \dots & Y_{t,N} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{T,1} & Y_{T,2} & \dots & Y_{T,N} \end{bmatrix}.$$

Multivarijatna regresija je zapisana kao:

$$\mathbf{Y} = \mathbf{X}B + \mathbf{E}, \tag{2.22}$$

gdje je $\mathbf{X} = T \times K$ matrica opservacija od K zavisnih varijabli,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_T \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,N} \\ x_{2,1} & x_{2,2} & \dots & x_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{t,1} & x_{t,2} & \dots & x_{t,N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T,1} & x_{T,2} & \dots & x_{T,N} \end{bmatrix},$$

$\mathbf{B} = K \times N$ matrica regresijskih koeficijenata,

$$\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} \beta_{0,1} & \beta_{0,2} & \dots & \beta_{0,N} \\ \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,N} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{K,1} & \beta_{K,2} & \dots & \beta_{K,N} \end{bmatrix},$$

$\mathbf{E} = T \times N$ matrica regresijskog šuma,

$$\mathbf{E} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_t \\ \vdots \\ \epsilon_T \end{bmatrix} = \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \dots & \epsilon_{1,N} \\ \epsilon_{2,1} & \epsilon_{2,2} & \dots & \epsilon_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{t,1} & \epsilon_{t,2} & \dots & \epsilon_{t,N} \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{T,1} & \epsilon_{T,2} & \dots & \epsilon_{T,N} \end{bmatrix}.$$

Prvi stupac od matrice $\mathbf{X}$ obično sadržava jedinice koje predstavljaju prisutnost slobodnog člana. U multivarijatnom slučaju, uobičajna pretpostavka linerane regresije da su pogreške nezavisno jednako distribuirane što znači da je svaki redak od matrice **E** nezavisna realizacija iz iste $N-$dimenzionalne multivarijatne distribucije. Pretpostavljamo da je ta distribucija multivarijatna normalna sa sredinom 0 i kovarijacijskom matricom Σ,

$$\epsilon_t \sim N(0, \Sigma),$$

za $t = 1, 2, \ldots, T$. Elementi matrice Σ koji nisu na dijagonali su općenito različiti od 0, budući da pretpostavljamo da su zavisne varijable u korelaciji i da kovarijacijska matrica sadržava N varijanci i $N(N-1)/2$ različitih kovarijanci.

Koristeći izraz za funkciju gustoće multivarijatne normalne distribucije (u dodatku), pišemo funkciju vjerodostojnosti za nepoznate parametre modela, **B** i $\mathbf{\Sigma}$, kao:

$$L(B, \Sigma | Y, \mathbf{X}) \propto |\Sigma|^{-T/2} exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} (Y_t - \mathbf{x}_t B) \Sigma^{-1} (Y_t - \mathbf{x}_t B)^t \right\},$$

gdje je sa $|\Sigma|$ označena determinanta kovarijacijske matrice. Sad ćemo odrediti apriornu distribucijsku pretpostavku za **B** i $\mathbf{\Sigma}$.

Nepravilana apriorna distribucija

Manjak apriornih informacija o **B** i $\mathbf{\Sigma}$ može se odraziti korištenjem Jeffreyeve apriorne distribucije, koja u multivarijatnom slučaju izgleda

$$\pi(B, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}.$$

Aposteriorna distribucija je kao i ona u univarijatnom slučaju. Primjetimo da je **B** slučajna matrica, premda, njena aposteriorna distribucija, uvjetno na $\mathbf{\Sigma} = \Sigma$, bit će generalizacija multivarijatne normalne aposteriorne distribucije u (2.9). Za opisati to, prvo ćemo zapisati vektor β, dimenzije $KN \times 1$, koji je nastao od matrice regresijeskih koeficijenata **B** tako da smo slagali okomito stupce od B^t

$$\beta = vec(B) = \begin{bmatrix} {\beta_0}^t \\ {\beta_1}^t \\ {\beta_2}^t \\ \vdots \\ {\beta_K}^t \end{bmatrix} \tag{2.23}$$

Može se pokazati da je aposteriorna distribucija od slučajnog vektora $\boldsymbol{\beta} = vec(\mathbf{B})$, uvjetno na $\mathbf{\Sigma} = \Sigma$, multivarijatna normala, dana sa:

$$\pi(\beta | \mathbf{y}, \mathbf{X}, \mathbf{\Sigma}) = N(\hat{\beta}, \Sigma \otimes (\mathbf{X}^t \mathbf{X}^{-1})^{-1}),$$

gdje je $\hat{\beta} = vec((\mathbf{X}^t\mathbf{X}^{-1})^{-1}(\mathbf{X}^t\mathbf{y}))$ procjenitelj od $\mathbf{B}$ dobiven MNK zapisan u vektorskom obliku, a $\otimes$ označava Kroneckerov produkt.

Može se pokazati da je aposteriorna distribucija od Σ inverzna Wisharova distribucija

$$\pi(\Sigma|\mathbf{Y}, \mathbf{X}) \sim InvW(\nu^*, S),$$

gdje su stupnjevi slobode označeni sa $\nu^* = T - K + N + 1$ i uzoračka kovarijacijska matrica je $S = (\mathbf{y} - \mathbf{X}\hat{B})^t(\mathbf{y} - \mathbf{X}\hat{B})$.

Potpuni bayesovski informativni apriorni pristup procjeni multivarijatnog linearnog regresijskog modela bi zahtjevao određivanje odgovarajuće apriorne distribucije za regresijske koeficijente $\boldsymbol{\beta}$ i kovarijacijske matrice $\boldsymbol{\Sigma}$. U slučaju konjugirane apriorne distribucije moramo pripaziti da regresijske parametre držimo unutar izvodivih analitičkih granica. U tom slučaju imamo multivarijatnu normalnu distribuciju za $\boldsymbol{B}$ i inverznu Wishartovu za $\boldsymbol{\Sigma}$.

Dodatak A

Uobičajne funkcije distribucije

U ovom poglavlju ćemo navesti neke uobičajne funkcije distribucije, njihove gustoće, te prvi i drugi moment tih distribucija. Najčešće distribucije diskretne slučajne varijable su binomna i poissonova, a najčešće distribucije neprekidne slučajne varijable su: uniformna, normalna, gama, beta, studentova t-distribucija, fisherova f-distribucija, χ^2-distribucija itd. Navest ćemo samo neke od njih.

A.1 Normalna distribucija

Univarijatna normalna distribucija, $N(\mu, \sigma^2)$

Neka su $\mu, \sigma \in \mathbb{R}, \sigma > 0$. Neprekidna slučajna varijabla X ima normalnu distribuciju s parametrima μ, σ^2 ako joj je gustoća f dana sa

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

- Oznaka: $X \sim N(\mu, \sigma^2)$.

- Očekivanje od X: $\mathbb{E}_{\mu,\sigma^2}(X) = \mu$.

- Varijanca od X: $Var_{\mu,\sigma^2}(X) = \sigma^2$.

Multivarijatna normalna distribucija, $N_p(\theta, \Sigma)$

Neka je $\theta \in \mathbb{R}^p$ i Σ $p \times p$ simetrična, pozitivno-definitna matrica. Za neprekidni p - dimenzionalni slučajni vektor X s vrijednostima u $\mathbb{R}^p$ kažemo da je normalno distribuiran

$N_p(\theta, \Sigma)$ ako vrijedi:

$$f(x|\theta, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(x-\theta)^t\Sigma^{-1}(x-\theta)\right\}.$$

- Oznaka: $X \sim N_p(\theta, \Sigma)$.

- Očekivanje od X: $\mathbb{E}_{\theta,\Sigma}(X) = \theta$.

- Kovarijacijska matrica od X: $Var_{\theta,\Sigma}(X) = \Sigma$.

Za $p = 1$, *log-normalna* distribucija je definirana kao distribucija slučajne varijable e^X kada je $X \sim N(\theta, \sigma^2)$.

A.2 Gama distribucija, $\Gamma(\alpha, \beta)$

Neka su $\alpha, \beta > 0$. Za neprekidnu slučajnu varijablu X s vrijednostima u $\mathbb{R}$, kažemo da ima gamma distribuciju $\Gamma(\alpha, \beta)$ ako vrijedi:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{[0,+\infty>}(x),$$

gdje je $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$ gama funkcija.

- Oznaka: $X \sim \Gamma(\alpha, \beta)$.

- Očekivanje od X: $\mathbb{E}_{\alpha,\beta}(X) = \frac{\alpha}{\beta}$.

- Varijanca od X: $Var_{\alpha,\beta}(X) = \frac{\alpha}{\beta^2}$.

Posebni slučaj gama distribucije su *Erlangova distribucija*, $\Gamma(\alpha, 1)$, *eksponencijalna distribucija*, $\Gamma(1, \beta)$, u oznaci $Exp(\beta)$ i χ^2−distribucija, $\Gamma(\nu/2, 1/2)$, u oznaci χ^2.

A.3 Beta distribucija, $Beta(\alpha,\beta)$

Neka su $\alpha,\beta > 0$. Za neprekidnu slučajnu varijablu X s vrijednostima u $\mathbb{R}$, kažemo da ima beta distribuciju $Beta(\alpha,\beta)$ ako vrijedi:

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \mathbb{1}_{[0,1]}(x),$$

gdje je

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

- Oznaka: $X \sim Beta(\alpha,\beta)$.

- Očekivanje od X: $\mathbb{E}_{\alpha,\beta}(X) = \frac{\alpha}{\alpha+\beta}$.

- Varijanca od X: $Var_{\alpha,\beta}(X) = \frac{\alpha\beta}{[(\alpha+\beta)^2(\alpha+\beta+1)]}$.

Beta distibucija se može dobiti kao distribucija $Y_1/(Y_1+Y_2)$ kada je $Y_1 \sim \Gamma(\alpha,1)$ i $Y_2 \sim \Gamma(\beta,1)$.

A.4 Studentova t-distribucija, $t_p(\nu,\theta,\Sigma)$

Neka je $\nu > 0$, $\theta \in \mathbb{R}^p$ i neka je Σ $(p \times p)$ simetrična pozitivno-definitna matrica. Za neprekidni p -dimenzionalni slučajni vektor X s vrijednostima u $\mathbb{R}^p$ kažemo da ima Studentovu distribuciju $t_p(\nu,\theta,\Sigma)$ ako vrijedi:

$$f(x|\nu,\theta,\Sigma) = \frac{\Gamma((\nu+p)/2)/\Gamma(\nu/2)}{|\Sigma|^{1/2}(\nu\pi)^{p/2}} \left(1 + \frac{(x-\theta)^t\Sigma^{-1}(x-\theta)}{\nu}\right)^{-(\nu+p)/2}.$$

- Oznaka: $X \sim t_p(\nu,\theta,\Sigma)$.

- Očekivanje od X: $\mathbb{E}_{\nu,\theta,\Sigma}(X) = \theta$, kada je $\nu > 1$.

- Kovarijacijska matrica od X: $Var_{\nu,\theta,\Sigma}(X) = \nu\Sigma/(\nu-2)$, za $\nu > 2$.

Poseban slučaj Studentove t-distribucije je kada je $p = 1$. Tada imamo *Cauchyjevu distribuciju*, $C(\theta,\sigma^2)$, što odgovara $\nu = 1$.

A.5 Fisherova F-distribucija, $F(\nu,\rho)$

Neka su $\nu,\rho > 0$. Za neprekidnu slučajnu varijablu X s vrijednostima u $\mathbb{R}$, kažemo da ima Fisherovu F-distribuciju $F(\nu,\rho)$ ako vrijedi:

$$f(x|\nu,\rho) = \frac{\Gamma((\nu+\rho)/2)\nu^{\rho/2}\rho^{\nu/2}}{\Gamma(\nu/2)\Gamma(\rho/2)} \frac{x^{(\nu-2)/2}}{(\nu+\rho x)^{(\nu+\rho)/2}} \mathbb{1}_{[0,+\infty>}(x).$$

- Oznaka: $X \sim F(\nu,\rho)$.
- Očekivanje od X: $\mathbb{E}_{\nu,\rho}(X) = \rho/(\rho-2)$, kada je $\rho > 2$.
- Varijanca od X: $Var_{\nu,\rho}(X) = 2\rho^2(\nu+\rho-2)/[\nu(\rho-4)(\rho-2)^2]$, za $\rho > 4$.

A.6 Inverzna gama distribucija, $Inv\Gamma(\alpha,\beta)$

Neka su $\alpha,\beta > 0$. Za neprekidnu slučajnu varijablu X s vrijednostima u $\mathbb{R}$, kažemo da ima inverznu gama distribuciju $Inv\Gamma(\alpha,\beta)$ ako vrijedi:

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{e^{-\beta/x}}{x^{\alpha+1}} \mathbb{1}_{[0,+\infty>}(x).$$

- Oznaka: $X \sim Inv\Gamma(\alpha,\beta)$.
- Očekivanje od X: $\mathbb{E}_{\alpha,\beta}(X) = \frac{\beta}{\alpha-1}$, za $\alpha > 1$.
- Varijanca od X: $Var_{\alpha,\beta}(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, za $\alpha > 2$.

Ova distribucija je zapravo distribucija od X^{-1} kada je $X \sim \Gamma(\alpha,\beta)$.

A.7 Inverzna χ^2-distribucija, $Inv\chi^2(\nu,c)$

Neka su $\nu > 0$, $c > 0$. Za neprekidnu slučajnu varijablu X s vrijednostima u $\mathbb{N}$, kažemo da ima inverznu χ^2-distribuciju $Inv\chi^2(\nu,c)$ ako vrijedi:

$$f(x|\nu,c) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{\nu/2} c^\nu x^{-\left(\frac{\nu}{2}+1\right)} exp\left(-\frac{\nu c}{2x}\right).$$

- Oznaka: $X \sim Inv\chi^2(\nu, c)$.
- Očekivanje od X: $\mathbb{E}_{\nu,c}(X) = \frac{\nu}{\nu-2}c$, za $\nu > 2$.
- Varijanca od X: $Var_{\nu,c}(X) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}c^2$, za $\nu > 4$.

A.8 Binomna distribucija, $\mathcal{B}(n, p)$

Neka je $0 \leq p \leq 1$. Za diskretnu slučajnu varijablu X s vrijednostima u $\mathbb{N}_0$, kažemo da ima binomnu distribuciju $\mathcal{B}(n, p)$ ako vrijedi:

$$f(x|p) = \binom{n}{k} p^x (1-p)^{n-x} \mathbb{1}_{\{0,1,\ldots,n\}}(x).$$

- Oznaka: $X \sim \mathcal{B}(n, p)$.
- Očekivanje od X: $\mathbb{E}_p(X) = np$.
- Varijanca od X: $Var_p(X) = np(1-p)$.

A.9 Negativna binomna distribucija, $Neg\mathcal{B}(n, p)$

Neka je $0 \leq p \leq 1$. Za diskretnu slučajnu varijablu X s vrijednostima u $\mathbb{N}$, kažemo da ima negativnu binomnu distribuciju $Neg\mathcal{B}(n, p)$ ako vrijedi:

$$f(x|p) = \binom{n+x+1}{x} p^n (1-p)^x \mathbb{1}_{\mathbb{N}}(x).$$

- Oznaka: $X \sim Neg\mathcal{B}(n, p)$.
- Očekivanje od X: $\mathbb{E}_p(X) = n(1-p)/p$.
- Varijanca od X: $Var_p(X) = n(1-p)/p^2$.

A.10 Poissonova distribucija, $\mathcal{P}(\lambda)$

Neka je $\lambda > 0$. Za diskretnu slučajnu varijablu X s vrijednostima u $\mathbb{N}_0$, kažemo da ima Poissonovu distribuciju $\mathcal{P}(\lambda)$ ako vrijedi:

$$f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}\mathbb{1}_{\mathbb{N}_0}(x).$$

- Oznaka: $X \sim P(\lambda)$.

- Očekivanje od X: $\mathbb{E}_\lambda(X) = \lambda$.

- Varijanca od X: $Var_\lambda(X) = \lambda$.

Dodatak B

Programski kod primjera u R-u

```
## integriranje
##chisq2 sa df=n, skalarni parametar=t2
mn.scIX2.sqrt<-function(x,n,t2)
{
    s2<-x^2
    n.2<-n/2.0
    lx<-log(x)
    2.0*x*x*exp(-n*t2/(2.0*s2)-lgamma(n.2)+
    n.2*(log(t2)+log(n.2))-(n+2)*lx)
}

bayesfitAnal<-function(lmfit){
    ## izvlacim koeficijente, stupnjeve slobode i varijancu
    ## matrica iz lm-a
        QR<-lmfit$qr
        df<-lmfit$df
        R<-qr.R(QR) ## R komponenta
        coef<-lmfit$coef
        Vb<-chol2inv(R) ## varijanca
        s2<-(t(lmfit$residuals)%*%lmfit$residuals)
        s2<-s2[1,1]/df
        scale<-sqrt(s2*diag(Vb))

    ## standardne greške za univarijatnu t
    se=scale*ifelse(df>2,
    sqrt(df/(df-2)),
```

```
    ifelse(df<1,NA,Inf))
    ## dataframe za dobivene vrijednosti
        ret<-data.frame(coef=coef,se=se,t=coef/se,
        mode=coef,median=coef,
        "CrI.2.5%"=qt(0.025,df=df),
        "CrI.97.5%"=qt(0.975,df=df))
    ## intervali pouzdanosti za t-distribuciju
    ret[,6:7]<-ret[,6:7]*se+coef

    ## sigma
    ret<-rbind(ret,rep(0,7))
    rownames(ret)[3]<-"sigma"
    ## sredina
    M1<-integrate(mn.scIX2.sqrt,n=df,
        t2=s2,lower=0,upper=Inf)$val
    S1<-ifelse(df<=2,NA,
        df*s2/(df-2))
    ret[3,1]<-M1 ## sredina
    ret[3,2]<-sqrt(S1-M1^2)  ## sd
    ret[3,3]<-ret[3,1]/ret[3,2]  ## t
    ret[3,4]<-sqrt(s2*df/(df+1)) ## mode
    ## računamo kvantile
    ret[3,5]<-uniroot(function(x) pgamma(x,
        shape=df/2,scale=2/(s2*df),
        lower.tail=FALSE)-0.5,
        lower=0,upper=1/s2)$root
    ## donja granica od 95% p.i
    ret[3,6]<-uniroot(function(x) pgamma(x,
        shape=df/2,scale=2/(s2*df),
        lower.tail=FALSE)-0.025,lower=0,
        upper=1/(M1-3*ret[3,2])^2)$root
    ## gornja granica od 95% p.i
    ret[3,7]<-uniroot(function(x) pgamma(x,
        shape=df/2,scale=2/(s2*df),
        lower.tail=FALSE)-0.975,lower=0,
        upper=1/s2)$root
    ret[3,5:7]<-sqrt(1/ret[3,5:7])
    ret
}
```

Bibliografija

[1] B. Basrak, *Aktuarska matematika II*, dostupno na http://web.math.pmf.unizg.hr/~bbasrak/pdf_files/AM2slides2dio.pdf, (lipanj 2015.).

[2] M. Huzak, *Matematička statistika*, PMF-MO predavanja, 2013.

[3] S.T. Rachev et al., *Bayesian Methods in Finance*, Wiley, 2008.

[4] C.P. Robert, *The Bayesian Choice, A Decision-Theoretic Motivation*, Springer, 1994.

[5] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 2002.

[6] M.J. Schervish, *Theory of Statistics*, Springer-Verlag, 1995.

Sažetak

U ovom diplomskom radu smo se bavili bayesovskom statistikom odnosno bayesovskim linearnim regresijskim modelima. Ukratko, u linearnoj regresiji pokušavamo objasniti varijabilnost zavisne varijable Y pomoću jedne ili više nezavisnih varijabli X. Koristimo pretpostavku da je regresijski šum nezavisno normalno distribuiran, zavisna varijabla Y također ima normalnu distribuciju, a $\mathbf{X}$ je matrica dizajna, prvi stupac sadrži jedinice. Sad kad je model postavljen, bayesovska analiza traži aposteriornu distribuciju parametara i prediktivnu distribuciju za model. Analiza započinje apriornom distribucijom. Parametre (β, σ^2) označimo sa (B, Σ^2) kao slučajne veličine.

U radu smo koristili pretpostavku da parametar (B, Σ^2) ima nepravilnu apriornu distribuciju $\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$, te dobili da je aposteriorna distribucija za β uvjetno na $\Sigma^2 = \sigma^2$ normalna, a aposteriorna distribucija za Σ^2 $Inv\chi^2$-distribucija.

Drugi slučaj koji smo pokazali je slučaj informativne apriorne distribucije. Pretpostavili smo da B ima normalnu apriornu distribuciju (uvjetnu na σ^2), a Σ^2 inverznu χ^2- apriornu distribuciju. Dobili smo da je i aposteriorna distribucija za B dana sa $\pi(\beta|\mathbf{y}, \mathbf{X}, \sigma^2)$ također normalna, a aposteriorna distribucija od Σ^2 ponovno inverzna χ^2 distribucija.

Pokazali smo i slučaj nejednakih varijanci, kad imamo dva modela, $\mathbf{Y_1} = \mathbf{X_1}\beta + \epsilon_\mathbf{1}$ i $\mathbf{Y_2} = \mathbf{X_2}\beta + \epsilon_\mathbf{2}$. Neinformativna apriorna distribucija se može utvrditi, pod pretpostavkom da su parametri nezavisni. Tada je apriorna distribucija zapisana kao: $\pi(\beta, \sigma_1, \sigma_2) \propto \frac{1}{\sigma_1 \sigma_2}$. Pokazuje se da je marginalna aposteriorna distribucija od B produkt dvije multivarijatne Studentove t-gustoće. Također je pokazano da se marginalna aposteriorna distribucija od B može aproksimirati sa normalnom distribucijom.

Slično smo pokazali i za multivarijatni slučaj.

Summary

This thesis is concerned with a Bayesian statistics and Bayesian linear regression models. In short, in linear regression we are trying to explain the variability in dependent variable Y with the help of one or more independent variables X. We assume that regression disturbances are independently and identically disturbed with normal distribution, the dependent variable Y has also normal distribution and $\mathbf{X}$ is design matrix, first column of $\mathbf{X}$ is a column of ones. Once the model is specified, the Bayesian analysis seeks the posterior distribution of the parameters and a predictive distribution for the model's prediction. The analysis begins with a prior distribution. Parameters (β, σ^2) are denoted with (B, Σ^2) like random variables.

We used assumption that parameter (B, Σ^2), has improper prior $\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$, it follows that posterior distribution of β conditional on $\Sigma^2 = \sigma^2$ is normal, and posterior distribution of Σ^2 is $Inv\chi^2$ distribution.

Another case, which we have shown is case of informative prior distribution. We assume that B has a normal prior distribution (conditional on σ^2), and Σ^2 inverted χ^2- prior distribution. It follows that posterior distribution for B is given with: $\pi(\beta|\mathbf{y}, \mathbf{X}, \sigma^2)$ is also normal and posterior distribution of Σ^2 is also inverted χ^2 distribution.

We also have shown the case of unequal variance, when we have two models,
$\mathbf{Y_1} = \mathbf{X_1}\beta + \epsilon_1$ and $\mathbf{Y_2} = \mathbf{X_2}\beta + \epsilon_2$. A non-informative prior distribution can be asserted, by assuming that the parameters are independent. The prior is written, then, as: $\pi(\beta, \sigma_1, \sigma_2) \propto \frac{1}{\sigma_1\sigma_2}$. It is shown that the marginal posterior distribution of B is the product of two multivariate Student's t-densities. Also, it is shown that the marginal posterior of B can be approximated with a normal distribution.

Similarly, we have shown for the multivariate case.

Životopis

- Rođena sam 17. svibnja 1989. godine u Zadru
- Od 1995. do 2003. godine pohađam osnovnu školu u Biogradu na Moru
- Od 2003. do 2007. godine pohađam opću gimnaziju u Biogradu na Moru
- Od 2007. do 2013. godine pohađam preddiplomski studij Matematika na PMF-MO u Zagrebu
- 2013. godine upisujem diplomski sveučilišni studij Matematička statistika na PMF-MO u Zagrebu