Generalizirani linearni modeli

Andrijević, Antonija

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Antonija Andrijević

GENERALIZIRANI LINEARNI MODELI

Diplomski rad

Voditelj rada:
prof. dr. sc. Miljenko Huzak

Zagreb, rujan 2024.

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. _____

Ovaj rad posvećujem svojim roditeljima, koji su mi svojom ljubavlju, strpljenjem i razumijevanjem bili najveća podrška tokom cijelog mog školovanja. Njihova neizmjerna podrška mi je bila oslonac u trenucima kada mi je bilo najteže.

Posebnu zahvalnost dugujem i svom mentoru prof. dr. sc. Miljenku Huzaku, koji me je svojim znanjem, savjetima i nesebičnom pomoći usmjeravao i motivirao kroz cijeli proces izrade ovog rada.

Također, želim se zahvaliti svim prijateljima, kolegama i profesorima koji su mi na bilo koji način olakšali put do ovog trenutka. Njihova podrška i ohrabrenje bili su od neprocjenjive vrijednosti.

Sadržaj

Uvod

U svakodnevnom nas životu često zanima kako različiti čimbenici utječu na neki događaj od interesa. Primjerice, može nas zanimati kako naše životne navike, poput načina prehrane, pušenja, količine sna i tjelovježbe utječu na naš godišnji broj posjeta liječniku. Kada bismo to znali, mogli bismo donijeti neke odluke o poboljšanju svojih životnih navika kako bismo smanjili broj posjeta liječniku. Također, financijske institucije, poput banaka, žele biti u mogućnosti procijeniti rizik vezan uz svakog svog klijenta, tj. procijeniti kolika je vjerojatnost da klijent neće moći otplatiti kredit. U skladu s tom vjerojatnosti, banka se može odlučiti ne odobriti kredit klijentu koji ima visok rizik te samim time izbjeći gubitke. U sportu, trenere i sportske analitičare obično zanima kako različite performanse igrača te stil igre utječu na konačan ishod utakmice. Rezultati tih analiza mogu pomoći trenerima da donesu odluke kojima će poboljšati uspjeh svog tima u utakmicama. Svi ovi, ali i mnogi drugi primjeri pomažu nam u boljem razumijevanju kako različiti čimbenici utječu na naš svakodnevni život te u lakšem donošenju odluka. Metoda u statistici koja nam omogućava analizu utjecaja raznih faktora (tzv. **nezavisnih varijabli**) na neki događaj od interesa (tzv. **zavisnu varijablu**) naziva se **regresijska analiza**. Regresijska se analiza opisuje matematičkom jednadžbom koja kvantificira povezanost između nezavisnih i zavisne varijable.

Najjednostavniji oblik kvantifikacije u regresijskoj analizi je linearan oblik, koji pretpostavlja da postoji linearna veza između nezavisnih i zavisne varijable. Dodatne pretpostavke na takav model uključuju, između ostalog, normalnost zavisne varijable. Takvu vrstu modela nazivamo **normalni linearni regresijski model**. Iako su ovi modeli korisni zbog svoje jednostavnosti i lakoće u interpretaciji rezultata, njihova primjena i preciznost može biti ograničena u slučaju kada radimo s podacima koji nisu normalno distribuirani ili su diskretni, kao što su to gore navedeni primjeri, te kada veza između nezavisnih i zavisne varijable nije linearna. Iz ovoga je razloga postojala potreba za proširenjem teorije normalnih linearnih modela na općenitije **generalizirane linearne modele** koji uključuju različite tipove zavisne varijable te različite veze između nezavisnih i zavisne varijable. Koncept generaliziranih linearnih modela prvi se put formalno opisuje u radu Johna Neldera i Roberta Wedderburna iz 1972. godine.

U ovome radu kroz četiri poglavlja opisujemo ključne koncepte i primjenu generalizirnih linearniih modela u statistici. U prvom ćemo se poglavlju upoznati s osnovnim sta-

tističkim pojmovima i pružiti kratak pregled regresijske analize s posebnim naglaskom na normalne linearne modele i njihova ograničenja. U drugom ćemo poglavlju uvesti pojam generaliziranih linearnih modela koji čine proširenje teorije normalnih linearnih modela te opisati njihove ključne komponente: zavisna varijabla dolazi iz šire klase distribucija koju nazivamo **eksponencijalna familija distribucija**, **linearni predviditelj** opisuje kako prediktori utječu na distribuciju zavisne varijable te **funkcija poveznica** koja povezuje linearni predviditelj i očekivanu vrijednost zavisne varijable. U trećem ćemo poglavlju koristiti metodu najveće vjerodostojnosti za procjenu parametara modela pomoću iterativne težinske metode najmanjih kvadrata. Konačno, u četvrtom ćemo se poglavlju baviti statističkim inferencijama kojima provjeravamo valjanost našega modela kako bismo se uvjerili da su procjene koje nam model daje dovoljno precizne te da ih možemo koristiti za donošenje odluka.

Poglavlje 1

Motivacija

U ovom ćemo se poglavlju posvetiti definiranju osnovnih statističkih pojmova koji su neophodni za dublje razumijevanje i izgradnju teorije generaliziranih linearnih modela. Uz to, predstavit ćemo najpoznatije distribucije i njihova svojstva te pružiti kratak pregled regresijske analize s posebnim naglaskom na normalne linearne modele. Ideja je ilustrirati kako ograničenja normalnih linearnih modela stvaraju potrebu za proširenjem teorije na generalizirane linearne modele.

1.1 Slučajne veličine

U području statistike i teorijske matematike, ključno je razumjeti slučajne veličine i njihova svojstva kako bismo mogli analizirati podatke i modelirati razne fenomene. *Slučajne veličine* predstavljaju osnovnu ideju koja nam pomaže u kvantifikaciji nesigurnosti i varijabilnosti u našim opažanjima. Tvrdnje navedene u ovom odjeljku prate [5], [6] i [9].

Definicija 1.1.1. *Neka je* $(\Omega, \mathcal{F}, \mathbb{P})$ *vjerojatnosni prostor. Funkcija* $Y : \Omega \to \mathbb{R}^n$ *je* n-**dimenzionalna slučajna veličina** *na* Ω *ako vrijedi*

$$(\forall y \in \mathbb{R}^n)\{Y \leq y\} = \{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F}.$$

Ako je $n = 1$*, onda kažemo da je* Y **slučajna varijabla***, a za* $n \geq 2$ *kažemo da je* Y n-**dimenzionalni slučajni vektor** *(ili, kraće,* **slučajni vektor***).*
Nadalje, funkciju $F_Y : \mathbb{R}^n \to [0, 1]$ *definiranu s*

$$F_Y(y) = \mathbb{P}(Y \leq y), \quad y \in \mathbb{R}^n$$

zovemo **funkcija razdiobe** *(ili* **funkcija distribucije***) slučajne veličine* Y*.*

Da bismo izbjegli zabune oko notacije slučajnih varijabli i slučajnih vektora, u ovome ćemo radu slučajne varijable označavati s velikim tiskanim slovima, npr. Y, dok ćemo slučajne vektore označavati podebljanim velikim tiskanim slovima, npr. $\boldsymbol{Y}$. Također, poznat statistički rezultat kaže da su komponente n-dimenzionalnog slučajnog vektora slučajne varijable, tj. vrijedi sljedeća propozicija:

Propozicija 1.1.1. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor te $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T : \Omega \to \mathbb{R}^n$. Tada je $\boldsymbol{Y}$ n-dimenzionalni slučajni vektor ako i samo ako su $Y_i : \Omega \to \mathbb{R}$ slučajne varijable, za svaki $i = 1, \ldots, n$.*

Još jedan koristan statistički rezultat koji će nam koristiti kasnije kod eksponencijalne familije distribucija kaže da je kompozicija izmjerivog preslikavanja i slučajne veličine ponovno slučajna veličina, tj. vrijedi sljedeća propozicija:

Propozicija 1.1.2. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor, $Y : \Omega \to \mathbb{R}^n$ slučajna veličina i $g : \mathbb{R}^n \to \mathbb{R}^n$ Borelova funkcija. Tada je $g(Y) = g \circ Y : \Omega \to \mathbb{R}^n$ slučajna veličina.*

Slučajne se veličine klasificiraju u dvije glavne kategorije ovisno o tome poprimaju li beskonačno ili konačno mnogo vrijednosti na temelju sljedeće dvije definicije:

Definicija 1.1.2. *Slučajna veličina $Y : \Omega \to \mathbb{R}^n$ definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ je* **aposlutno neprekidna** *(ili, kraće,* **neprekidna***) slučajna veličina ako postoji Borelova funkcija $f_Y : \mathbb{R}^n \to [0, +\infty\rangle$ takva da za sve $y \in \mathbb{R}^n$ vrijedi*

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^{y} f_Y(t)\, dt,$$

gdje je $t \in \mathbb{R}^n$. Funkciju f_Y iz gornje jednadžbe zovemo **funkcija gustoće neprekidne slučajne veličine** *Y.*

Definicija 1.1.3. *Slučajna veličina $Y : \Omega \to \mathbb{R}^n$ definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ je* **diskretna** *slučajna veličina ako postoji prebrojiv skup $D \subseteq \mathbb{R}^n$ takav da je*

$$\mathbb{P}(Y \in D) = 1.$$

Funkcija gustoće diskretne slučajne veličine *je funkcija $f_Y : \mathbb{R}^n \to [0, 1]$ takva da za sve $y \in \mathbb{R}^n$ vrijedi*

$$f_Y(y) = \mathbb{P}(Y = y).$$

Iz svojstava funkcija razdiobe slijedi da funkcija gustoće slučajne veličine $Y : \Omega \to \mathbb{R}^n$ zadovoljava sljedeće jako važno svojstvo

$$\textit{(i)}\ \int_{\mathbb{R}^n} f_Y(t)\, dt = 1, \text{ ako je slučajna veličina } Y \text{ neprekidna}, \tag{1.1}$$

(ii) $\sum_{y\in D} f_Y(y) = 1$, ako je slučajna veličina Y diskretna. (1.2)

Pored razumijevanja pojma slučajnih veličina, važno je znati i nešto o njihovim očekivanjima. *Matematičko očekivanje*, ili očekivana vrijednost, predstavlja neku vrstu "prosječne" vrijednosti koju bismo mogli očekivati ako bismo isti eksperiment ponavljali mnogo puta.

Definicija 1.1.4. *Neka je* $Y : \Omega \to \mathbb{R}$ *slučajna varijabla definirana na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$. **Matematičko očekivanje slučajne varijable** *Y, u oznaci* $\mathbb{E}[Y]$,

(i) za neprekidnu slučajnu varijablu takvu da je $\int_{\mathbb{R}} |y| \cdot f_Y(y)\, dy < +\infty$ *definira se kao*

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y \cdot f_Y(y)\, dy, \tag{1.3}$$

(ii) za diskretnu slučajnu varijablu takvu da je $\sum_{y\in D} |y| \cdot f_Y(y) < +\infty$ *definira se kao*

$$\mathbb{E}[Y] = \sum_{y\in D} y \cdot f_Y(y).$$

Nadalje, za Borelovu funkciju $g : \mathbb{R} \to \mathbb{R}$ *je* **očekivanje slučajne varijable** *$g(Y)$, u oznaci* $\mathbb{E}[g(Y)]$,

(i) za neprekidnu slučajnu varijablu takvu da je $\int_{\mathbb{R}} |g(y)| \cdot f_Y(y)\, dy < +\infty$ *definirano kao*

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(y) \cdot f_Y(y)\, dy, \tag{1.4}$$

(ii) za diskretnu slučajnu varijablu takvu da je $\sum_{y\in D} |g(y)| \cdot f_Y(y) < +\infty$ *definirano kao*

$$\mathbb{E}[g(Y)] = \sum_{y\in D} g(y) \cdot f_Y(y).$$

Kako su komponente n-dimenzionalnog slučajnog vektora slučajne varijable potpuno je intuitivna i sljedeća definicija *očekivanja slučajnog vektora*:

Definicija 1.1.5. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T : \Omega \to \mathbb{R}^n$ *n-dimenzionalni slučajni vektor na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$. **Matematičko očekivanje slučajnog vektora *$\boldsymbol{Y}$*** *takvog da je* $\mathbb{E}[|Y_i|] < +\infty$ *za sve* $i = 1, \ldots, n$, *u oznaci* $\mathbb{E}[\boldsymbol{Y}]$, *definira se*

$$\mathbb{E}[\boldsymbol{Y}] = (\mathbb{E}[Y_1], \ldots, \mathbb{E}[Y_n])^T.$$

Sljedeća nam propozicija daje neka od najvažnijih svojstava matematičkog očekivanja koja ćemo koristiti u ovome radu.

Propozicija 1.1.3. *Neka su* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T : \Omega \to \mathbb{R}^n$ *n-dimenzionalan slučajni vektor definiran na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$ *takav da* $\mathbb{E}[|Y_i|] < +\infty$ *za sve* $i = 1, \ldots, n$. *Nadalje, neka je* $\boldsymbol{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ *konstantan vektor. Tada vrijedi*

(i) očekivanje konstante: $\mathbb{E}[\boldsymbol{a}] = \boldsymbol{a}$, (1.5)

(i) linearnost očekivanja: $\mathbb{E}[\boldsymbol{a}^T\boldsymbol{Y}] = \mathbb{E}\left[\sum_{i=1}^n a_i Y_i\right] = \sum_{i=1}^n a_i \mathbb{E}[Y_i] = \boldsymbol{a}^T\mathbb{E}[\boldsymbol{Y}]$, (1.6)

(iii) očekivanje produkta: $\mathbb{E}[Y_jY_k] = \mathbb{E}[Y_j]\mathbb{E}[Y_k]$ *za sve* $j \neq k$ *takve da su* Y_j *i* Y_k *nezavisne.* (1.7)

Također, uz pojam matematičkog očekivanja često vežemo pojam *varijance* i *standardne devijacije* slučajne veličine. Varijanca mjeri koliko se vrijednosti slučajne veličine u prosjeku razlikuju od očekivane vrijednosti te nam time pomaže razumjeti disperziju podataka. Standardna devijacija, koja je korijen varijance, daje intuitivniju sliku o rasprostranjenosti podataka jer je izražena u istoj jedinici kao i sama varijabla. *Kovarijanca*, s drugu stranu, mjeri kako se dvije slučajne veličine zajednički ponašaju u odnosu na svoje očekivane vrijednosti. Pozitivna kovarijanca sugerira da, kada jedna veličina raste, druga također ima tendenciju rasta, dok negativna kovarijanca implicira suprotnu težnju. Kovarijanca je temelj za razumijevanje odnosa između više slučajnih veličina i igra ključnu ulogu u regresijskoj analizi.

Definicija 1.1.6. *Neka su* $X, Y : \Omega \to \mathbb{R}$ *slučajne varijable definirane na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$ *takve da* $\mathbb{E}[X^2] < +\infty$ *i* $\mathbb{E}[Y^2] < +\infty$.

(i) **Varijanca slučajne varijable** *Y, u oznaci* Var*(Y), definira se kao*

$$\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2]. \tag{1.8}$$

(ii) **Standardna devijacija slučajne varijable** *Y, u oznaci* std(*Y*)*, definira se kao*

$$\mathrm{std}(Y) = \sqrt{\mathrm{Var}(Y)}.$$

(iii) **Kovarijanca slučajnih varijabli** *X i Y, u oznaci* Cov(X, Y)*, definira se kao*

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

U nastavku slijede najčešće korištena svojstva varijance i kovarijance koja će se koristiti u ovome radu.

Propozicija 1.1.4. *Neka su $X, Y : \Omega \to \mathbb{R}$ slučajne varijable definirane na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Nadalje, neka su $\mathbb{E}[X^2] < +\infty$ i $\mathbb{E}[Y^2] < +\infty$ te $a, b \in \mathbb{R}$. Tada vrijedi:*

(i) $$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \tag{1.9}$$

(ii) $$\mathrm{Var}(aY + b) = a^2\mathrm{Var}(Y) \tag{1.10}$$

(iii) $$\mathrm{Cov}(Y, Y) = \mathrm{Var}(Y) \tag{1.11}$$

(iv) $$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{1.12}$$

Definicija 1.1.7. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T : \Omega \to \mathbb{R}^n$ n-dimenzionalni slučajni vektor definiran na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ takav da $\mathbb{E}[Y_i^2] < +\infty$ za sve $i = 1, \ldots, n$.* **Kovarijacijska matrica slučajnog vektora $\boldsymbol{Y}$**, *u oznaci* $\mathrm{Cov}(\boldsymbol{Y})$, *je $n \times n$ matrica definirana kao*

$$\mathrm{Cov}(\boldsymbol{Y}) = \mathbb{E}\left[(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])^T\right]. \tag{1.13}$$

U nastavku su dana svojstva kovarijacijske matrice koja ćemo koristiti u ovome radu.

Propozicija 1.1.5. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T : \Omega \to \mathbb{R}^n$ n-dimenzionalni slučajni vektor definiran na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$, $\boldsymbol{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ konstantan vektor te $A = [A_{jk}]_{j,k=1,\ldots,n}$ konstantna $n \times n$ matrica. Tada vrijedi:*

(i) $$\mathrm{Cov}(\boldsymbol{Y} + \boldsymbol{a}) = \mathrm{Cov}(\boldsymbol{Y}) \tag{1.14}$$

(ii) $$\mathrm{Cov}(A\boldsymbol{Y}) = A\mathrm{Cov}(\boldsymbol{Y})A^T \tag{1.15}$$

1.2 Razdiobe slučajnih veličina

Fokusirajmo se na razdiobe slučajnih veličina, koje se opisuju funkcijom gustoće vjerojatnosti. Postoji mnogo različitih razdioba koje se koriste u statistici, no za potrebe ovog rada, dat ćemo kratak opis i najvažnija svojstva *Bernoullijeve*, *binomne* i *Poissonove razdiobe* koje su diskretne te *normalne*, *eksponencijalne* i *gama razdiobe* koje su neprekidne.

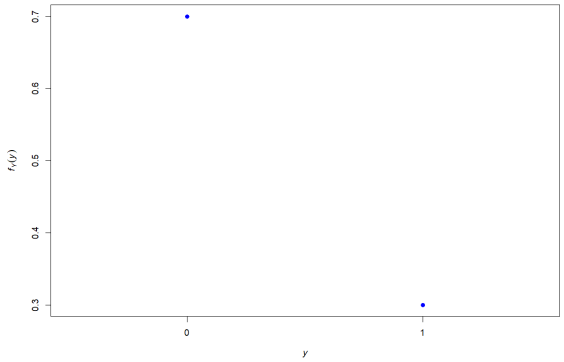
Bernoullijeva razdioba

Bernoullijeva distribucija je diskretna vjerojatnosna distribucija koja opisuje rezultat eksperimenta s dva moguća ishoda: "uspjeh", koji se događa s vjerojatnošću p, te "neuspjeh" koji se događa s vjerojatnošću $1 - p$.

Ova se distribucija koristi za modeliranje događaja s binarnim ishodom, kao što su bacanje simetričnog novčića, pitanja na koja se odgovara s "da" ili "ne", testiranje proizvoda na ispravnost (ispravan ili neispravan proizvod) i mnogih drugih.

Definicija 1.2.1. *Neka je Y diskretna slučajna varijabla definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Kažemo da slučajna varijabla Y ima* **Bernoullijevu distribuciju** *s parametrom $p \in \langle 0, 1\rangle$, u oznaci $Y \sim \text{Bernoulli}(p)$, ako joj je funkcija gustoće*

$$f_Y(y; p) = p^y(1-p)^{1-y}, \quad y \in \{0, 1\}. \tag{1.16}$$



Slika 1.1: Funkcija gustoće slučajne varijable $Y \sim \text{Bernoulli}(0.3)$

Propozicija 1.2.1. *Neka je $Y \sim \text{Bernoulli}(p)$, $p \in \langle 0, 1\rangle$. Tada vrijedi*

$$\mathbb{E}[Y] = p, \tag{1.17}$$

$$\text{Var}(Y) = p(1-p). \tag{1.18}$$

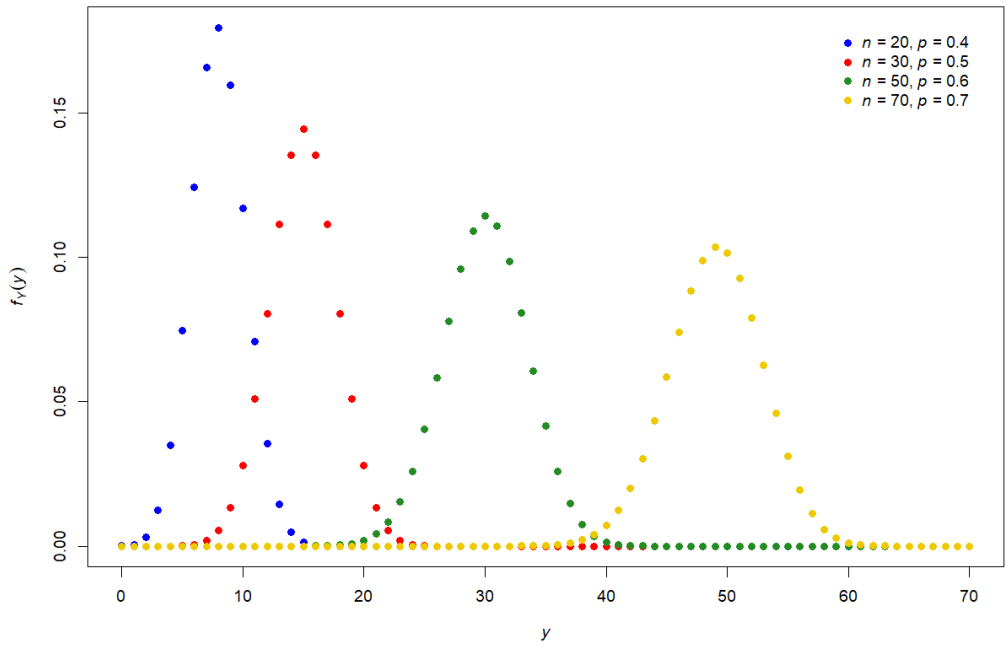
Binomna razdioba

Binomna distribucija je diskretna vjerojatnosna distribucija koja opisuje broj uspjeha u nizu od n nezavisnih Bernoullijevih eksperimenata s vjerojatnošću uspjeha p.

Ovu distribuciju koristimo za modeliranje situacija kao što su broj dobivenih "glava" u bacanju simetričnog novčića, broj točnih odgovora na testu s pitanjima tipa "točno/netočno", broj uspješnih prodaja u kampanji te drugih.

Definicija 1.2.2. *Neka je Y diskretna slučajna varijabla definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Kažemo da slučajna varijabla Y ima* **binomnu distribuciju** *s parametrima $n \in \mathbb{N}$ i $p \in \langle 0, 1\rangle$, u oznaci $Y \sim B(n, p)$, ako joj je funkcija gustoće*

$$f_Y(y; n, p) = \binom{n}{y} p^y(1-p)^{n-y}, \quad y = 0, 1, \ldots, n. \tag{1.19}$$



Slika 1.2: Funkcija gustoće slučajne varijable $Y \sim B(n, p)$

Propozicija 1.2.2. *Neka je* $Y \sim B(n, p), n \in \mathbb{N}, p \in \langle 0, 1 \rangle$. *Vrijedi*

$$\mathbb{E}[Y] = np, \tag{1.20}$$

$$\mathrm{Var}(Y) = np(1-p). \tag{1.21}$$

Poissonova razdioba

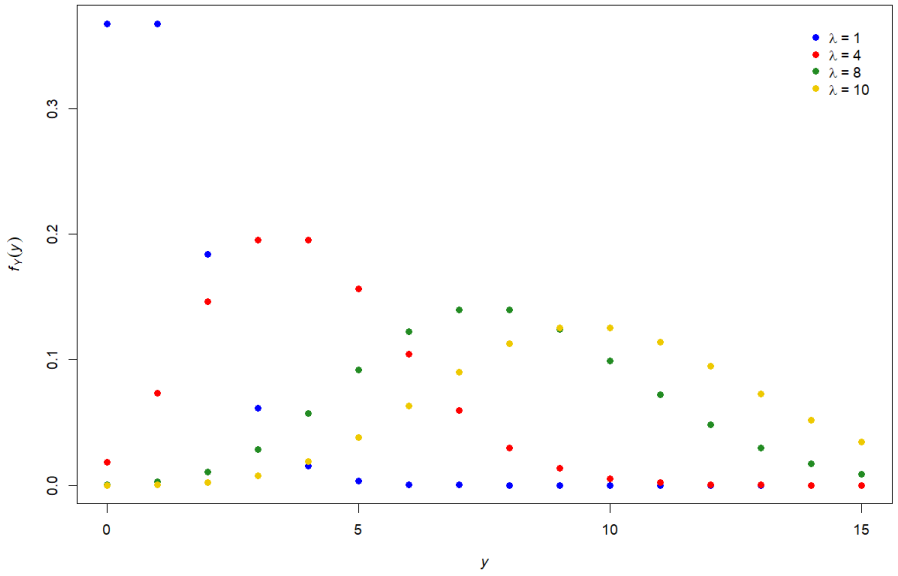
Poissonova distribucija je diskretna vjerojatnosna distribucija koja se najčešće koristi za modeliranje broja rijetkih i nezavisnih događaja u fiksnom vremenskom periodu ili prostornom okruženju, kao što su broj telefonskih poziva u određenom vremenskom intervalu, broj automobila koji prolaze kroz određeno raskrižje u određenom vremenskom intervalu, broj pravopisnih grešaka na stranici novina, itd.

Parametar Poissonove distribucije je λ koji predstavlja očekivani broj događaja u nekom vremenskom periodu ili prostornom području.

Definicija 1.2.3. *Neka je* Y *diskretna slučajna varijabla definirana na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$. *Kažemo da slučajna varijabla* Y *ima* **Poissonovu distribuciju** *s pa-*

rametrom $\lambda > 0$, u oznaci $Y \sim P(\lambda)$, ako joj je funkcija gustoće

$$f_Y(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}. \tag{1.22}$$



Slika 1.3: Funkcija gustoće slučajne varijable $Y \sim P(\lambda)$

Propozicija 1.2.3. *Neka je $Y \sim P(\lambda)$, $\lambda > 0$. Vrijedi*

$$\mathbb{E}[Y] = \lambda, \tag{1.23}$$
$$\mathrm{Var}(Y) = \lambda. \tag{1.24}$$

Normalna razdioba

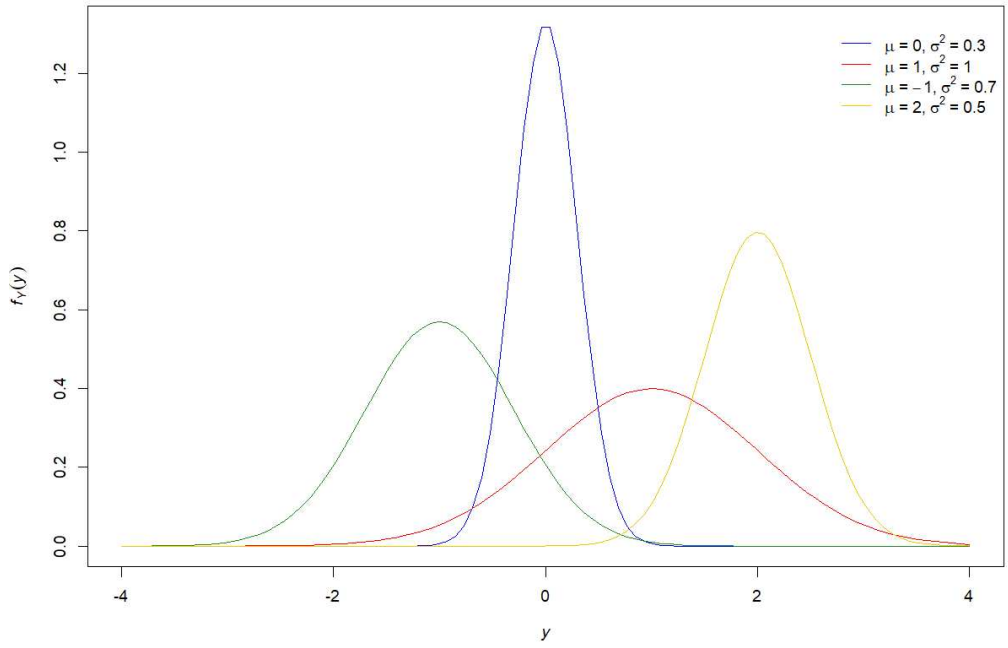
Normalna distribucija jedna je od najvažnijih i najpoznatijih neprekidnih vjerojatnosnih distribucija koja se široko primjenjuje u statistici. Njena važnost proizlazi iz činjenice da se mnogi prirodni i društveni fenomeni, poput visine, težine te krvnog tlaka osobe, mogu opisati normalnom distribucijom. Osim toga, normalna je distribucija temelj mnogih statističkih teorija i metoda. Jedan od najvažnijih rezultata u teoriji vjerojatnosti je *Centralni granični teorem*, koji navodi da će distribucija prosjeka slučajnog uzorka biti približno normalno distribuirana, čak i kada originalni podaci nisu normalni. Ovaj teorem igra ključnu

ulogu u statistici, omogućujući primjenu normalne distribucije u različitim statističkim analizama i testovima.

Normalna je distribucija opisana s dva parametra: srednjom vrijednošću μ i standardnom devijacijom σ. Srednja vrijednost označava aritmetičku sredinu distribucije i predstavlja središte krivulje normalne distribucije. Standardna devijacija mjeri raspršenost podataka oko srednje vrijednosti i određuje širinu krivulje.

Definicija 1.2.4. *Neka je Y neprekidna slučajna varijabla definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Kažemo da slučajna varijabla Y ima* **normalnu razdiobu** *s parametrima $\mu \in \mathbb{R}$ i $\sigma^2 > 0$, u oznaci $Y \sim N(\mu, \sigma^2)$, ako joj je funkcija gustoće*

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right), \quad y \in \mathbb{R}. \tag{1.25}$$



Slika 1.4: Funkcija gustoće slučajne varijable $Y \sim N(\mu, \sigma^2)$

Propozicija 1.2.4. *Neka je $Y \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Vrijedi*

$$\mathbb{E}[Y] = \mu, \tag{1.26}$$

$$\mathrm{Var}(Y) = \sigma^2. \tag{1.27}$$

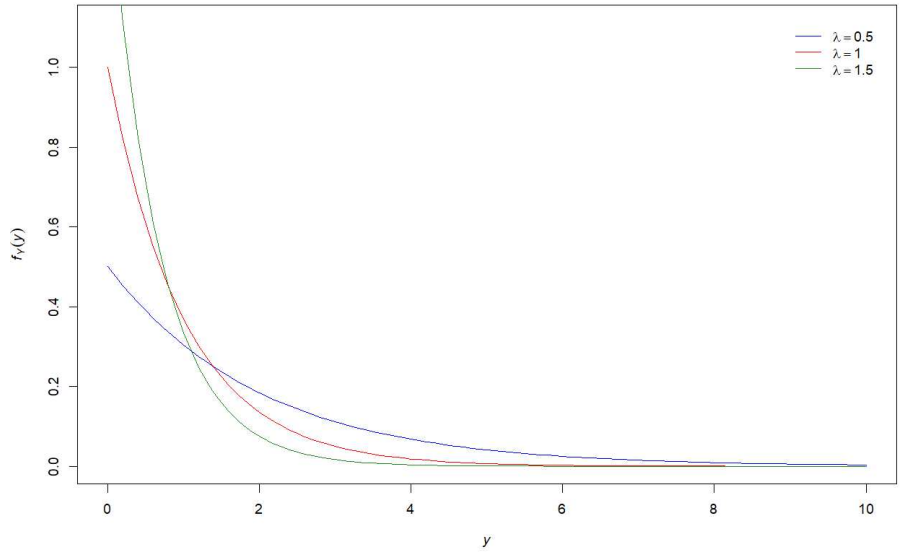
Eksponencijalna razdioba

Eksponencijalna distribucija je neprekidna vjerojatnosna distribucija koja opisuje vrijeme između događaja u Poissonovom procesu, gdje se događaji događaju kontinuirano i nezavisno jedno od drugog s konstantnom srednjom brzinom. Koristimo ju za modeliranje vremena čekanja, poput vremena do kvara uređaja, trajanja poziva ili vremena do sljedećeg dolaska kupca.

Eksponencijalnu distribuciju opisujemo parametrom λ koji predstavlja brzinu (ili intenzitet) dolaska događaja.

Definicija 1.2.5. *Neka je Y neprekidna slučajna varijabla definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Kažemo da slučajna varijabla Y ima* **eksponencijalnu razdiobu** *s parametrom $\lambda > 0$, u oznaci $Y \sim \text{Exp}(\lambda)$, ako joj je funkcija gustoće*

$$f_Y(y; \lambda) = \lambda e^{-\lambda y}, \quad y \geq 0. \tag{1.28}$$



Slika 1.5: Funkcija gustoće slučajne varijable $Y \sim \text{Exp}(\lambda)$

Propozicija 1.2.5. *Neka je $Y \sim \text{Exp}(\lambda)$, $\lambda > 0$. Vrijedi*

$$\mathbb{E}[Y] = \frac{1}{\lambda}, \tag{1.29}$$

$$\text{Var}(Y) = \frac{1}{\lambda^2}. \tag{1.30}$$

Gama razdioba

Razmotrimo situaciju u kojoj pratimo događaje koji se javljaju u nekom vremenskom razdoblju, a brzina tih događaja je stalna i ne mijenja se, kao što su vrijeme između pojava zemljotresa, vrijeme između promjena cijena, vrijeme između stvaranja potomstva ili vrijeme između doziranja lijeka. Ako nas zanima koliko će vremena proći prije nego se dogodi određeni broj tih događaja, možemo koristiti gama distribuciju.

Ova se distribucija opisuje s dva parametra: jedan nam parametar govori koliko događaja trebamo da bi se postigao cilj (taj broj nazivamo "broj događaja" ili "količina" i u ovome radu označavamo s α), dok drugi parametar opisuje koliko su događaji učestali ili brzi (taj parametar nazivamo "intenzitet" i označavamo s β).

Definicija 1.2.6. *Neka je Y neprekidna slučajna varijabla definirana na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Kažemo da slučajna varijabla Y ima* **gama razdiobu** *s parametrima $\alpha > 0$ i $\beta > 0$, u oznaci $Y \sim \Gamma(\alpha, \beta)$, ako joj je funkcija gustoće*

$$f_Y(y; \alpha, \beta) = \frac{y^{\alpha-1}\beta^{\alpha}e^{-y\beta}}{\Gamma(\alpha)}, \quad y > 0, \tag{1.31}$$

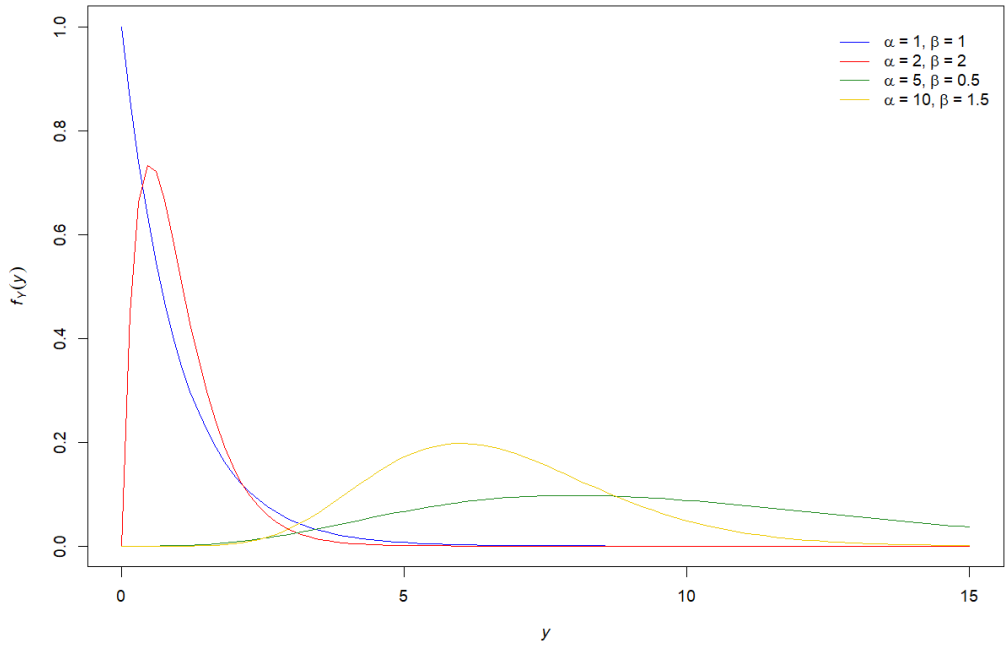
gdje funkcija $\Gamma(\alpha)$ označava gama funkciju evaluiranu u α, tj.

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} \cdot e^{-t}\, dt.$$

Propozicija 1.2.6. *Neka je $Y \sim \Gamma(\alpha, \beta), \alpha > 0, \beta > 0$. Vrijedi*

$$\mathbb{E}[Y] = \frac{\alpha}{\beta}, \tag{1.32}$$

$$\mathrm{Var}(Y) = \frac{\alpha}{\beta^2}. \tag{1.33}$$



Slika 1.6: Funkcija gustoće slučajne varijable $Y \sim \Gamma(\alpha, \beta)$

1.3 Regresijska analiza

U statistici često proučavamo kako jedna mjerena veličina, poput visine plaće, ovisi o drugim mjerenim veličinama, kao što su spol, razina obrazovanja i broj godina iskustva. Veličinu od interesa modeliramo kao slučajnu varijablu koju nazivamo **zavisnom varijablom** ili **odgovorom** (engl. *response variable*) jer se pretpostavlja da njena vrijednost ovisi o vrijednostima drugih varijabli. S druge strane, varijable koje utječu na zavisnu varijablu nazivamo **nezavisne varijable**, **prediktori** ili **regresori** (engl. *explanatory variables*).

Da bismo formalizirali naš pristup, uvodimo notaciju za zavisne i nezavisne varijable koju ćemo koristiti kroz ovaj rad. Pretpostavimo da imamo n mjerenja (često kažemo i opservacija). Zavisnu varijablu označavamo kao n-dimenzionalni slučajni vektor $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, a njene pripadne realizacije (tj. mjerenja) označavamo s $\boldsymbol{y} = (y_1, \ldots, y_n)^T$. Prediktore možemo predstaviti u $n \times (p+1)$ matrici X koju nazivamo **matrica dizajna**, gdje je p broj prediktora. Svaka opservacija $i = 1, \ldots, n$ ima $p+1$ vrijednost nezavisnih

varijabli koje označavamo s $\boldsymbol{x}_i^T = (1, x_{i1}, \dots x_{ip})$ pa matricu dizajna možemo zapisati kao

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix},$$

U ovom kontekstu, svaki redak matrice X predstavlja jednu opservaciju, dok svaki stupac predstavlja jednu nezavisnu varijablu.

Otkrivanje načina na koji prediktori utječu na zavisnu varijablu predstavlja jedan od ključnih izazova u statistici. Prema [2], regresijska analiza je metoda ispitivanja i analize ovisnosti jedne zavisne varijable o jednoj ili više nezavisnih varijabli. Kao rezultat regresijske analize, razvija se regresijski model, koji se može opisati kao matematička jednadžba koja kvantificira povezanost između zavisne varijable i prediktora. Ova je veza obično izražena funkcijom f koja se definira kroz regresijske parametre $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$. U najopćenitijem obliku, regresijsku jednadžbu možemo zapisati kao:

$$Y_i = f(1, x_{i1}, \dots, x_{ip}; \beta_0, \beta_1, \dots, \beta_p) + \epsilon_i, \quad i = 1, \dots, n,$$

gdje su

- Y_i zavisna varijabla za i-tu opservaciju,
- $x_{i1}, \dots, x_{ip}$ nezavisne varijable za i-tu opservaciju,
- $f(\cdot)$ funkcija koja modelira odnos između zavisne i nezavisnih varijabli,
- ϵ_i slučajne varijable koje zovemo **slučajnim pogreškama** ili **šumovima**.

Zavisne varijable Y_i tretiraju se kao slučajne varijable koje slijede određenu distribuciju vjerojatnosti jer njihove vrijednosti mogu varirati zbog različitih nepredvidljivih čimbenika ϵ_i, kao što su slučajni utjecaji, pogreške u mjerenju ili fluktuacije koje nije moguće unaprijed odrediti. Na primjer, visina plaće može ovisiti o spolu, dobi, razini obrazovanja i broju godina iskustva. Međutim, čak i kada uzmemo u obzir sve te informacije, varijacije u visini plaće mogu nastati zbog dodatnih faktora, kao što su individualne pregovaračke vještine ili trenutno stanje tržišta rada. Zbog ovih se utjecaja, zavisna varijabla tretira kao slučajna varijabla.

S druge strane, prediktori ili nezavisne varijable $x_{i1}, \dots, x_{ip}$ obično se tretiraju kao neslučajna mjerenja ili opažanja. To znači da se njihove vrijednosti ne mijenjaju slučajno za pojedinačne opservacije unutar našeg uzorka podataka. Prediktori su često fiksne ili unaprijed određene vrijednosti koje su prikupljene na temelju određenog dizajna eksperimenta ili studije. Na primjer, ako provodimo anketu o plaćama, razina obrazovanja (osnovna, srednja škola ili fakultet) i spol (muškarac, žena) ispitanika su unaprijed poznate

i ne mijenjaju se tijekom prikupljanja podataka. Pretpostavka da su prediktori neslučajni pomaže u pojednostavljivanju modela i analize, omogućujući nam da se usredotočimo na proučavanje utjecaja tih prediktora na slučajne odgovore.

Funkcija f koja modelira odnos između zavisne varijable i prediktora može poprimiti razne oblike. U ovome radu fokusirat ćemo se na regresijske modele čija je funkcija f linearna u parametrima $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. To znači da ćemo modelirati odnos zavisne varijable Y_i i nezavisnih varijabli $x_{i1}, \ldots x_{ip}$ pomoću linearnog izraza. Regresijska se jednadžba u ovom slučaju može zapisati kao

$$Y_i = f(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) + \epsilon_i = f(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \epsilon_i, \quad i = 1, \ldots, n.$$

Osnovni je cilj svake regresijske analize predviđanje očekivane vrijednosti zavisne varijable. Za $i = 1, \ldots, n$ i fiksne $x_{i1}, \ldots, x_{ip}$ iz svojstava matematičkog očekivanja slijedi:

$$\begin{aligned} \mu_i = \mathbb{E}[Y_i] &= \mathbb{E}[f(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \epsilon_i] \\ &= f(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \mathbb{E}[\epsilon_i] \\ &= f(\boldsymbol{x}_i^T \boldsymbol{\beta}). \end{aligned}$$

Ključna je pretpostavka regresijske analize da je očekivanje slučajnih pogrešaka $\mathbb{E}[\epsilon_i] = 0$. To nam osigurava da model ne podcjenjuje ili precjenjuje zavisnu varijablu. Na primjer, pretpostavimo da modeliramo visinu plaće na temelju broja godina iskustva, razine obrazovanja i spola. Kada bi očekivanje pogrešaka $\mathbb{E}[\epsilon_i] = 500$, to bi značilo da naš model sustavno promašuje pravu vrijednost plaće za 500, što dovodi do netočnije procjene plaće svakog zaposlenika. Time bi u prosjeku naš model bio u pravilu lošiji, jer bi izostavljao relevantne informacije koje utječu na plaću. S druge strane, kada pretpostavimo da je $\mathbb{E}[\epsilon_i] = 0$, to znači da su pogreške ravnomjerno raspoređene oko nule - uz povremeno podcjenjivanje i povremeno precjenjivanje.

Normalni linearni modeli

Kada se statističari susretnu s podacima gdje žele modelirati jednu zavisnu varijablu na osnovu jedne ili više nezavisnih varijabli, obično prvi pokušaj uključuje korištenje normalne linearne regresije. Ova se metoda često koristi jer je relativno jednostavna za interpretaciju i primjenu.

U daljnjem ćemo tekstu istražiti koncept normalnih linearnih modela, počevši od njihove matematičke definicije i osnovnih pretpostavki. Istražit ćemo i ograničenja koja se mogu javiti kada pretpostavke linearnih modela nisu zadovoljene, naglašavajući važnost razumijevanja uvjeta pod kojima ovi modeli pružaju valjane rezultate. Cilj nam je pružiti čvrst temelj za daljnje istraživanje generaliziranih linearnih modela, čija se primjena proširuje na širi spektar distribucija i situacija koje normalni linearni modeli ne mogu riješiti.

Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ vektor nezavisnih slučajnih varijabli za koje pretpostavljamo da ovise o vrijednostima $x_1, \ldots, x_p$. Normalni linearni regresijski modeli su modeli koje opisujemo jednadžbom

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n.$$

Dodatne pretpostavke na normalni linearni regresijski model usmjerene su na karakteristike slučajnih pogrešaka unutar modela te prema [5] one uključuju da su slučajne pogreške:

(i) centrirane: $\mathbb{E}[\epsilon_i] = 0$ za sve $i = 1, \ldots, n$,

(ii) homoskedastične: $\mathrm{Var}(\epsilon_i) = \sigma^2$ za sve $i = 1, \ldots, n$,

(iii) nekorelirane: $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ za sve $i \neq j$,

(iv) normalno distribuirane: $\epsilon_i \sim N(0, \sigma^2)$ za sve $i = 1, \ldots, n$.

Nadalje, pretpostavka o normalnosti slučajnih pogrešaka implicira normalnost zavisne varijable, tj. $Y_i \sim N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2)$.

Osnovni je cilj linearne regresije predviđanje vrijednosti zavisne varijable Y_i na temelju formule

$$\mathbb{E}[Y_i] = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \boldsymbol{x}_i^T \boldsymbol{\beta}, \quad i = 1, \ldots, n.$$

Međutim, ako pretpostavke *(i)-(iv)* nisu zadovoljene, a pretpostavimo ih u procesu statističkog zaključivanja, tada mogu utjecati na točnost naših procjena i dovesti do pogrešnih zaključaka.

Primjer 1.3.1. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ vektor nezavisnih slučajnih varijabli takvih da $Y_i \sim B(n_i, p_i)$, gdje su $n_i \in \mathbb{N}$ i $p_i \in \langle 0, 1\rangle$ za sve $i = 1, \ldots, n$. Nadalje, pretpostavimo da su Y_i zavisne varijable koje ovise o vrijednostima $x_{i1}, \ldots, x_{ip}$ linearno, tj.*

$$\mathbb{E}[Y_i] = \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

Za $Y_i \sim B(n_i, p_i)$ je iz (1.20) $\mathbb{E}[Y_i] = n_i p_i$, *odakle slijedi*

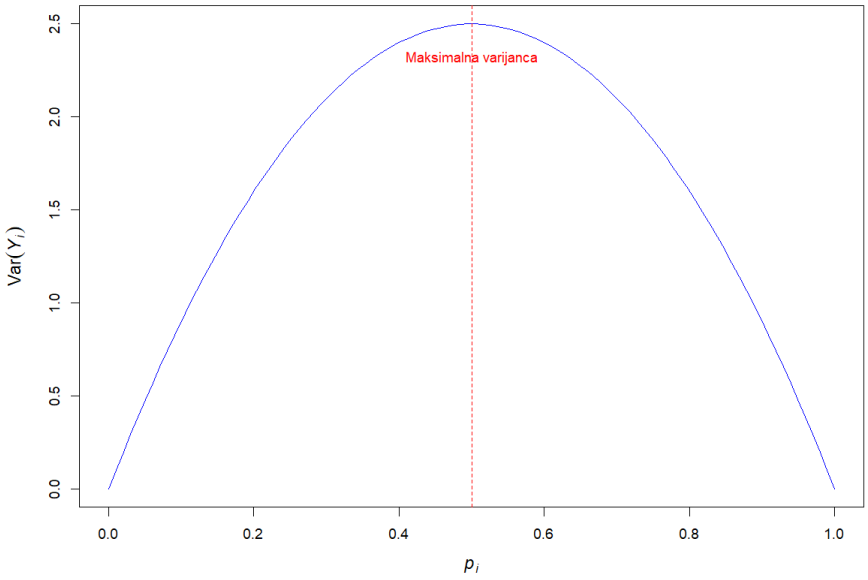
$$n_i p_i = \boldsymbol{x}_i^T \boldsymbol{\beta}. \tag{1.34}$$

Kako je $p_i \in \langle 0, 1\rangle$ vjerojatnost uspjeha, onda je lijeva strana jednakosti $n_i p_i \in \langle 0, n_i\rangle$. Međutim, linearna kombinacija prediktora $\boldsymbol{x}_i^T \boldsymbol{\beta}$ nije ograničena na $\langle 0, n_i\rangle$, nego može poprimiti bilo koju realnu vrijednost, uključujući negativne vrijednosti ili vrijednosti veće od n_i. Stoga, modeliranje normalnom linearnom regresijom u ovom slučaju ne bi bilo prikladno. Probleme ovakvog tipa riješit ćemo definiranjem odgovarajuće funkcije poveznice kod generaliziranih linearnih modela.

Nadalje, iz (1.21) *slijedi da je varijanca slučajne varijable* Y_i *jednaka*

$$\text{Var}(Y_i) = n_i p_i (1 - p_i).$$

Ovdje je jasno da varijanca ovisi o vjerojatnosti p_i *koja može poprimiti različite vrijednosti ovisno o nezavisnim varijablama, što je vidljivo i na slici 1.7. Plava krivulja prikazuje kako se varijanca mijenja s* p_i *- doseže svoj maksimum kada je* $p_i = 0.5$ *i smanjuje se kako se* p_i *približava* 0 *ili* 1*. Dakle, svojstvo (ii) nije zadovoljeno što nas ponovno dovodi do zaključka da modeliranje normalnom linearnom regresijom nije prikladno za ovakav tip podataka.*



Slika 1.7: Heteroskedastičnost varijance slučajne varijable $Y_i \sim B(n_i, p_i)$

Osnova za izradu ovog rada došla je iz prepoznavanja upravo ovakvih ograničenja normalnih linearnih modela. Ti nedostaci često su se odnosili na probleme s normalnošću slučajnih grešaka, heteroskedastičnost i teškoće u modeliranju zavisnih varijabli koje nisu neprekidne. *Generalizirani linearni modeli*, koje ćemo detaljnije opisati u sljedećem poglavlju, čine proširenje teorije normalnih linearnih modela, pružaju veću prilagodljivost u proučavanju različitih tipova zavisnih varijabli i omogućuju korištenje različitih distribucija pogrešaka. Ovim pristupom nastojimo bolje razumjeti odnose između varijabli i dobiti preciznije rezultate u analizi podataka.

Poglavlje 2

Generalizirani linearni modeli

Generalizirani linearni modeli (skraćeno *GLM*) predstavljaju fleksibilan okvir za modeliranje širokog spektra podataka i odnosa između varijabli. Osnovna je ideja *GLM*-a proširenje normalnog linearnog modela kako bi se omogućilo modeliranje podataka u kojima pretpostavke za modeliranje normalnom linearnom regresijom nisu zadovoljene.

Kao što smo vidjeli, u normalnim linearnim modelima polazimo od pretpostavke da postoji linearna veza između očekivanja varijable odziva Y_i i kovarijata $x_{i1}, \ldots, x_{ip}$, tj. za $i = 1, \ldots, n$ su

$$\begin{cases} Y_i \overset{nez.}{\sim} N(\mu_i, \sigma^2) \\ \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}. \end{cases}$$

Međutim, dosadašanje analize pokazale su nam da određene vrste varijabli odziva ne zadovoljavaju prvu pretpostavku o normalnosti pa samim time ni nisu prikladne za modeliranje normalnom linearnom regresijom. Takve varijable odziva uključuju binomnu, Poissonovu, eksponencijalnu, gama distribuciju i mnoge druge. Osim toga, veza između varijable odziva i prediktora ne mora nužno biti linearna, tj. oblika $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$.
Generalizirani linearni modeli ove probleme rješavaju kroz sljedeće tri komponente:

(i) distribucija varijable odziva pripada široj klasi distribucija koju zovemo *eksponencijalna familija distribucija*,

(ii) *linearni predviditelj* opisuje kako prediktori utječu na distribuciju zavisne varijable putem linearne kombinacije prediktora,

(iii) *funkcija poveznica* povezuje očekivanu vrijednost zavisne varijable s linearnim predviditeljem.

Prije nego što damo definiciju generaliziranih linearnih modela, upoznat ćemo se s gore uvedenim pojmovima eksponencijalne familije distribucija, linearnog predviditelja i funkcije poveznice.

2.1 Eksponencijalna familija distribucija

Eksponencijalna familija distribucija čini jednu od najvažnijih klasa distribucija u statistici. Mnoge poznate i često korištene distribucije, poput normalne, binomne, Poissonove ili gama distribucije, imaju neka zajednička svojstva te pripadaju ovoj klasi distribucija.

U narednim odlomcima dat ćemo definiciju eksponencijalnih familija distribucija, s posebnim naglaskom na njihovu *standardnu* (*kanonsku*) formu. Ova nam je forma od posebne važnosti jer igra ključnu ulogu u uvođenju i razumijevanju generaliziranih linearnih modela. Iako ova familija distribucija posjeduje brojne zanimljive karakteristike, fokus ćemo staviti na ona svojstva koja će nam biti ključna za razvoj teorije generaliziranih linearnih modela - izvest ćemo formulu za očekivanje i varijancu slučajne varijable iz eksponencijalne familije distribucija. Za opsežniju razradu ove teme, čitatelj može konzultirati [3].

Definicija

Funkcija gustoće svake slučajne varijable koja pripada eksponencijalnoj familiji distribucija ima specifičan oblik. U različitim literaturama može se pronaći više varijanti te funkcije gustoće, no važno je naglasiti da su sve te varijante međusobno ekvivalentne. U ovome radu slijedimo definiciju eksponencijalne familije distribucija po [1].

Definicija 2.1.1. *Neka je $Y : \Omega \to \mathbb{R}$ slučajna varijabla čija vjerojatnosna distribucija (diskretna ili neprekidna) ovisi o jednom parametru θ. Distribucija slučajne varijable Y pripada* **eksponencijalnoj familiji distribucija** *ako se gustoća f_Y može zapisati kao*

$$f_Y(y;\theta) = s(y)t(\theta)\exp\left[a(y)b(\theta)\right], \tag{2.1}$$

za $y \in S$ gdje je $S \subseteq \mathbb{R}$ takav da je $\int_S f_Y(y;\theta)dy = 1$ u neprekidnom slučaju, odnosno $\sum_{y\in S} f_Y(y;\theta) = 1$ u diskretnom slučaju te $s : S \to \langle 0, +\infty\rangle$, $t : \Theta \to \langle 0, +\infty\rangle$, $a : S \to \mathbb{R}$ i $b : \Theta \to \mathbb{R}$ poznate funkcije.
Ukoliko je $a(y) = y$ za distribuciju kažemo da je u **standardnoj (kanonskoj) formi** *s* **prirodnim parametrom** *$b(\theta)$. Tada pišemo $Y \sim EFD(\theta)$.*

U slučaju da uz parametar od interesa θ postoje i drugi parametri distribucije, njih ćemo smatrati poznatima te ih nećemo procjenjivati.

Nadalje, ako stavimo $s(y) = \exp d(y)$ za neku funkciju $d : S \to \mathbb{R}$ i $t(\theta) = \exp c(\theta)$ za neku funkciju $c : \Theta \to \mathbb{R}$ tada se jednadžba (2.1) može zapisati na sljedeći način

$$f_Y(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]. \tag{2.2}$$

U daljnjem tekstu rada koristit ćemo taj oblik funkcije gustoće, budući da je ovaj zapis analitički povoljniji za izvođenje i daljnje formulacije statističkih rezultata koje ćemo prikazati.

S obzirom da nastojimo generalizirati teoriju normalnih linearnih modela, pokažimo prvo da normalna razdioba pripada eksponencijalnoj familiji distribucija u standardnoj (kanonskoj) formi.

Primjer 2.1.1. *Neka je* $Y \sim N(\mu, \sigma^2)$*, gdje je* $\mu \in \mathbb{R}$ *parametar od interesa i* $\sigma^2 > 0$ *poznata. Funkciju gustoće iz* (1.25) *moguće je napisati u obliku* (2.2) *na sljedeći način:*

$$\begin{aligned} f_Y(y;\mu) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) \\ &= \exp\left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu^2\right)\right)\right] \\ &= \exp\left[\log\left((2\pi\sigma^2)^{-\frac{1}{2}}\right) + \log\left[\exp\left(-\frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right)\right]\right] \\ &= \exp\left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right] \\ &= \exp\left[\frac{\mu}{\sigma^2}y - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right]. \end{aligned}$$

Vidimo da su

$$\begin{aligned} a(y) &= y, \\ b(\mu) &= \frac{1}{\sigma^2}\mu, \\ c(\mu) &= -\frac{1}{2\sigma^2}\mu^2, \\ d(y) &= -\frac{1}{2\sigma^2}y^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right). \end{aligned}$$

Dakle, normalna razdioba pripada eksponencijalnoj familiji distribucija u standardnoj formi.

Pogledajmo i primjer distribucije iz eksponencijalne familije koja nije u standardnoj formi te distribucije koja ne pripada eksponencijalnoj familiji distribucija.

Primjer 2.1.2. *Neka je* $Y : \Omega \to \mathbb{R}$ *neprekidna slučajna varijabla kojoj funkcija gustoće ovisi o jednom parametru* $\alpha > 0$ *te je dana jednadžbom*

$$f_Y(y;\alpha) = \alpha y^{-\alpha-1}, \quad y > 0. \tag{2.3}$$

Tada vrijedi

$$f_Y(y;\alpha) = \exp\left[(-\alpha - 1)\log y + \log\alpha\right].$$

Za

$$\begin{aligned} a(y) &= \log y, \\ b(\alpha) &= -\alpha - 1, \\ c(\alpha) &= \log \alpha, \\ d(y) &= 0 \end{aligned}$$

je dana funkcija gustoće u obliku (2.2) *pa zaključujemo da pripada eksponencijalnoj familiji distribucija. Kako je* $a(y) = \log y$*, gustoća nije u standardnoj formi.*

Primjer 2.1.3. *Neka je* $Y : \Omega \to \mathbb{R}$ *neprekidna slučajna varijabla kojoj funkcija gustoće ovisi o jednom parametru* $n > 0$ *te je dana jednadžbom*

$$f_Y(y; n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1 + \frac{1}{n}y^2\right)^{-\frac{n+1}{2}}, \quad y \in \mathbb{R}. \tag{2.4}$$

Tada vrijedi

$$f_Y(y; n) = \exp\left[\log\Gamma\left(\frac{n+1}{2}\right) - \log\Gamma\left(\frac{n}{2}\right) - \frac{1}{2}\log(n\pi) - \frac{n+1}{2}\log\left(1 + \frac{1}{n}y^2\right)\right].$$

Kako $\log\left(1 + \frac{1}{n}y^2\right)$ *ne možemo raspisati u obliku* $a(y)b(n)$*, očito ovakva distribucija ne pripada eksponencijalnoj familiji distribucija.*

Distribuciju kojoj je gustoća dana s (2.4) *nazivamo* **Studentova t-distribucija** *s n stupnjeva slobode i označavamo s* $Y \sim t(n)$*. Ona nam je bitna za analizu podataka u situacijama s malim uzorcima, kod procjene varijance populacije te testiranja statističkih hipoteza.*

Tablica 2.1 prikazuje poznate distribucije koje pripadaju eksponencijalnoj familiji u standardnoj formi te izraze za funkcije b, c i d iz (2.2). Ove se distribucije najčešće koriste za modeliranje varijable odziva kod generaliziranih linearnih modela.

Distribucija	b	c	d
$Y \sim \text{Bernoulli}(p), p \in [0,1]$	$\log \frac{p}{1-p}$	$\log(1-p)$	0
$Y \sim B(n,p), p \in [0,1], n \in \mathbb{N}$ poznat	$\log \frac{p}{1-p}$	$n\log(1-p)$	$\log \binom{n}{y}$
$Y \sim P(\lambda), \lambda > 0$	$\log \lambda$	$-\lambda$	$-\log y!$
$Y \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > o$ poznat	$\frac{1}{\sigma^2}\mu$	$-\frac{1}{2\sigma^2}\mu^2$	$-\frac{1}{2\sigma^2}y^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right)$
$Y \sim \text{Exp}(\lambda), \lambda > 0$	$-\lambda$	$\log \lambda$	0
$Y \sim \Gamma(\alpha,\beta), \beta > 0, \alpha > 0$ poznata	$-\beta$	$\alpha \log \beta$	$(\alpha-1)\log y - \log \Gamma(\alpha)$

Tablica 2.1: Distribucije unutar eksponencijalne familije distribucija

Napomena 2.1.1. *U ovom radu, većina analiza i rezultata bit će predstavljena korištenjem integrala koji impliciraju pretpostavku o neprekidnosti slučajne varijable. Međutim, kako eksponencijalna familija distribucija, pa samim time i generalizirani linearni modeli, uključuju i diskretne distribucije, isti zaključci i rezultati mogu se analogno primijeniti na diskretne slučajne varijable odgovarajućom zamjenom integrala sumom.*

Očekivanje i varijanca

Jedno od svojstava eksponencijalne familije distribucija je da očekivanje ovisi jedino o parametru od interesa θ, dok varijanca predstavlja funkciju očekivanja, što ćemo pokazati u sljedećem teoremu.

Teorem 2.1.1. *Neka je $Y : \Omega \longrightarrow \mathbb{R}$ slučajna varijabla čija distribucija pripada eksponencijalnoj familiji distribucija. Tada je*

$$(i)\ \ \mathbb{E}[a(Y)] = -\frac{c'(\theta)}{b'(\theta)} \tag{2.5}$$

$$(ii)\ \ \text{Var}(a(Y)) = -\frac{b''(\theta)\mathbb{E}[a(Y)]+c''(\theta)}{[b'(\theta)]^2} = \frac{b''(\theta)c'(\theta)-c''(\theta)b'(\theta)}{[b'(\theta)]^3}, \tag{2.6}$$

gdje su $b'(\theta) = \frac{db}{d\theta}$, $c'(\theta) = \frac{dc}{d\theta}$, $b''(\theta) = \frac{d^2b}{d\theta^2}$ i $c''(\theta) = \frac{d^2c}{d\theta^2}$.

Dokaz. Dokaz ovog teorema prati [1]. Neka je $Y : \Omega \to \mathbb{R}$ slučajna varijabla kojoj distribucija pripada eksponencijalnoj familiji distribucija.

Iz (1.1) znamo da funkcija gustoće slučajne varijable Y zadovoljava

$$\int_{\mathbb{R}} f_Y(y;\theta)\, dy = 1.$$

Pretpostavimo da gornji izraz možemo dvaput diferencirati po θ uvođenjem znaka diferenciranja pod integral (što je uvijek moguće kod eksponencijalnih familija), odakle slijedi

$$(a)\ \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f_Y(y;\theta)\,dy = 0, \tag{2.7}$$

$$(b)\ \int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2} f_Y(y;\theta)\,dy = 0. \tag{2.8}$$

(i) Pokazat ćemo da je $\mathbb{E}[a(Y)] = -\frac{c'(\theta)}{b'(\theta)}$ pomoću (2.7).

Iz (2.2) slijedi

$$\frac{\partial}{\partial\theta} f_Y(y;\theta) = \left[a(y)b'(\theta) + c'(\theta)\right] f_Y(y;\theta). \tag{2.9}$$

Uvrštavanjem ovog rezultata u (2.7) i korištenjem svojstva linearnosti integrala dobivamo

$$b'(\theta)\int_{\mathbb{R}} a(y) f_Y(y;\theta)\,dy + c'(\theta)\int_{\mathbb{R}} f_Y(y;\theta)\,dy = 0.$$

Na kraju, iz definicije matematičkog očekivanja slučajne varijable $a(Y)$ slijedi da je $\int_{\mathbb{R}} a(y) f_Y(y;\theta)\,dy \overset{(1.4)}{=} \mathbb{E}[a(Y)]$ pa je

$$\mathbb{E}[a(Y)] = -\frac{c'(\theta)}{b'(\theta)}.$$

(ii) Pokazat ćemo da je $\mathrm{Var}(a(Y)) = \frac{b''(\theta)\mathbb{E}[a(Y)] + c''(\theta)}{[b'(\theta)]^2}$ pomoću (2.8).

Iz (2.9) slijedi

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2} f_Y(y;\theta) &= \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta} f_Y(y;\theta)\right) \\
&= \frac{\partial}{\partial\theta}\left[\left(a(y)b'(\theta) + c'(\theta)\right) f_Y(y;\theta)\right] \\
&= \left(a(y)b''(\theta) + c''(\theta)\right) f_Y(y;\theta) + \left(a(y)b'(\theta) + c'(\theta)\right)^2 f_Y(y;\theta) \\
&\overset{(2.5)}{=} \left(a(y)b''(\theta) + c''(\theta)\right) f_Y(y;\theta) + \left(b'(\theta)\right)^2 \left(a(y) - \mathbb{E}[a(Y)]\right)^2 f_Y(y;\theta),
\end{aligned}$$

Uvrštavanjem ovog rezultata u (2.8) i korištenjem svojstva linearnosti integrala dobivamo

$$b''(\theta)\int_{\mathbb{R}} a(y) f_Y(y;\theta)\,dy + c''(\theta)\int_{\mathbb{R}} f_Y(y;\theta)\,dy + \left(b'(\theta)\right)^2 \int_{\mathbb{R}} \left(a(y) - \mathbb{E}[a(Y)]\right)^2 f_Y(y;\theta)\,dy = 0.$$

Iz definicije matematičkog očekivanja i varijance slučajne varijable $a(Y)$ slijedi da je

$$\int_{\mathbb{R}} \left(a(y) - \mathbb{E}[a(Y)]\right)^2 f_Y(y;\theta,\varphi)\,dy \overset{(1.4)}{=} \mathbb{E}[(a(Y) - \mathbb{E}[a(Y)])^2] \overset{(1.8)}{=} \mathrm{Var}(a(Y))$$

pa je

$$\begin{aligned}\mathrm{Var}(a(Y)) &= -\frac{b''(\theta)\mathbb{E}[a(Y)] + c''(\theta)}{[b'(\theta)]^2} \\ &\overset{(2.5)}{=} \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}.\end{aligned}$$

□

Uvjerimo se da ovako dobivene formule za očekivanje i varijancu zaista daju dobre rezultate na primjeru normalne razdiobe.

Primjer 2.1.4. *Neka je* $Y \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$ *poznata. Iz* (2.5)*,* (2.6) *i tablice 2.1 slijedi*

$$\mathbb{E}[Y] = -\frac{c'(\mu)}{b'(\mu)} = -\frac{-\frac{\mu}{\sigma^2}}{\frac{1}{\sigma^2}} = \mu,$$

$$\mathrm{Var}(Y) = \frac{b''(\mu)c'(\mu) - c''(\mu)b'(\mu)}{[b'(\mu)]^3} = \frac{\frac{1}{\sigma^2}\cdot\frac{1}{\sigma^2}}{\left[\frac{1}{\sigma^2}\right]^3} = \sigma^2.$$

Dakle, formule daju očekivane rezultate za normalnu razdiobu.

Slično se možemo uvjeriti i za ostale distribucije iz tablice 2.1 koristeći odgovarajuće izraze za funkcije b i c.

2.2 Linearni predviditelj

Linearni predviditelj predstavlja linearnu kombinaciju nezavisnih varijabli koje utječu na zavisnu varijablu, određujući time njezinu razdiobu. U nastavku slijedi definicija linearnih predviditelja u najopćenitijem obliku.

Definicija 2.2.1. *Neka su* $\boldsymbol{x} = (1, x_1, \ldots, x_p)^T$ *prediktori te* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ *parametri. Linearnu kombinaciju prediktora*

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p = \boldsymbol{x}^T\boldsymbol{\beta} \tag{2.10}$$

nazivamo **linearnim predviditeljem**.

Ovako definirani, linearni predviditelji omogućuju analiziranje utjecaja nezavisnih varijabli na zavisnu varijablu. Međutim, linearni predviditelji pružaju i fleksibilnost u modeliranju složenijih odnosa jer njihovom primjenom možemo jednostavno uključiti interakcijske efekte između varijabli ili modelirati nelinearne odnose. Pogledajmo kako bismo to napravili kroz sljedećih nekoliko primjera.

Pretpostavimo da želimo modelirati visinu plaće Y pomoću dobi x_1, razine obrazovanja x_2 (gdje su: osnovna škola = 1, srednja škola = 2, fakultet = 3) i broja godina iskustva x_3. U nastavku dajemo primjere mogućih linearnih predviditelja za modeliranje visine plaće pomoću navedenih varijabli.

Primjer 2.2.1. *Najjednostavniji linearni predviditelj.*

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Ovdje se svaki prediktor linearno kombinira s odgovarajućim koeficijentom kako bi se modelirala visina plaće.

Primjer 2.2.2. *Model s interakcijama između prediktora.*

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 (x_1 \cdot x_2) + \beta_5 (x_1 \cdot x_3) + \beta_6 (x_2 \cdot x_3)$$

Ovaj model uključuje sve moguće interakcije između prediktora, omogućujući istraživanje njihovih međusobnih utjecaja na plaću.

Primjer 2.2.3. *Model s nelinearnim efektom.*

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2.$$

Kvadratni član $\beta_4 x_1^2$ omogućuje modeliranje situacije u kojoj visina plaće raste do određene dobi, a zatim stagnira ili opada.

Primjer 2.2.4. *Model s indikatorskom varijablom.*

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_{2,2} + \beta_3 x_{2,3} + \beta_4 x_3$$

U ovome su modelu varijable $x_{2,2}, x_{2,3} \in \{0, 1\}$ indikatorske (engl. dummy) varijable koje predstavljaju razine obrazovanja. Konkretno, $x_{2,2}$ će imati vrijednost 1 za srednju školu (tj. kada je $x_2 = 2$) i 0 za ostale razine obrazovanja, dok će $x_{2,3}$ imati vrijednost 1 ako je razina obrazovanja fakultet (tj. kada je $x_2 = 3$), a u suprotnom 0. Osnovna je škola referentna kategorija, pa se efekti srednje škole i fakulteta mjere u odnosu na osnovnu školu. Model nam tako omogućuje procjenu utjecaja različitih razina obrazovanja na visinu plaće, uzimajući u obzir i kako dob te broj godina iskustva doprinosi plaći.

U ostatku rada linearni ćemo predviditelj zapisivati u obliku (2.10), pri čemu podrazumijevamo da to može uključivati sve moguće transformacije prediktora, kao što su interakcije i nelinearni odnosi, prikazane u prethodnim primjerima.

2.3 Funkcija poveznica

Funkcija poveznica igra ključnu ulogu u povezivanju očekivane vrijednosti zavisne varijable s linearnim predviditeljem. Ona omogućuje modelima da se prilagode različitim distribucijama podataka, poput binomne, eksponencijalne ili Poissonove distribucije, jer uspostavlja odnos između parametara distribucije i linearnog predviditelja kroz specifičnu funkciju.

Definicija 2.3.1. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *vektor nezavisnih slučajnih varijabli za koje pretpostavljamo da ovise o vrijednostima* $x_1, \ldots, x_p$*. Nadalje, neka je* $\eta_i = \boldsymbol{x_i}^T\boldsymbol{\beta}$ *linearni predviditelj i* $\mu_i = \mathbb{E}[Y_i]$*, za* $i = 1, \ldots, n$*. Monotono diferencijabilnu funkciju* $g : \mathbb{R} \to \mathbb{R}$ *takvu da je*

$$\eta_i = g(\mu_i), \quad i = 1, \ldots, n \tag{2.11}$$

nazivamo **funkcijom poveznicom**.

Primjer 2.3.1. *(Normalni linearni modeli.) Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *vektor nezavisnih slučajnih varijabli koje ovise o vrijednostima* $x_1, \ldots, x_p$ *tako da za* $i = 1, \ldots, n$ *vrijedi*

$$\begin{cases} Y_i \sim N(\mu_i, \sigma^2) \\ \mu_i = \boldsymbol{x}_i^T\boldsymbol{\beta}. \end{cases}$$

Ovdje je funkcija poveznica funkcija identiteta, tj. $g \equiv id$.

U praksi često želimo razumjeti kako linearni predviditelj utječe na očekivanu vrijednost zavisne varijable pa bi iz (2.11) bilo korisno izračunati μ_i. Kako svaka monotono diferencijabila funkcija ima svoj inverz, tako za funkciju g iz prethodne definicije postoji inverz $g^{-1} : \mathbb{R} \to \mathbb{R}$ takav da

$$\mu_i = g^{-1}(\eta_i). \tag{2.12}$$

Ovaj nam je rezultat koristan jer omogućuje interpretaciju rezultata modela u smislu očekivane vrijednosti zavisne varijable, a ne samo u smislu linearnog predviditelja.

Primjer 2.3.2. *Vratimo se na primjer 1.3.1 te pretpostavimo da su* $Y_i \sim B(m, p_i)$, $p_i \in \langle 0, 1\rangle$, $m \in \mathbb{N}$ *i* $\mathbb{E}[Y_i] = \mu_i = mp_i$*. U tom smo primjeru zapravo pokazali da funkcija poveznica ne može biti funkcija identiteta jer bi tada iz jednadžbe* (1.34) *slijedilo da vjerojatnosti* p_i *mogu biti negativne ili veće od jedan, što znamo da nije moguće. Sada ćemo pokazati da ovaj problem možemo riješiti tako da umjesto funkcije identiteta uzmemo odgovarajuću funkciju poveznicu.*
Funkciju $g : \langle 0, m\rangle \to \mathbb{R}$ *definiranu s*

$$g(x) = \log \frac{x}{m - x}$$

možemo uzeti kao funkciju poveznicu i pretpostaviti da vrijedi (2.11)*, tj.*

$$g(mp_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

Kako je funkcija g bijekcija, onda postoji inverz $g^{-1} : \mathbb{R} \to \langle 0, m\rangle$ *koji je dan s*

$$g^{-1}(x) = \frac{me^x}{1 + e^x},$$

odakle iz (2.12) *slijedi*

$$mp_i = g^{-1}(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \frac{me^{\boldsymbol{x}_i^T \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}} \in \langle 0, m\rangle.$$

Vidimo da smo korištenjem odgovarajuće funkcije poveznice osigurali da očekivane vrijednosti ostanu unutar intervala $\langle 0, m\rangle$ *što smo i htjeli postići. Dakle, funkcija poveznica nam je omogućila da binomnu zavisnu varijablu modeliramo kao odgovarajuću funkciju linearnog predviditelja.*

Općenito, bilo koja monotono diferencijabilna funkcija može biti funkcija poveznica, ali praktičnost i razumljivost modela ključni su pri njenom izboru. Važno je razmotriti kako ta funkcija odgovara prirodi podataka i vrsti distribucije zavisne varijable. Izbor funkcije poveznice može utjecati na preciznost i stabilnost rezultata, pa je često najbolje koristiti standardne funkcije poveznice prikazane u tablici 2.2 jer olakšavaju razumijevanje modela i daju pouzdanije rezultate.

Distribucija	$\mathbb{E}[Y] = \mu$	**Standardna funkcija poveznica** $g(\mu)$
$Y \sim \text{Bernoulli}(p), p \in \langle 0, 1\rangle$	$\mu = p$	$\log \frac{\mu}{1-\mu}$
$Y \sim B(n, p), p \in \langle 0, 1\rangle, n \in \mathbb{N}$ poznat	$\mu = np$	$\log \frac{\mu}{n-\mu}$
$Y \sim P(\lambda), \lambda > 0$	$\mu = \lambda$	$\log \mu$
$Y \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$ poznat	$\mu = \mu$	μ
$Y \sim \text{Exp}(\lambda), \lambda > 0$	$\mu = \frac{1}{\lambda}$	$\log \mu$
$Y \sim \Gamma(\alpha, \beta), \beta > 0, \alpha > 0$ poznata	$\mu = \frac{\alpha}{\beta}$	$\log \frac{\mu}{\alpha}$

Tablica 2.2: Standardne funkcije poveznice nekih distribucija

2.4 Formulacija generaliziranih linearnih modela

Prisjetimo se, u uvodu ovoga poglavlja naglasili smo da generalizirani linearni modeli imaju tri važne komponente: varijabla odziva pripada eksponencijalnoj familiji distribucija, linearni predviditelj opisuje kako prediktori utječu na distribuciju zavisne varijable te

funkcija poveznica koja povezuje očekivanu vrijednost zavisne varijable s linearnim predviditeljem. Sada možemo dati i formalnu definiciju generaliziranih linearnih modela koju ćemo koristiti dalje u radu.

Definicija 2.4.1. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *vektor nezavisnih slučajnih varijabli koje ovise o vrijednostima* $x_1, \ldots, x_p$*. Također, neka je* $g : \mathbb{R} \to \mathbb{R}$ *funkcija poveznica te* $\boldsymbol{x}_i^T\boldsymbol{\beta}$ *linearni predviditelj za* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$*.* **Generalizirani linearni modeli** *su modeli koje možemo prikazati na sljedeći način:*

$$\begin{cases} Y_i \overset{nez.}{\sim} EFD(\theta_i) \\ \mu_i = g^{-1}(\boldsymbol{x}_i^T\boldsymbol{\beta}), \quad i = 1, \ldots, n. \end{cases} \tag{2.13}$$

Primjer 2.4.1. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *vektor nezavisnih slučajnih varijabli koje ovise o vrijednostima* $x_1, \ldots, x_p$ *te* $\boldsymbol{x}_i^T\boldsymbol{\beta}$ *linearni predviditelj. Za* $i = 1, \ldots, n$ *su sljedeći modeli generalizirani linearni modeli:*

(i) $\begin{cases} Y_i \sim \text{Bernoulli}(p_i) \\ \mu_i = p_i = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} \end{cases}$ *(iii)* $\begin{cases} Y_i \sim P(\lambda_i) \\ \mu_i = \lambda_i = e^{\boldsymbol{x}_i^T\boldsymbol{\beta}} \end{cases}$ *(v)* $\begin{cases} Y_i \sim \text{Exp}(\lambda_i) \\ \mu_i = \frac{1}{\lambda_i} = e^{\boldsymbol{x}_i^T\boldsymbol{\beta}} \end{cases}$

(ii) $\begin{cases} Y_i \sim B(m, p_i) \\ \mu_i = mp_i = \frac{me^{x_i^T\beta}}{1+e^{x_i^T\beta}} \end{cases}$ *(iv)* $\begin{cases} Y_i \sim N(\mu_i, \sigma^2) \\ \mu_i = \boldsymbol{x}_i^T\boldsymbol{\beta}. \end{cases}$ *(vi)* $\begin{cases} Y_i \sim \Gamma(\alpha, \lambda_i) \\ \mu_i = \frac{\alpha}{\lambda_i} = \alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}} \end{cases}$

Da bismo mogli predviđati očekivane vrijednosti zavisne varijable, ključno je znati vrijednosti parametara $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. Međutim, u praksi te parametre ne poznajemo unaprijed te ih stoga moramo procijeniti koristeći statističke metode. U sljedećem ćemo se poglavlju fokusirati na procjenu parametara generaliziranih linearnih modela primjenom *metode najveće vjerodostojnosti*. Ova nam metoda omogućava da na temelju dostupnih podataka odredimo parametre koji najbolje objašnjavaju promatrane podatke.

Poglavlje 3

Procjena parametara

Nakon što smo definirali generalizirane linearne modele i razumjeli ključne komponente koje ih čine, sljedeći je korak odrediti kako ćemo procijeniti parametre $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ ovih modela. Metoda najveće vjerodostojnosti predstavlja jedan od najpouzdanijih i najčešće korištenih pristupa u statistici za procjenu nepoznatih parametara u modelima.

Ovo ćemo poglavlje započeti kratkim pregledom osnovnih principa metode najveće vjerodostojnosti i ključnih svojstava koje ćemo koristiti. Zatim ćemo tu metodu primijeniti na generalizirane linearne modele, kako bismo dobili precizne procjene parametara.

3.1 Metoda najveće vjerodostojnosti

Metoda najveće vjerodostojnosti je statistička tehnika koja se koristi za procjenu parametara određenih vjerojatnosnih distribucija na temelju dostupnih podataka.

Ponavljanjem eksperimenta n puta dolazimo do skupa podataka $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ pri čemu svaki podatak y_i predstavlja realizaciju slučajne varijable Y_i, za $i = 1, \ldots, n$, koja modelira naš eksperiment. Pretpostavimo da su sva ponavljanja eksperimenta međusobno nezavisna te da pripadaju istoj klasi distribucija. Nadalje, neka gustoće f_{Y_i} slučajnih varijabli Y_i ovise o nepoznatom parametru θ_i, za $i = 1, \ldots, n$. Važno je naglasiti da parametri θ_i, mogu, ali i ne moraju biti međusobno jednaki.

Da bismo procijenili parametar distribucije, važno je uzeti u obzir sve opservacije y_i. Zbog toga ćemo definirati funkciju gustoće slučajnog vektora $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, koja opisuje zajedničku distribuciju svih Y_i.

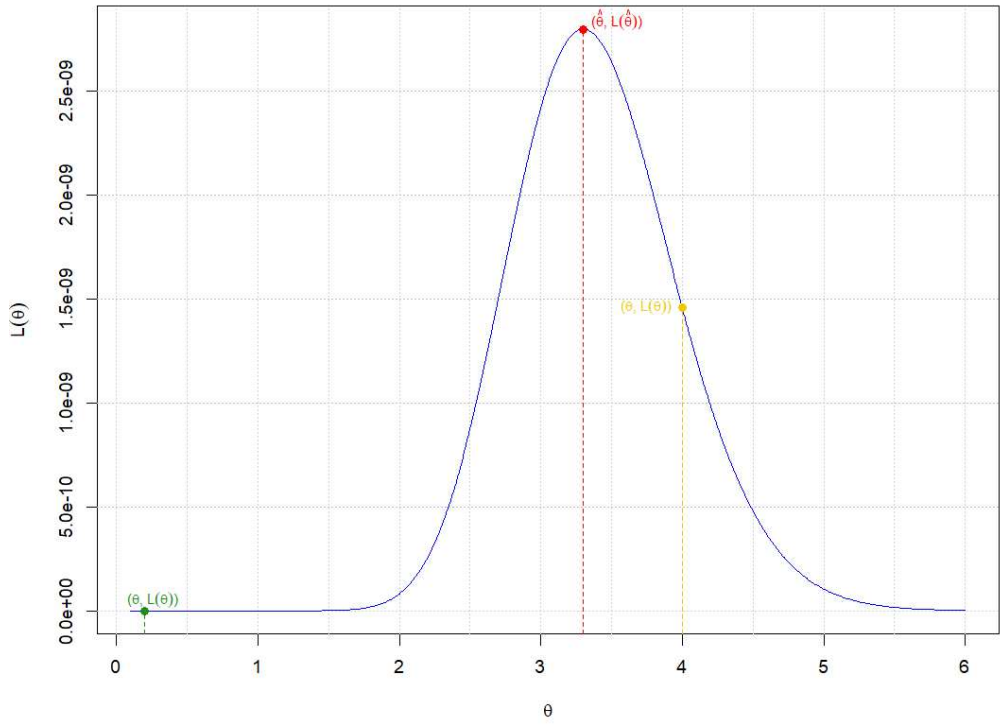
S obzirom na nezavisnost slučajnih varijabli Y_i, funkcija gustoće slučajnog vektora $\boldsymbol{Y}$ može se izraziti kao produkt marginalnih funkcija gustoće, tj.

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) \overset{nez.}{=} \prod_{i=1}^{n} f_{Y_i}(y_i; \theta_i),$$

gdje je $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ opaženi uzorak, a $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ vektor parametara.

Ideja metode najveće vjerodostojnosti je definirati funkciju vjerodostojnosti koja će biti funkcija parametra od interesa $\boldsymbol{\theta}$ te koja će procijeniti koja je najvjerojatnija vrijednost parametra koja opisuje opažene podatke $\boldsymbol{y}$. Drugim riječima, tražit ćemo maksimum te funkcije.

Graf na slici 3.1 prikazuje ideju procjene parametra θ metodom najveće vjerodostojnosti u jednodimenzionalnom slučaju, vizualizirajući funkciju vjerodostojnosti i naglašavajući ključne točke.



Slika 3.1: Metoda najveće vjerodostojnosti nalazi točku $\hat{\boldsymbol{\theta}}$ koja je najvjerodostojnija obzirom na opažene podatke $\boldsymbol{y}$.

Funkcija vjerodostojnosti i log-vjerodostojnosti

Definicije navedene u ovome poglavlju prate [7].

Definicija 3.1.1. *Za n-dimenzionalni slučajni vektor* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *nezavisnih slučajnih varijabli sa zajedničkom gustoćom* $f_{\boldsymbol{Y}}$ *koja ovisi o parametru* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ *funkciju* $L : \Theta^n \to \mathbb{R}$ *definiranu s*

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{Y_i}(y_i; \theta_i), \tag{3.1}$$

nazivamo **vjerodostojnost** *parametra* $\boldsymbol{\theta}$ *s obzirom na opaženi uzorak* $\boldsymbol{y} = (y_1, \ldots, y_n)^T$.

Nadalje, **procjenitelj najveće vjerodostojnosti** *parametra* $\boldsymbol{\theta}$ *je statistika* $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ *koja maksimizira funkciju vjerodostojnosti, tj.*

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta^n} L(\boldsymbol{\theta}\,; \boldsymbol{Y}),$$

dok je $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ **procjena metodom najveće vjerodostojnosti** *od* $\boldsymbol{\theta}$ *s obzirom na opaženi uzorak* $\boldsymbol{y}$.

Vidimo da je funkcija vjerodostojnosti ista kao i funkcija gustoće slučajnog vektora $\boldsymbol{Y}$, samo što sada promatramo parametar $\boldsymbol{\theta}$, a $\boldsymbol{y}$ smatramo fiksnim.

U praksi se često koristi prirodni logaritam funkcije vjerodostojnosti, koji se naziva log-vjerodostojnost.

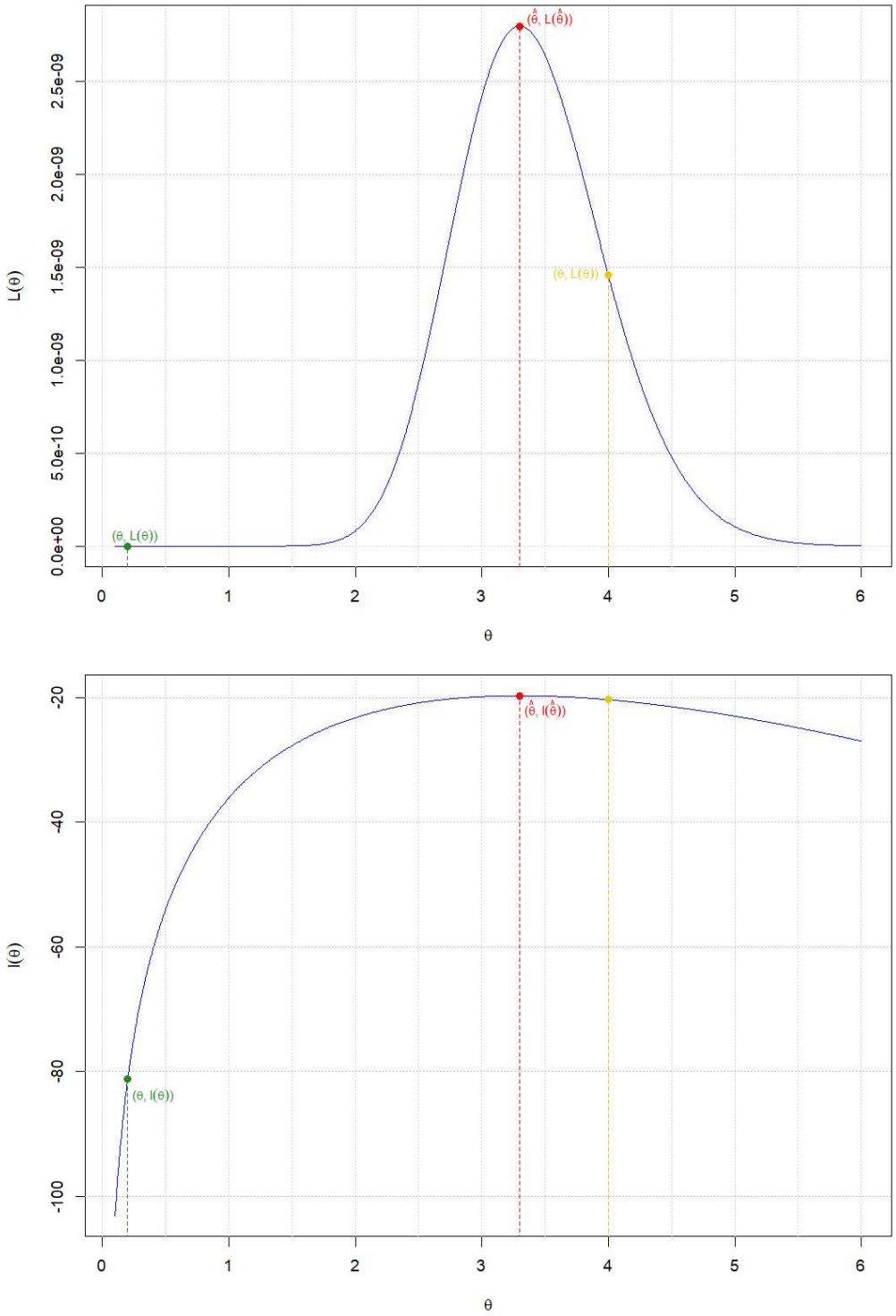
Definicija 3.1.2. *Za n-dimenzionalni slučajni vektor* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *nezavisnih slučajnih varijabli sa zajedničkom gustoćom* $f_{\boldsymbol{Y}}$ *koja ovisi o parametru* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ *funkciju* $l : \Theta^n \to \mathbb{R}$ *definiranu s*

$$l(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}) = \log \left[\prod_{i=1}^{n} f_{Y_i}(y_i; \theta_i) \right] = \sum_{i=1}^{n} \log f_{Y_i}(y_i; \theta_i). \tag{3.2}$$

nazivamo **log-vjerodostojnost** *parametra* $\boldsymbol{\theta}$ *s obzirom na opaženi uzorak* $\boldsymbol{y}$.

Traženje maksimuma funkcije vjerodostojnosti

Traženje maksimuma log-vjerodostojnosti često je jednostavnije od traženja maksimuma funkcije vjerodostojnosti, budući da derivacija sume iz jednadžbe (3.2) često daje jednostavnije jednadžbe nego derivacija produkta iz jednadžbe (3.1). Kako je logaritamska funkcija monotona, ona raste i opada u istim točkama kao i funkcija vjerodostojnosti pa je maksimum funkcije log-vjerodostojnosti uvijek u istoj točki kao i maksimum funkcije vjerodostojnosti. U to se možemo uvjeriti na slici 3.2.



Slika 3.2: Funkcije vjerodostojnosti i log-vjerodostojnosti daju istu procjenu parametra $\boldsymbol{\theta}$

Iz tih ćemo se razloga ovdje fokusirati na traženje maksimuma funkcije log-vjerodostojnosti, tj. tražimo vektor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)^T$ koji zadovoljava

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta^n} l(\boldsymbol{\theta}\,; \boldsymbol{y}).$$

Maksimizacija log-vjerodostojnosti često se svodi na analizu njezinih parcijalnih derivacija. U okviru teorije analize, izračunavanje prve derivacije funkcije log-vjerodostojnosti omogućuje nam identifikaciju kritičnih točaka, odnosno onih točaka gdje bi se mogao nalaziti maksimum funkcije.

Prema osnovnim teoretskim rezultatima analize, procjena maksimalne vjerodostojnosti parametra $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ je vektor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)^T$ koji je rješenje sustava jednadžbi

$$\left.\frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad i = 1, \ldots, n. \tag{3.3}$$

Da bismo se uvjerili da se zaista radi o točki maksimuma, potrebno je utvrditi da je funkcija log-vjerodostojnosti konkavna funkcija. Važno je napomenuti da su log-vjerodostojnosti većine vjerojatnosnih distribucija, posebno onih iz eksponencijalne familije, konkavne. U ovome radu nećemo u detalje ulaziti u dokaz ovog svojstva.

U nekim je slučajevima moguće eksplicitno riješiti sustav (3.3). Međutim, u većini slučajeva situacija nije tako jednostavna. Često se suočavamo s modelima za koje ne postoji analitičko rješenje za procjenu parametara. U takvim situacijama koristimo numeričke metode optimizacije koje nam omogućuju da aproksimiramo rješenja tako što iterativno tražimo maksimalnu vrijednost funkcije vjerodostojnosti. U nastavku ćemo opisati kako dolazimo do procjene parametra $\boldsymbol{\theta}$ tzv. *Fisherovom iterativnom metodom*.

U tu svrhu definirat ćemo pojmove *skor statistike* i *Fisherove informacijske matrice*.

Definicija 3.1.3. *Za n-dimenzionalni slučajni vektor $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ nezavisnih slučajnih varijabli sa zajedničkom gustoćom $f_{\boldsymbol{Y}}$ koja ovisi o parametru $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$,* **skor statistika** *u parametru θ_i je slučajna varijabla definirana s*

$$U_i = U(\theta_i) = \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial \theta_i}, \quad i = 1, \ldots, n. \tag{3.4}$$

U nastavku ćemo pokazati da očekivana vrijednost skor statistike za pravu vrijednost parametra $\boldsymbol{\theta}$ iznosi 0.

Propozicija 3.1.1. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ niz nezavisnih slučajnih varijabli sa zajedničkom gustoćom $f_{\boldsymbol{Y}}$ koja ovisi o parametru $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$. Tada je za sve $i = 1, \ldots, n$ očekivanje skor statistike $U_i = U(\theta_i)$ u pravoj vrijednosti parametra θ_i jednako nuli, odnosno vrijedi*

$$\mathbb{E}[U_i] = 0, \quad za\ i = 1, \ldots, n. \tag{3.5}$$

Dokaz. Neka je U_i za $i = 1, \ldots, n$ skor statistika definirana s (3.4). Iz definicije matematičkog očekivanja slučajne varijable U_i slijedi

$$\mathbb{E}[U_i] = \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_i} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y}. \tag{3.6}$$

Nadalje, kako je $l(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}) = \log f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, onda je

$$\frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_i} = \frac{1}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_i}. \tag{3.7}$$

Konačno, slijedi

$$\begin{aligned}
\mathbb{E}[U_i] &\overset{(3.6)}{=} \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_i} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} \\
&\overset{(3.7)}{=} \int_{\mathbb{R}^n} \frac{1}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_i} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} \\
&= \int_{\mathbb{R}^n} \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_i}\, d\boldsymbol{y} \\
&= \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} \\
&= \frac{\partial}{\partial \theta_i} 1 \\
&= 0,
\end{aligned}$$

pri čemu pretpostavljamo da možemo mijenjati poredak diferenciranja i integriranja kao do sada. □

Definicija 3.1.4. *Za n-dimenzionalni slučajni vektor $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ nezavisnih slučajnih varijabli sa zajedničkom gustoćom $f_{\boldsymbol{Y}}$ koja ovisi o parametru $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$* **Fisherova informacijska matrica** *$\mathcal{I} = (\mathcal{I}_{jk})_{j,k=1,\ldots,n}$ je $n \times n$ matrica kojoj su elementi dani s*

$$\mathcal{I}_{jk} = \mathrm{Cov}(U_j, U_k), \quad j, k = 1, \ldots, n. \tag{3.8}$$

Budući da očekivanje skora za pravu vrijednost parametra $\boldsymbol{\theta}$ iznosi nula po prethodnoj propoziciji, jednažbu (3.8) možemo pojednostavniti

$$\begin{aligned}
\mathcal{I}_{jk} &= \mathrm{Cov}(U_j, U_k) \\
&\overset{(1.12)}{=} \mathbb{E}[U_j U_k] - \mathbb{E}[U_j]\mathbb{E}[U_k] \\
&\overset{(3.5)}{=} \mathbb{E}[U_j U_k] \qquad (3.9) \\
&\overset{(3.4)}{=} \mathbb{E}\left[\frac{\partial l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial \theta_j} \cdot \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial \theta_k}\right]. \qquad (3.10)
\end{aligned}$$

Propozicija 3.1.2. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *niz nezavisnih slučajnih varijabli sa zajedničkom gustoćom* $f_{\boldsymbol{Y}}$ *koja ovisi o parametru* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ *te* $\mathcal{I} = (\mathcal{I}_{jk})_{j,k=1,\ldots,n}$ *Fisherova informacijska matrica. Elementi Fisherove informacijske matrice mogu se izraziti kao*

$$\mathcal{I}_{jk} = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial\theta_j \partial\theta_k}\right], \quad j, k = 1, \ldots, n. \tag{3.11}$$

Dokaz. Iz definicije matematičkog očekivanja slučajne varijable U_k i prethodne propozicije znamo da za sve $k = 1, \ldots, n$ vrijedi

$$\mathbb{E}[U_k] = \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} = 0.$$

Diferenciramo gornju jednakost po θ_j i pretpostavimo da možemo zamijeniti redoslijed diferenciranja i integriranja. Tada imamo

$$\int_{\mathbb{R}^n} \frac{\partial}{\partial\theta_j}\left[\frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\right] d\boldsymbol{y} = 0,$$

odnosno

$$\int_{\mathbb{R}^n} \left[\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_j \partial\theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) + \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial\theta_j}\right] d\boldsymbol{y} = 0,$$

odakle iz svojstva linearnosti integrala slijedi

$$\int_{\mathbb{R}^n} \frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_j \theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} + \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial\theta_j}\, d\boldsymbol{y} = 0.$$

Dakle, imamo

$$\int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial\theta_j}\, d\boldsymbol{y} = -\int_{\mathbb{R}^n} \frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_j \theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y}. \tag{3.12}$$

Primijetimo da je

$$\int_{\mathbb{R}^n} \frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_j \theta_k} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y} = \mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial\theta_j \theta_k}\right].$$

Da bismo dokazali tvrdnju propozicije preostaje pokazati da lijeva strana jednakosti (3.12) iznosi $\mathcal{I}_{jk}$, tj.

$$\int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial\theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})}{\partial\theta_j}\, d\boldsymbol{y} = \mathcal{I}_{jk}.$$

Naime,

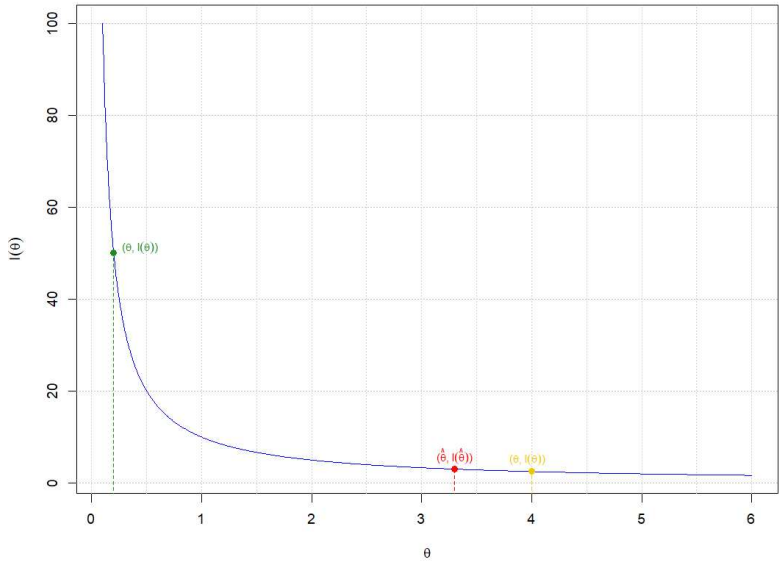
$$\begin{aligned}\int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta};\boldsymbol{y})}{\partial \theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}{\partial \theta_j}\, d\boldsymbol{y} &= \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta};\boldsymbol{y})}{\partial \theta_k} \cdot \frac{\partial f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}{\partial \theta_j} \cdot \frac{f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}{f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}\, d\boldsymbol{y} \\ &\overset{(3.7)}{=} \int_{\mathbb{R}^n} \frac{\partial l(\boldsymbol{\theta};\boldsymbol{y})}{\partial \theta_k} \cdot \frac{\partial l(\boldsymbol{\theta};\boldsymbol{y})}{\partial \theta_j} \cdot f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})\, dy \\ &= \mathbb{E}\left[\frac{\partial l(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \theta_j} \cdot \frac{\partial l(\boldsymbol{\theta};\boldsymbol{Y})}{\partial \theta_k}\right] \overset{(3.10)}{=} \mathcal{I}_{jk}.\end{aligned}$$

□

Fisherova informacijska matrica daje nam uvid u količinu informacija koju skup podataka $\boldsymbol{y}$ nosi o nepoznatom parametru $\boldsymbol{\theta}$.

Primjerice, na slici 3.3 prikazan je primjer funkcije Fisherove informacije $\mathcal{I}(\theta)$ u odnosu na jednodimenzionalni parametar θ. Graf funkcije pokazuje da se količina informacija o parametru θ značajno mijenja kada su vrijednosti θ male, dok se za veće vrijednosti θ količina informacija stabilizira i gotovo ne varira. To znači da su male promjene u parametru θ u blizini manjih vrijednosti povezane s velikim promjenama u informacijama koje možemo izvući iz podataka, dok su promjene u informacijama manje izražene kada θ postane veća.

Također, visoka Fisherova informacija $\mathcal{I}(\theta)$ implicira da je $\left|\frac{\partial^2 l(\theta;\boldsymbol{y})}{\partial \theta^2}\right|$ visoka, što znači da se log-vjerodostojnost brzo mijenja u okolini tog parametra. To sugerira oštriji vrh funkcije log-vjerodostojnosti i samim time precizniju procjenu parametra.

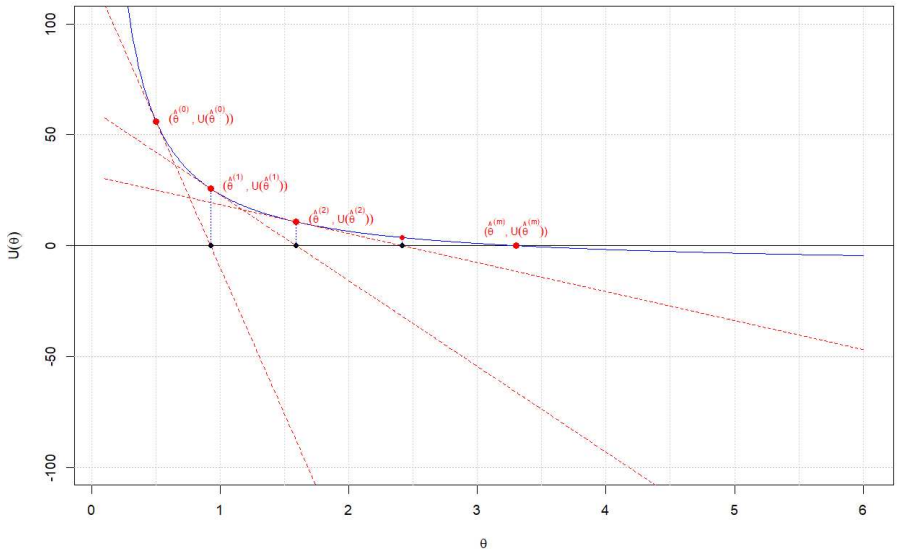


Slika 3.3: Graf funkcije Fisherove informacije $\mathcal{I}(\theta)$ u odnosu na parametar θ

Vratimo se sada na sustav

$$U_i = U(\theta_i) = \frac{\partial l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_i} = 0, \quad i = 1, \ldots, n.$$

Vidjeli smo da ovaj sustav predstavlja uvjet za maksimizaciju funkcije log-vjerodostojnosti, ali pretpostavimo da ne možemo analitički pronaći rješenje tog sustava, nego ga tražimo iterativno pomoću Newton-Rapshonovog algoritma.



Slika 3.4: Newton-Raphsonov algoritam u jednodimenzionalnom slučaju

Na slici 3.4 prikazana je ilustracija Newton-Raphsonove metode, koja pokazuje postupak iterativnog približavanja nultočki funkcije. Naime, u jednodimenzionalnom slučaju, tražimo $\hat{\theta}$ takav da $U(\hat{\theta}) = 0$. Metoda počinje odabirom početne točke $\hat{\theta}^{(0)}$ u blizini očekivane nultočke i crtanjem tangente na graf funkcije U u odabranoj točki. Ova tangenta daje novu aproksimaciju $\hat{\theta}^{(1)}$ nultočke kao sjecište tangente s osi apscisa. Algoritam se nastavlja sve dok apsolutna razlika između uzastopnih aproksimacija ne padne ispod unaprijed definiranog praga ϵ, čime se osigurava da su iteracije dovoljno blizu stvarnoj nultočki funkcije.

Sada, s obzirom na to da želimo riješiti sustav od n jednadžbi s n nepoznanica, Newton-Raphsonov algoritam generaliziramo na vektorske funkcije. U višedimenzionalnom slučaju, funkcija $\boldsymbol{U} = U(\boldsymbol{\theta}; \boldsymbol{y}) = (U_1, \ldots, U_n)^T$ postaje vektorska funkcija sa n komponenti, a umjesto tangente koristimo Jacobijevu $n \times n$ matricu $\mathcal{J} = (\mathcal{J}_{jk})_{j,k=1,\ldots,n}$ koja sadrži sve parcijalne derivacije funkcije $\boldsymbol{U} = (U_1, \ldots, U_n)^T$ s obzirom na parametre $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$, tj. elementi Jacobijeve matrice su oblika:

$$\mathcal{J}_{jk} = \frac{\partial U_j}{\partial \theta_k} \overset{(3,4)}{=} \frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \theta_k \partial \theta_j}, \quad j, k = 1, \ldots, n. \tag{3.13}$$

Za $m \in \mathbb{N}$, u m-toj iteraciji, umjesto pronalaženja nultočke tangente kao u jednodimenzionalnom slučaju, izračunavamo novu aproksimaciju za procjenu $\hat{\boldsymbol{\theta}}^{(m)} = (\hat{\theta}_1^{(m)}, \ldots, \hat{\theta}_n^{(m)})^T$ rješavanjem linearnog sustava:

$$\hat{\boldsymbol{\theta}}^{(m)} = \hat{\boldsymbol{\theta}}^{(m-1)} - [\mathcal{J}^{(m-1)}]^{-1} \boldsymbol{U}^{(m-1)},$$

gdje su $\boldsymbol{U}^{(m-1)} = (U_1^{(m-1)}, \ldots, U_n^{(m-1)})^T$ takvi da $U_i^{(m-1)} = U(\hat{\theta}_i^{(m-1)})$ za sve $i = 1, \ldots, n$ te $\mathcal{J}^{(m-1)} = (\mathcal{J}_{jk}^{(m-1)})_{j,k=1,\ldots,n}$ Jacobijeva matrica s elementima iz (3.13) evaluirana u točki $\hat{\boldsymbol{\theta}}^{(m-1)}$, tj.

$$\mathcal{J}_{jk}^{(m-1)} = \left.\frac{\partial U_i}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(m-1)}}.$$

Iteracije se ponavljaju dok norma razlika između uzastopnih aproksimacija ne postane dovoljno mala, tj. dok ne postignemo željenu točnost $\boldsymbol{\epsilon}$.

Kada se procjenjuju parametri metodom najveće vjerodostojnosti, često se koristi Fisherova informacijska matrica umjesto Jacobijeve matrice za poboljšanje procjena i bržu konvergenciju algoritma.

Konkretno, kod Newton-Raphsonovog algoritma, Jacobijeva matrica $\mathcal{J}^{(m-1)}$ iz prethodnih iteracija može se zamijeniti s Fisherovom informacijskom matricom $\mathcal{I}^{(m-1)} = (I_{jk}^{(m-1)})_{j,k=1,\ldots,n}$, jer vrijedi

$$\mathcal{I}_{jk} \overset{(3.11)}{=} -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial\theta_j \partial\theta_k}\right] \overset{(3.13)}{=} -\mathbb{E}[\mathcal{J}_{jk}], \quad j,k = 1, \ldots, n \tag{3.14}$$

pa se m-ta iteracija može izraziti kao:

$$\hat{\boldsymbol{\theta}}^{(m)} = \hat{\boldsymbol{\theta}}^{(m-1)} + [\mathcal{I}^{(m-1)}]^{-1}\boldsymbol{U}^{(m-1)}. \tag{3.15}$$

Ova metoda, poznata kao **Fisherova iterativna metoda**, služi kao temelj za procjenu parametara u generaliziranim linearnih modelima.

3.2 Procjena parametara generaliziranog linearnog modela

Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ niz nezavisnih slučajnih varijabli za koje pretpostavljamo da ovise o vrijednostima $x_1, \ldots, x_p$. Nadalje, pretpostavimo da $\boldsymbol{Y}$ zadovoljava pretpostavke generaliziranih linearnih modela, tj. za $i = 1, \ldots, n$ vrijedi

$$\begin{cases} Y_i \overset{nez.}{\sim} EFD(\theta_i) \\ \mu_i = \mathbb{E}[Y_i] = g^{-1}(\boldsymbol{x}_i^T\boldsymbol{\beta}), \end{cases} \tag{3.16}$$

gdje su $\boldsymbol{x}_i^T = (1, x_{i1}, \ldots, x_{ip})$ te $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ vektor parametara. Kako su parametri $\boldsymbol{\beta}$ nepoznati, želimo ih procijeniti metodom najveće vjerodostojnosti kako je opisano u prethodnom odjeljku. Dakle, moramo definirati funkciju log-vjerodostojnosti te pronaći parametre $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ takve da

$$\left.\frac{\partial \boldsymbol{l}}{\partial \beta_j}\right|_{\boldsymbol{\beta}=\boldsymbol{b}} = 0, \quad j = 0, 1, \ldots, p$$

korištenjem analitčkog pristupa ili Fisherovom iterativnom metodom. U tu ćemo se svrhu, slijedeći [1], u nastavku fokusirati na derivaciju formula za funkciju gustoće slučajnog vektora $\boldsymbol{Y}$, kao i na formulacije funkcije vjerodostojnosti, log-vjerodostojnosti, skor funkcija i Fisherove informacijske matrice za generalizirane linearne modele. Ovi će nam rezultati poslužiti kao osnova za procjenu parametara u našem modelu od interesa.

Funkcija gustoće

Kako je po pretpostavci generaliziranih linearnih modela $Y_i \sim EFD(\theta_i)$, za $i = 1, \ldots, n$ su funkcije gustoće svake slučajne varijable Y_i dane s

$$f_{Y_i}(y_i; \theta_i) = \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right] \tag{3.17}$$

pa je zajednička gustoća $f_{\boldsymbol{Y}}$ dana s

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) &\overset{nez.}{=} \prod_{i=1}^{n} f_{Y_i}(y_i; \theta_i) \\ &= \prod_{i=1}^{n} \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right] \\ &= \exp\left[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)\right]. \end{aligned} \tag{3.18}$$

Funkcija vjerodostojnosti i log-vjerodostojnosti

Za $i = 1, \ldots, n$ vjerodostojnost parametra θ_i je funkcija

$$L_i = L(\theta_i; y_i) = f_{Y_i}(y_i; \theta_i) \overset{(3.17)}{=} \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right], \tag{3.19}$$

dok je vjerodostojnost parametra $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ s obzirom na opaženi uzorak $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ dana s

$$\boldsymbol{L} = L(\boldsymbol{\theta}; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) \overset{(3.18)}{=} \exp\left[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)\right]. \tag{3.20}$$

Nadalje, za $i = 1, \ldots, n$ log-vjerodostojnost parametra θ_i je oblika

$$l_i = l(\theta_i; y_i) = \log L(\theta_i; y_i) \overset{(3.19)}{=} y_i b(\theta_i) + c(\theta_i) + d(y_i), \tag{3.21}$$

dok je log-vjerodostojnost parametra $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ s obzirom na opaženi uzorak $\boldsymbol{y}$

$$\boldsymbol{l} = l(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}) \overset{(3.20)}{=} \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i).$$

Napomena 3.2.1. *U nastavku ćemo istraživati oblik skor funkcije i Fisherove informacijske matrice generaliziranih linearnih modela. S obzirom da smo u vektor parametara* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ *uključili presječnu točku (engl. intercept)* β_0 *kao prvi parametar te samim time promijenili indeksiranje tako da počinje od 0 umjesto 1, prilagodit ćemo indeksiranje i za skor funkciju i Fisherovu informacijsku matricu. Ova prilagođena notacija olakšava jasno predstavljanje svih parametara u analizi.*

Skor funkcija

Nakon određivanja log-vjerodostojnosti, potrebno je pronaći parcijalne derivacije funkcije log-vjerodostojnosti s obzirom na parametre $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ kako bismo pronašli točke $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ u kojima funkcija doseže svoj maksimum. Budući da funkcija log-vjerodostojnosti ovisi o parametrima β_j kroz različite funkcije, potrebno je primijeniti lančano pravilo za deriviranje kojim dobivamo sljedeći izraz za skor funkcije U_j u parametrima β_j

$$U_j = U(\beta_j) = \frac{\partial \boldsymbol{l}}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} \right], \quad j = 0, 1, \ldots, p, \tag{3.22}$$

pri čemu smo u trećoj jednakosti koristili svojstvo linearnosti parcijalne derivacije. Svaku parcijalnu derivaciju s desne strane jednakosti razmotrit ćemo sada zasebno. Prvo, vrijedi

$$\frac{\partial l_i}{\partial \theta_i} \overset{(3.21)}{=} y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i) \left(y_i + \frac{c'(\theta_i)}{b'(\theta_i)} \right) \overset{(2.5)}{=} b'(\theta_i) \left(y_i - \mu_i \right). \tag{3.23}$$

Nadalje,

$$\frac{\partial \mu_i}{\partial \theta_i} \overset{(2.5)}{=} \frac{c'(\theta_i) b''(\theta_i) - c''(\theta_i) b'(\theta_i)}{[b'(\theta_i)]^2} \overset{(2.6)}{=} b'(\theta_i) \mathrm{Var}(Y_i)$$

pa je

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b'(\theta_i) \mathrm{Var}(Y_i)}. \tag{3.24}$$

Na kraju, imamo

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}. \tag{3.25}$$

Sada iz (3.23), (3.24) i (3.25) uvrštavanjem u (3.22) dobivamo

$$U_j = \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{\mathrm{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right], \quad j = 0, 1, \ldots, p. \tag{3.26}$$

Fisherova informacijska matrica

Fisherova informacijska matrica $\mathcal{I} = (\mathcal{I}_{jk})_{j,k=0,1,\ldots,p}$ je $(p+1)\times(p+1)$ matrica čiji su elementi oblika

$$\begin{aligned}
\mathcal{I}_{jk} &\overset{(3.9)}{=} \mathbb{E}[U_j U_k] \overset{(3.26)}{=} \mathbb{E}\left\{\sum_{i=1}^{n}\left[\frac{Y_i-\mu_i}{\mathrm{Var}(Y_i)}x_{ij}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right]\sum_{l=1}^{n}\left[\frac{Y_l-\mu_l}{\mathrm{Var}(Y_l)}x_{lk}\left(\frac{\partial\mu_l}{\partial\eta_l}\right)\right]\right\} \\
&= \mathbb{E}\left\{\sum_{i=1}^{n}\sum_{l=1}^{n}\left[\frac{(Y_i-\mu_i)(Y_l-\mu_l)}{\mathrm{Var}(Y_i)\mathrm{Var}(Y_l)}x_{ij}x_{lk}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\mu_l}{\partial\eta_l}\right)\right]\right\} \\
&= \sum_{i=1}^{n}\sum_{l=1}^{n}\left[\frac{\mathbb{E}[(Y_i-\mu_i)(Y_l-\mu_l)]}{\mathrm{Var}(Y_i)\mathrm{Var}(Y_l)}x_{ij}x_{lk}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\mu_l}{\partial\eta_l}\right)\right] \\
&= \sum_{i=1}^{n}\left[\frac{\mathbb{E}[(Y_i-\mu_i)^2]}{(\mathrm{Var}(Y_i))^2}x_{ij}x_{ik}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\right] \\
&= \sum_{i=1}^{n}\frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2, \qquad (3.27)
\end{aligned}$$

gdje smo u petoj jednakosti za $l \neq i$ koristili svojstvo $\mathbb{E}[(Y_i-\mu_i)(Y_l-\mu_l)] = 0$ zbog nezavisnosti varijabli Y_i za $i = 1,\ldots,n$ što slijedi iz

$$\mathbb{E}[(Y_i-\mu_i)(Y_l-\mu_l)] = \mathbb{E}[Y_iY_l - \mu_lY_i - \mu_iY_l + \mu_i\mu_l] = \mathbb{E}[Y_iY_l] - \mathbb{E}[Y_i]\mathbb{E}[Y_l] \overset{(1.7)}{=} 0$$

te smo u šestoj jednakosti koristili definiciju varijance, tj. $\mathbb{E}[(Y_i-\mu_i)^2] = \mathrm{Var}(Y_i)$.

Maksimizacija funkcije log-vjerodostojnosti

Nakon određivanja skor funkcije i Fisherove informacijske matrice prelazimo na rješavanje sustava

$$\boldsymbol{U} = U(\boldsymbol{\beta}) = 0,$$

gdje je $\boldsymbol{U} = (U_0, U_1, \ldots, U_p)^T$ vektor skor funkcija definiranih s (3.26). Kao što smo u prethodnom odjeljku pokazali, ovisno o kompleksnosti skor funkcije rješenje sustava možemo naći analitički ili Fisherovom iterativnom metodom

$$\boldsymbol{b}^{(m)} = \boldsymbol{b}^{(m-1)} + \left[\mathcal{I}^{(m-1)}\right]^{-1}\boldsymbol{U}^{(m-1)} \qquad (3.28)$$

gdje su

(i) $\boldsymbol{b}^{(m)} = (b_0^{(m)}, b_1^{(m)}, \ldots, b_p^{(m)})^T$ vektor procjene parametara $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ u m-toj iteraciji,

(ii) $\boldsymbol{U}^{(m-1)} = (U_0^{(m-1)}, U_1^{(m-1)}, \ldots, U_p^{(m-1)})^T$ je vektor skor funkcija (3.26) evaluiran u $\boldsymbol{b}^{(m-1)}$,

(iii) $\mathcal{I}^{(m-1)} = (\mathcal{I}_{jk}^{(m-1)})_{j,k=0,1,\ldots,p}$ Fisherova informacijska matrica čiji su elementi dani s (3.27) evaluirani u $\boldsymbol{b}^{(m-1)}$.

Iterativna težinska metoda najmanjih kvadrata

Iterativna težinska metoda najmanjih kvadrata koristi se za poboljšanje procjena parametara putem Fisherovog algoritma. Ključne komponente ove metode uključuju težinsku matricu W, koja prilagođava utjecaj opažanja prema njihovoj pouzdanosti, i pseudo-odgovor z, koji pomaže modelu da se prilagodi na osnovi trenutnih procjena.

Težinska matrica $W = (W_{ij})_{i,j=1,\dots,n}$ je $n \times n$ matrica čiji su elementi definirani s

$$W_{ij} = \begin{cases} \frac{1}{\mathrm{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2, & i = j, \\ 0, & i \neq j \end{cases} \tag{3.29}$$

Ova formula pokazuje da je težinska matrica dijagonalna, pri čemu svaki njen dijagonalni element, W_{ii}, predstavlja koliko je važno svako pojedinačno opažanje Y_i. Opažanja s manjom varijancijom i većom osjetljivošću srednje vrijednosti μ_i na promjene u predvidi-telju η_i dobivaju veću težinu. Ovo omogućuje modelu da bolje uvaži pouzdana opažanja i smanji utjecaj manje pouzdanih podataka.

S druge strane, pseudo-odgovor predstavlja niz vrijednosti $\boldsymbol{z} = (z_1, \dots, z_n)^T$ definiranih

$$z_i = \sum_{k=0}^{p} x_{ik} b_k + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right), \quad i = 1, \dots, n. \tag{3.30}$$

Iz gornje formule vidimo da pseudo-odgovor z_i pomaže modelu da se prilagodi na osnovi trenutnih procjena. Prvi sumand u izrazu predstavlja osnovnu procjenu kako se prediktori kombiniraju sa koeficijentima, dok drugi sumand uzima u obzir razliku između stvarne i očekivane vrijednosti, čime se ispravlja predikcija. Ovo omogućava modelu da se bolje uklopi u stvarne podatke, tako da stalno uči i poboljšava svoje procjene.

Sada ćemo ove dvije komponente uklopiti u Fisherovu metodu. Jednadžbu (3.28) možemo zapisati u ekvivalentnom obliku

$$\mathcal{I}^{(m-1)} \boldsymbol{b}^{(m)} = \mathcal{I}^{(m-1)} \boldsymbol{b}^{(m-1)} + \boldsymbol{U}^{(m-1)}.$$

Za $j, k = 0, 1, \dots, p$ elemente lijeve strane jednakosti možemo zapisati kao

$$\begin{aligned} \mathcal{I}_{jk}^{(m-1)} \boldsymbol{b}^{(m)} &\overset{(3.27)}{=} \left[\sum_{i=1}^{n} \frac{x_{ij} x_{ik}}{\mathrm{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right] \boldsymbol{b}^{(m)} \\ &= \left[\sum_{i=1}^{n} x_{ij} \frac{1}{\mathrm{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 x_{ik}\right] \boldsymbol{b}^{(m)} \\ &= \left[\sum_{i=1}^{n} x_{ij} W_{ii}^{(m-1)} x_{ik}\right] \boldsymbol{b}^{(m)} \\ &= (X^T W X)_{jk}^{(m-1)} \boldsymbol{b}^{(m)}, \end{aligned}$$

gdje je $(\partial\mu_i/\partial\eta_i)^2$ evaluirano u $\boldsymbol{b}^{(m-1)}$. Dakle, lijevu stranu jednakosti možemo zapisati kao

$$\mathcal{I}^{(m-1)}\boldsymbol{b}^{(m)} = (X^T W X)^{(m-1)}\boldsymbol{b}^{(m)} \tag{3.31}$$

Slično, za $j = 0, 1, \ldots, p$ elemente desne strane jednakosti možemo zapisati kao

$$\begin{aligned}
(\mathcal{I}^{(m-1)}\boldsymbol{b}^{(m-1)} + \boldsymbol{U}^{(m-1)})_j &= \sum_{k=0}^{p} \mathcal{I}_{jk}^{(m-1)} b_k^{(m-1)} + U_j^{(m-1)} \\
&= \sum_{k=0}^{p}\sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 b_k^{(m-1)} + \sum_{i=1}^{n}\frac{y_i-\mu_i}{\mathrm{Var}(Y_i)}x_{ij}\left(\frac{\partial\mu_i}{\partial\eta_i}\right) \\
&= \sum_{i=1}^{n} x_{ij}\frac{1}{\mathrm{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \sum_{k=0}^{p} x_{ik}b_k^{(m-1)} + \\
&+ \sum_{i=1}^{n} x_{ij}\frac{1}{\mathrm{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 (y_i-\mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right) \\
&= \sum_{i=1}^{n} x_{ij}\frac{1}{\mathrm{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \left[\sum_{k=0}^{p} x_{ik}b_k^{(m-1)} + (y_i-\mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)\right] \\
&= \sum_{i=1}^{n} x_{ij} W_{ii}^{(m-1)} z_i^{(m-1)} \\
&= (X^T W \boldsymbol{z})_j^{(m-1)},
\end{aligned} \tag{3.32}$$

gdje su $(\partial\mu_i/\partial\eta_i)^2$, μ_i i $\partial\eta_i/\partial\mu_i$ evaluirani u $\boldsymbol{b}^{(m-1)}$. Dakle, desnu stranu jednakosti možemo zapisati kao

$$\mathcal{I}^{(m-1)}\boldsymbol{b}^{(m-1)} + \boldsymbol{U}^{(m-1)} = (X^T W \boldsymbol{z})^{(m-1)}. \tag{3.33}$$

Sada jednadžbu (3.28) možemo zapisati u obliku

$$\boldsymbol{b}^{(m)} = [(X^T W X)^{(m-1)}]^{-1}(X^T W \boldsymbol{z})^{(m-1)}. \tag{3.34}$$

U sljedećem ćemo primjeru pokazati kako ove teorijske rezultate primjenjujemo u praksi.

Primjer 3.2.1. *Vlasnik male trgovine u gradu želi istražiti kako prosječna dnevna temperatura i broj kampanja u danu utječe na broj kupaca koji posjećuju njegovu trgovinu. U tablici 3.1 zabilježeni su podaci o broju kupaca koji su posjećivali trgovinu u zadnjih 30 dana, zajedno s prosječnim dnevnim temperaturama i brojem kampanja za te dane.*

n	y_i	x_{i1}	x_{i2}	n	y_i	x_{i1}	x_{i2}
1	7	3.88	4	16	29	8.57	2
2	9	4.54	1	17	23	6.00	2
3	25	8.12	2	18	2	1.07	3
4	7	5.14	2	19	12	6.40	2
5	18	5.26	4	20	10	4.05	1
6	37	8.43	3	21	7	2.86	2
7	21	5.92	4	22	16	4.56	4
8	12	2.47	4	23	7	2.95	2
9	11	3.63	4	24	13	3.54	4
10	6	4.11	3	25	12	3.75	1
11	25	7.45	4	26	4	1.63	3
12	12	5.72	3	27	46	6.68	7
13	21	5.80	4	28	17	5.31	5
14	9	5.22	0	29	7	2.72	5
15	8	3.89	3	30	21	7.51	1

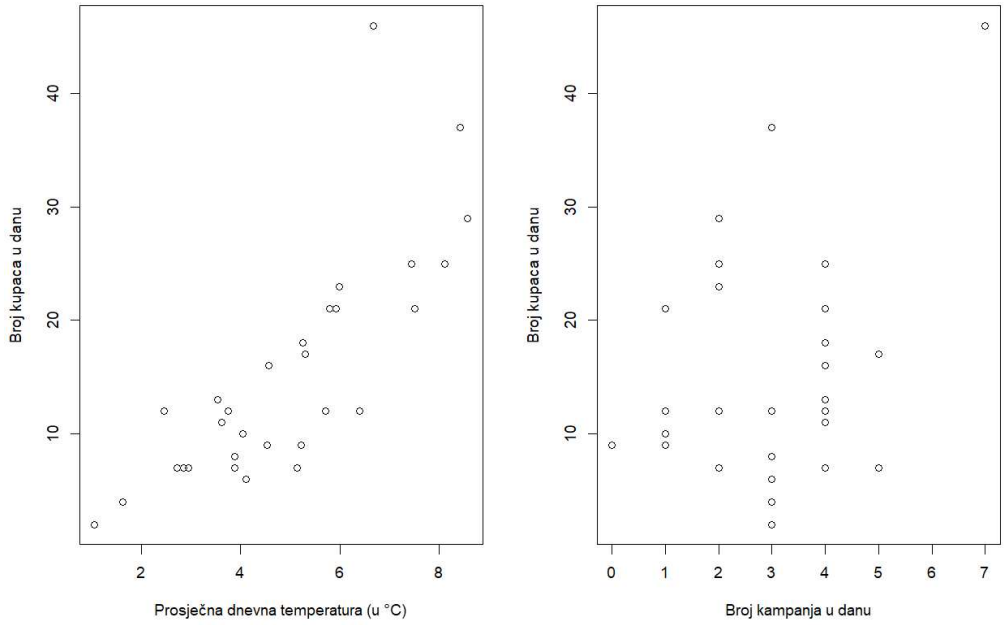
Tablica 3.1: Broj kupaca y_i, prosječne dnevne temperature x_{i1} i broj kampanja u danu x_{i2}

Napomena 3.2.2. *Primjer prikazan u tablici 3.1 je ilustrativne prirode. Slučajne varijable su generirane koristeći programski jezik R, pri čemu je x_{i1} (prosječna dnevna temperatura) generirana pomoću normalne razdiobe s parametrimma $\mu = 5$ i $\sigma^2 = 4$ te x_{i2} (broj promotivnih kampanja) pomoću Poissonove distribucije s parametrom $\lambda = 3$. Nadalje, zavisna je varijabla y_i (broj kupaca) generirana pomoću Poissonove distribucije s parametrima prilagođenim simuliranim podacima, tj. $\lambda_i = e^{0.5+0.3x_{i1}+0.2x_{i2}}$.*

Pretpostavimo da je broj kupaca u danu međusobno nezavisan te neka su $Y = (Y_1, \ldots, Y_n)^T$ slučajne varijable koje predstavljaju broj kupaca, pri čemu je $n = 30$, takve da je $Y_i \sim P(\lambda_i)$, $\lambda_i > 0$, $i = 1, \ldots, n$. Nadalje, pretpostavimo da broj kupaca u danu Y_i ovisi o prosječnoj dnevnoj temperaturi x_{i1} te broju kampanja u danu x_{i2} za $i = 1, \ldots, n$.

Lijevi grafikon na slici 3.5 prikazuje ovisnost broja kupaca koji posjećuju trgovinu (na y-osi) o prosječnoj dnevnoj temperaturi (na x-osi). Iz ovog grafičkog prikaza vidimo da se broj kupaca u danu povećava kako raste temperatura. Međutim, s grafa vidimo da ta veza nije linearna, što implicira da se promjene u broju kupaca ne mogu adekvatno opisati normalnom linearnom regresijom. Desni grafikon prikazuje odnos između broja kampanja u danu (na x-osi) i broja kupaca (na y-osi). Iako se čini da postoji neki utjecaj broja kampanja na broj kupaca (posebno u slučaju manjeg broja kampanja), veza nije tako jasna kao kod temperature. Veći broj kampanja možda može privući više kupaca, ali

podaci sugeriraju da ovaj utjecaj zahtijeva daljnju analizu koju ćemo napraviti u sljedećem poglavlju.



Slika 3.5: Ovisnost broja kupaca u trgovini o prosječnoj dnevnoj temperaturi i broju kampanja u danu

Zbog uočene nelinearnosti primijenit ćemo generalizirane linearne modele, tj. za $i = 1, \ldots, n$ *pretpostavljamo*

$$\begin{cases} Y_i \sim P(\lambda_i), \lambda_i > 0 \\ \mu_i = \lambda_i = \exp[\eta_i] = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta}\right], \end{cases} \tag{3.35}$$

gdje su:

(i) $\mu_i = \mathbb{E}[Y_i] \overset{(1.23)}{=} \lambda_i$,
(ii) $g : \langle 0, +\infty \rangle \to \mathbb{R}, g(\lambda_i) = \log \lambda_i$ *standardna funkcija poveznica iz tablice 2.2,*
(iii) $\boldsymbol{x}_i^T = (1, x_{i1}, x_{i2})$ *vektor prediktora,*
(iv) $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ *vektor nepoznatih parametara.*

Kako su parametri $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ *nepoznati, procjenjujemo ih metodom najveće vjerodostojnosti kako je opisano u prethodnim odjeljcima, tj. tražimo* $\boldsymbol{b} = (b_0, b_1, b_2)^T$ *takve da je*

$$U(b_j) = 0, \quad j = 0, 1, 2.$$

Iz prethodnog odjeljka znamo da su skor funkcije U_j u parametrima β_j za $j = 0, 1, 2$ oblika

$$\begin{aligned} U_j = U(\beta_j) &\overset{(3.26)}{=} \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{Var(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \\ &= \sum_{i=1}^{n} \left[\frac{y_i - \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right]}{\exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right]} x_{ij} \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right] \right] \\ &= \sum_{i=1}^{n} \left(y_i - \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right] \right) x_{ij}, \end{aligned} \tag{3.36}$$

gdje smo u drugoj jednakosti koristili:

(i) $\mu_i = \mathbb{E}[Y_i] \overset{(1.23)}{=} \lambda_i = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right]$,

(ii) $\mathrm{Var}(Y_i) \overset{(1.24)}{=} \lambda_i = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right]$,

(iii) $\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial}{\partial \eta_i} (\exp \eta_i) = \exp \eta_i = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta} \right]$.

Zbog kompleksnosti funkcije U_j rješenje sustava tražimo iterativnom težinskom metodom najmanjih kvadrata gdje je aproksimacija u m-toj iteraciji

$$\boldsymbol{b}^{(m)} \overset{(3.34)}{=} [(X^T W X)^{(m-1)}]^{-1} (X^T W \boldsymbol{z})^{(m-1)}.$$

Dijagonalni elementi težinske matrice $W^{(m-1)} = (W_{ij}^{(m-1)})_{i,j=1,\ldots,n}$ u iteraciji $m-1$ dani su s

$$\begin{aligned} W_{ii}^{(m-1)} &\overset{(3.29)}{=} \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \\ &= \frac{1}{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right]} \left(\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right] \right)^2 \\ &= \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right], \end{aligned}$$

dok je pseudo-odgovor $\boldsymbol{z}^{(m-1)} = (z_1^{(m-1)}, \ldots, z_n^{(m-1)})^T$ u iteraciji $m-1$ dan s

$$\begin{aligned} z_i^{(m-1)} &\overset{(3.30)}{=} \sum_{k=0}^{2} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \\ &= \boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} + \frac{y_i - \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right]}{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right]} \\ &= \frac{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right] (\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + y_i}{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} \right]}. \end{aligned}$$

Nadalje, elementi 3×3 *matrice* $(X^T WX)^{(m-1)}$ *u iteraciji* $m-1$ *za* $j,k = 0,1,2$ *dani su s*

$$\begin{aligned}(X^T WX)^{(m-1)}_{jk} &= \sum_{i=1}^{n} x_{ij} W^{(m-1)}_{ii} x_{ik} \\ &= \sum_{i=1}^{n} x_{ij} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] x_{ik}, \end{aligned} \tag{3.37}$$

pa je

$$\begin{aligned}&(X^T WX)^{(m-1)} = \\ = &\begin{pmatrix} \sum_{i=1}^n \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i1} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i2} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] \\ \sum_{i=1}^n x_{i1} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i1}^2 \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i1}x_{i2} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] \\ \sum_{i=1}^n x_{i2} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i1}x_{i2} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] & \sum_{i=1}^n x_{i2}^2 \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] \end{pmatrix}. \end{aligned} \tag{3.38}$$

Elementi 3*-dimenzionalnog vektora* $(X^T W\boldsymbol{z})^{(m-1)}$ *u iteraciji* $m-1$ *za* $j = 0,1,2$ *su*

$$\begin{aligned}(X^T W\boldsymbol{z})^{(m-1)}_j &= \sum_{i=1}^{n} x_{ij} W^{(m-1)}_{ii} z_i^{(m-1)} \\ &= \sum_{i=1}^{n} x_{ij} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right] \frac{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right](\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + y_i}{\exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right]} \\ &= \sum_{i=1}^{n} x_{ij} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right](\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + x_{ij} y_i. \end{aligned} \tag{3.39}$$

pa je

$$(X^T W\boldsymbol{z})^{(m-1)} = \begin{pmatrix} \sum_{i=1}^n \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right](\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + y_i \\ \sum_{i=1}^n x_{i1} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right](\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + x_{i1} y_i \\ \sum_{i=1}^n x_{i2} \exp\left[\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)}\right](\boldsymbol{x}_i^T \boldsymbol{b}^{(m-1)} - 1) + x_{i2} y_i. \end{pmatrix} \tag{3.40}$$

Algoritam počinjemo s proizvoljno odabranom točkom $\boldsymbol{b}^{(0)} = (0.7, 0.2, 0.1)^T$ *u kojoj je*

$$(X^T WX)^{(0)} = \begin{pmatrix} 235.5016 & 1326.786 & 744.9241 \\ 1326.7859 & 8335.111 & 4148.3251 \\ [744.9241 & 4148.325 & 2941.6834 \end{pmatrix} \quad (X^T W\boldsymbol{z})^{(0)} = \begin{pmatrix} 723.1991 \\ 4361.9652 \\ 2397.3561 \end{pmatrix}$$

pa je

$$\begin{aligned}\boldsymbol{b}^{(1)} &= [(X^T W X)^{(0)}]^{-1}(X^T W \boldsymbol{z})^{(0)} \\ &= \begin{pmatrix} 0.06194 & -0.00689 & -0.00597 \\ -0.00689 & 0.00117 & 0.00010 \\ -0.00597 & 0.00010 & 0.00172 \end{pmatrix} \begin{pmatrix} 723.1991 \\ 4361.9652 \\ 2397.3561 \end{pmatrix} \\ &= \begin{pmatrix} 0.43487 \\ 0.34649 \\ 0.21623 \end{pmatrix} \end{aligned} \tag{3.41}$$

Ovaj postupak nastavljamo sve dok apsolutna razlika dvije uzastopne iteracije ne postigne željenu preciznost $\boldsymbol{\epsilon} = 10^{-6}$*, tj.* $|\boldsymbol{b}^{(m)} - \boldsymbol{b}^{(m-1)}| < \boldsymbol{\epsilon} = 10^{-6}$*. Tablica 3.2 prikazuje aproksimacije procjenitelja maksimalne vjerodostojnosti* $\boldsymbol{b}$ *u svakoj od* 6 *iteracija koje su bile potrebne za postizanje željene preciznosti.*

m**-ta iteracija**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
$b_0^{(m)}$	0.7	0.43487	0.61443	0.71103	0.71649	0.7165	0.71650
$b_1^{(m)}$	0.2	0.34649	0.29486	0.27875	0.27802	0.27802	0.27802
$b_2^{(m)}$	0.1	0.21623	0.17550	0.15978	0.1588	0.1588	0.15880

Tablica 3.2: Aproksimacije procjenitelja maksimalne vjerodostojnosti

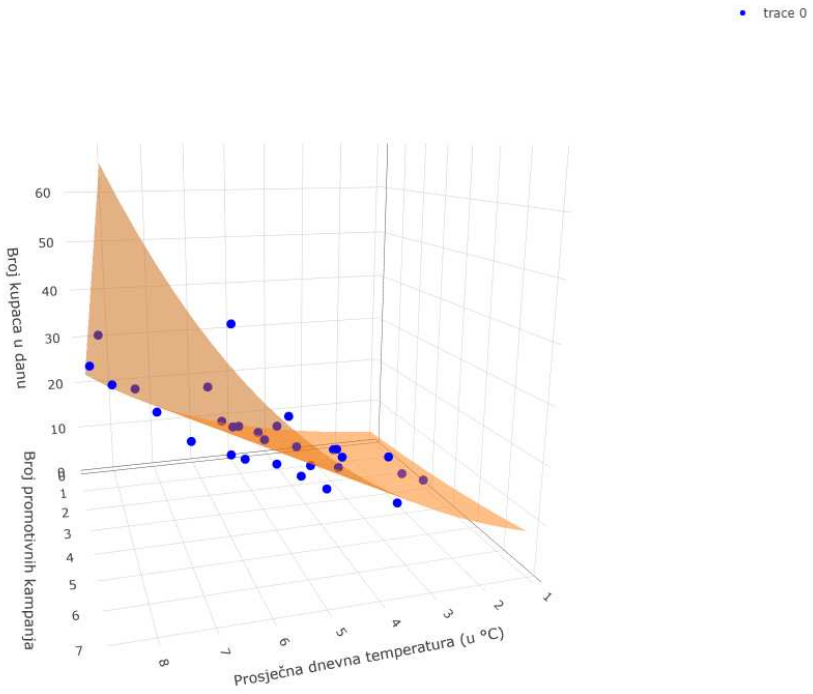
Dakle, procjena parametra $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ *je vektor*

$$\boldsymbol{b} = (b_0, b_1, b_2)^T = (0.71650, 0.27802, 0.15880)^T$$

odakle slijedi

$$\hat{y}_i = \hat{\lambda}_i = \exp\left[0.71650 + 0.27802 x_{i1} + 0.15880 x_{i2}\right]. \tag{3.42}$$

Na slici 3.6 plave točke predstavljaju opažene vrijednosti y_i*, dok je model prikazan crvenom krivuljom. Vidimo da model dobro opisuje opažene podatke jer je većina plavih točaka blizu crvene krivulje, ili na njoj.*



Slika 3.6: Grafički prikaz modela $\hat{\lambda}_i = \exp[b_0 + b_1 x_{i1} + b_2 x_{i2}]$

Iz ove jednadžbe vlasnik trgovine može donositi razne zaključke. Konkretno, ako su procijenjene vrijednosti $\beta_0 \approx 0.71650$ *i* $\beta_1 \approx 0.27802$ *i* $\beta_2 \approx 0.15880$*, možemo reći da osnovni broj kupaca, kada je temperatura 0°C i broj kampanja iznosi 0, iznosi približno* $e^{0.71650} \approx 2$ *kupca. Nadalje, svaki stupanj porasta temperature rezultira povećanjem broja kupaca za oko 32% jer je* $e^{0.27802} \approx 1.32051$*, dok svaka kampanja u danu rezultira povećanjem broja kupaca za oko 17% jer je* $e^{0.15880} \approx 1.172104$*. Također, ako sutra javljaju prosječnu dnevnu temperaturu od* $x_{i1} = 8°C$ *te vlasnik planira imati* $x_{i2} = 1$ *kampanju u danu, tada on može očekivati*

$$\hat{\lambda}_i = \exp[0.71650 + 0.27802 \cdot 8 + 0.15880 \cdot 1] \approx 22$$

kupaca. Ovi rezultati pomažu razumjeti kako promjene temperature i broj kampanja utječu na očekivani broj kupaca u trgovini, omogućujući bolje planiranje resursa i poslovnih aktivnosti.

Poglavlje 4

Statističke inferencije

U prethodnom smo poglavlju na osnovi uzorka $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ procijenili nepoznate parametre $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ našeg modela od interesa. Opaženi nam je uzorak omogućio dobivanje procjena $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ koje su ključne za predviđanje budućih vrijednosti. Sada nas zanima koliko su naše procjene parametara "dobre", odnosno koliko su blizu stvarnim vrijednostima parametara, jer o tome ovisi koliko će "dobro" naš model predviđati buduće vrijednosti. Drugim riječima, želimo procijeniti preciznost i pouzdanost dobivenih procjena kao i ukupnu sposobnost našeg modela da adekvatno opisuje vezu između varijabli.

Stoga se u ovom poglavlju usmjeravamo na statističke inferencije, koje obuhvaćaju *testiranje statističkih hipoteza* i *izračunavanje pouzdanih intervala*, kako bismo donijeli zaključke o pravim vrijednostima parametara $\boldsymbol{\beta}$ na temelju procjena $\boldsymbol{b}$ iz opaženog uzorka $\boldsymbol{y}$.

Za oba su nam tipa statističkih inferencija potrebne uzoračke distribucije. Iz tog ćemo se razloga u prvom dijelu ovog poglavlja fokusirati na derivaciju uzoračkih distribucija potrebnih statistika: skor statistike, procjenitelja maksimalne vjerodostojnosti te devijance. Zatim ćemo u nastavku poglavlja vidjeti kako te distribucije koristimo za računanje intervala pouzdanosti i testiranje statističkih hipoteza.

4.1 Uzoračke distribucije

Za deriviranje potrebnih uzoračkih distribucija definirat ćemo multivarijatnu normalnu razdiobu te χ^2 razdiobu, prateći metodologiju i opisana svojstva u [10].

Definicija 4.1.1. *Neka je* $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ *vektor nezavisnih i jednakodistribuiranih slučajnih varijabli takvih da* $Z_i \sim N(0, 1)$ *za sve* $i = 1, \ldots, n$. *Kažemo da slučajni vektor* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *ima* **multivarijatnu normalnu razdiobu** *s parametrom očekivanja*

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ i kovarijcijskom matricom $\Sigma = (\Sigma_{jk})_{j,k=1,\ldots,n}$, u oznaci $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$, ako se može zapisati u obliku

$$\boldsymbol{Y} = A\boldsymbol{Z} + \boldsymbol{\mu}, \tag{4.1}$$

gdje je $A = [A_{jk}]_{j,k=1,\ldots,n}$ $n \times n$ matrica takva da $\Sigma = AA^T$ (ovo je svojstvo poznato i kao Choleskyjeva dekompozicija).

Korisno svojstvo multivarijatne normalne razdiobe je svojstvo invarijantnosti.

Propozicija 4.1.1. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ n-dimenzionalni slučajni vektor takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$ te neka je $B = (B_{jk})_{j,k=1,\ldots,n}$ konstantna matrica. Tada vrijedi*

$$B\boldsymbol{Y} \sim N(B\boldsymbol{\mu}, B\Sigma B^T) \tag{4.2}$$

Dokaz. Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ n-dimenzionalni vektor takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$. Tada po definiciji multivarijatne normalne razdiobe postoji vektor $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ nezavisnih i jednakodistribuiranih slučajnih varijabli takvih da $Z_i \sim N(0, 1)$ za sve $i = 1, \ldots, n$ te matrica $A = (A_{jk})_{j,k=1,\ldots,n}$ takva da je $\Sigma = AA^T$ te vrijedi

$$\boldsymbol{Y} = A\boldsymbol{Z} + \boldsymbol{\mu}. \tag{4.3}$$

Za $n \times n$ konstantnu matricu $B = (B_{jk})_{j,k=1,\ldots,n}$ imamo

$$B\boldsymbol{Y} = B(A\boldsymbol{Z} + \boldsymbol{\mu}) = BA\boldsymbol{Z} + B\boldsymbol{\mu} \tag{4.4}$$

te iz svojstava kovarijacijskih matrica u (1.15) slijedi

$$\mathrm{Cov}(B\boldsymbol{Y}) = B\mathrm{Cov}(\boldsymbol{Y})B^T = B\Sigma B^T = BAA^T B^T = BA(BA)^T. \tag{4.5}$$

Sada iz definicije multivarijatne normalne razdiobe slijedi $B\boldsymbol{Y} \sim N(B\boldsymbol{\mu}, B\Sigma B^T)$. □

Definicija 4.1.2. *Neka je $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ vektor nezavisnih slučajnih varijabli takvih da za $i = 1, \ldots, n$ vrijedi $Z_i \sim N(\mu_i, 1)$ za $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$. Kažemo da slučajna varijabla*

$$X^2 = \sum_{i=1}^{n} Z_i^2 = \boldsymbol{Z}^T\boldsymbol{Z} \tag{4.6}$$

ima **necentralnu χ^2 distribuciju s n stupnjeva slobode i parametrom necentralnosti λ,** *u oznaci $X^2 \sim \chi^2(n, \lambda)$, gdje je parametar necentralnosti λ dan s*

$$\lambda = \sum_{i=1}^{n} \mu_i^2 = \boldsymbol{\mu}^T\boldsymbol{\mu}. \tag{4.7}$$

Nadalje, ako je $\mu_i = 0$ za sve $i = 1, \ldots, n$, tj. $Z_i \sim N(0, 1)$, onda kažemo da slučajna varijabla X^2 definirana s (4.6) *ima* **centralnu χ^2 distribuciju s n stupnjeva slobode** *(ili, kraće, χ^2* **distribuciju s n stupnjeva slobode***), u oznaci $X^2 \sim \chi^2(n)$.*

U nastavku navodimo nekoliko svojstava χ^2 distribucije koja ćemo koristiti kasnije u radu.

Propozicija 4.1.2. *Neka su $X_1^2 \sim \chi^2(m)$ i $X_2^2 \sim \chi^2(k)$ za $m, k \in \mathbb{N}$ takve da $k > m$ i $X_2^2 = Y^T P_2 Y$, $X_1^2 = Y^T P_1 Y$ za neki standardni normalni slučajni vektor Y i ortogonalne projektore P_1 i P_2 takve da P_1 projicira u potprostor sadržan u potprostoru u koji projicira P_2. Tada vrijedi*

$$X^2 = X_2^2 - X_1^2 \sim \chi^2(k - m) \tag{4.8}$$

i X^2 i X_1^2 su nezavisne slučajne varijable.

Prethodna je propozicija posljedica poznatog Fisher-Cochranovog teorema (vidjeti [1]).

Propozicija 4.1.3. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ n-dimenzionalni slučajni vektor takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$, gdje je Σ pozitivno definitna (regularna) matrica. Tada*

$$\boldsymbol{Y}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y} \sim \chi^2(n, \lambda),$$

gdje je parametar necentralnosti λ dan s

$$\lambda = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}.$$

Dokaz. Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ vektor (ne nužno nezavisnih) slučajnih varijabli takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$. Da bismo pokazali tvrdnju teorema, pretpostavimo da postoji nesingularna $n \times n$ matrica $A = (A_{jk})_{j,k=1,\ldots,n}$ takva da vrijedi $\Sigma = AA^T$. Definiramo

$$\boldsymbol{Z} = A^{-1} \boldsymbol{Y}. \tag{4.9}$$

Tada je

$$\boldsymbol{Y} = A\boldsymbol{Z},$$

odakle slijedi

$$\begin{aligned} \boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} &= \boldsymbol{Z}^T A^T \Sigma^{-1} A \boldsymbol{Z} \\ &= \boldsymbol{Z}^T A^T (AA^T)^{-1} A \boldsymbol{Z} \\ &= \boldsymbol{Z}^T A^T (A^T)^{-1} A^{-1} A \boldsymbol{Z} \\ &= \boldsymbol{Z}^T \boldsymbol{Z}, \end{aligned}$$

Nadalje, iz propozicije 4.1.1 slijedi

$$\begin{aligned} \boldsymbol{Z} &\sim N(A^{-1}\boldsymbol{\mu}, A^{-1}\Sigma(A^{-1})^T) \\ &\sim N(A^{-1}\boldsymbol{\mu}, A^{-1}AA^T(A^T)^{-1}) \\ &\sim N(A^{-1}\boldsymbol{\mu}, I) \end{aligned}$$

gdje je I jedinična $n \times n$ matrica. Dakle, za sve $i = 1, \ldots, n$ su Z_i nezavisne te imamo $Z_i \sim N(\tau_i, 1)$, gdje je $\boldsymbol{\tau} = A^{-1}\boldsymbol{\mu}$. Sada iz definicije necentralne χ^2 distribucije slijedi $\boldsymbol{Y}^T\Sigma^{-1}\boldsymbol{Y} \sim \chi^2(n, \lambda)$, gdje je parametar necentralnosti $\lambda = \boldsymbol{\tau}^T\boldsymbol{\tau} = \boldsymbol{\mu}^T(A^T)^{-1}A^{-1}\boldsymbol{\mu} = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$. □

Propozicija 4.1.4. *Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ n-dimenzionalni slučajni vektor takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$ i Σ je regularna matrica. Tada*

$$(\boldsymbol{Y} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2(n). \tag{4.10}$$

Dokaz. Neka je $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ vektor slučajnih varijabli takav da $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$. Da bismo pokazali tvrdnju teorema, pretpostavimo da postoji nesingularna $n \times n$ matrica $A = (A_{jk})_{j,k=1,\ldots,n}$ takva da vrijedi $\Sigma = AA^T$. Definiramo

$$\boldsymbol{Z} = A^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}). \tag{4.11}$$

Tada je

$$\boldsymbol{Y} = A\boldsymbol{Z} + \boldsymbol{\mu},$$

odakle slijedi

$$\begin{aligned}(\boldsymbol{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) &= \boldsymbol{Z}^T A^T\Sigma^{-1}A\boldsymbol{Z} \\ &= \boldsymbol{Z}^T A^T(AA^T)^{-1}A\boldsymbol{Z} \\ &= \boldsymbol{Z}^T A^T(A^T)^{-1}A^{-1}A\boldsymbol{Z} \\ &= \boldsymbol{Z}^T\boldsymbol{Z},\end{aligned}$$

Nadalje, iz propozicije 4.1.1 slijedi

$$\begin{aligned}\boldsymbol{Z} &\sim N(\boldsymbol{0}, A^{-1}\Sigma(A^{-1})^T) \\ &\sim N(\boldsymbol{0}, A^{-1}AA^T(A^T)^{-1}) \\ &\sim N(\boldsymbol{0}, I)\end{aligned}$$

gdje je I jedinična $n \times n$ matrica. Dakle, za sve $i = 1, \ldots, n$ su Z_i nezavisne te imamo $Z_i \sim N(0, 1)$. Sada iz definicije centralne χ^2 distribucije slijedi $(\boldsymbol{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2(n)$. □

Ako su varijable odgovora normalno distribuirane, uzoračke distribucije koje se koriste za inferenciju često se mogu točno odrediti, dok se za druge distribucije, kao što su one u ekponencijalnim familijama, moramo osloniti na asimptotske rezultate velikih uzoraka koji se temelje na Centralnom graničnom teoremu. U ovakvim slučajevima, asimptotske procjene mogu pružiti pouzdane rezultate za inferenciju čak i kada nije moguće točno odrediti uzoračke distribucije.

Iako strogi razvoj i primjena asimptotske teorije zahtijevaju pažljivo razmatranje uvjeta regularnosti, u ovome radu pretpostavljamo da su uvjeti regularnosti zadovoljeni za eksponencijalne familije distribucija te samim time i generalizirane linearne modele. Ova nam pretpostavka omogućuje da se usredotočimo na primjenu rezultata asimptotske teorije na inferenciju bez dubljeg ulaska u uvjete koji osiguravaju njihovu valjanost. Detalje čitatelj može pronaći na [4].

Prema [1], osnovna ideja uzoračke distribucije statistike od interesa S temelji se na pretpostavci da, ako su ispunjeni odgovarajući uvjeti, tada za dovoljno velike uzorke vrijedi

$$\frac{S - \mathbb{E}[S]}{\sqrt{\mathrm{Var}(S)}} \sim AN(0,1) \Longleftrightarrow \frac{(S - \mathbb{E}[S])^2}{\mathrm{Var}(S)} \sim A\chi^2(1). \tag{4.12}$$

Ako je $\boldsymbol{S} = (S_0, S_1, \ldots, S_p)^T$ $(p+1)$-dimenzionalna statistika s očekivanjem $\mathbb{E}[\boldsymbol{S}]$ i $(p+1)\times(p+1)$ pozitivno definitnom kovarijacijskom matricom Σ takva da $\boldsymbol{S} \sim AN(\mathbb{E}[\boldsymbol{S}], \Sigma)$ te ako su ispunjeni odgovarajuću uvjeti, onda za dovoljno velike uzorke vrijedi

$$(\boldsymbol{S} - \mathbb{E}[\boldsymbol{S}])^T \Sigma^{-1} (\boldsymbol{S} - \mathbb{E}[\boldsymbol{S}]) \sim A\chi^2(p+1). \tag{4.13}$$

Skor statistika

U prethodnom smo poglavlju pokazali da su skor statistike generaliziranih linearnih modela oblika

$$U_j = \sum_{i=1}^{n} \left[\frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right], \quad j = 0, 1, \ldots, p.$$

Uzoračka distribucija skor statistika U_j za $j = 0, 1, \ldots, p$ za dovoljno velike uzorke je

$$\frac{U_j - \mathbb{E}[U_j]}{\sqrt{\mathrm{Var}(U_j)}} = \frac{U_j}{\sqrt{\mathcal{I}_{jj}}} \sim AN(0,1) \Longleftrightarrow \frac{U_j^2}{\mathcal{I}_{jj}} \sim A\chi^2(1) \tag{4.14}$$

jer je $\mathbb{E}[U_j] = 0$ za sve $j = 0, 1, \ldots, p$ prema propoziciji 3.1.1 te

$$\mathrm{Var}(U_j) \overset{(1.11)}{=} \mathrm{Cov}(U_j, U_j) \overset{(3.8)}{=} \mathcal{I}_{jj} \overset{(3.27)}{=} \sum_{i=1}^{n} \frac{x_{ij}^2}{\mathrm{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2.$$

Nadalje, $(p+1)$-dimenzionalna skor statistika $\boldsymbol{U} = (U_0, U_1, \ldots, U_p)^T$ za dovoljno velike uzorke ima multivarijatnu normalnu razdiobu, tj. $\boldsymbol{U} \sim AN(\boldsymbol{0}, \mathcal{I})$ pa je

$$\boldsymbol{U}^T \mathcal{I}^{-1} \boldsymbol{U} \sim A\chi^2(p+1). \tag{4.15}$$

Procjenitelj maksimalne vjerodostojnosti

Taylorov razvoj prvog reda skor funkcije $\boldsymbol{U} = (U_0, U_1, \ldots, U_p)^T$ oko procjene $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ koja je dovoljno blizu pravoj vrijednosti parametra $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ dan je s

$$\boldsymbol{U}(\boldsymbol{\beta}) \approx \boldsymbol{U}(\boldsymbol{b}) + \mathcal{J}(\boldsymbol{b})(\boldsymbol{\beta} - \boldsymbol{b}),$$

gdje je $\mathcal{J}(\boldsymbol{b})$ $(p+1)\times(p+1)$ Jacobijeva matrica čiji su elementi dani (3.13) i evaluirani u $\boldsymbol{b}$. Ako Jacobijevu matricu aproksimiramo njenim očekivanjem, $\mathcal{J}(\boldsymbol{b}) \approx \mathbb{E}[\mathcal{J}(\boldsymbol{b})] \overset{(3.14)}{=} -\mathcal{I}(\boldsymbol{b})$ dobivamo

$$\boldsymbol{U}(\boldsymbol{\beta}) \approx \boldsymbol{U}(\boldsymbol{b}) - \mathcal{I}(\boldsymbol{b})(\boldsymbol{\beta} - \boldsymbol{b}). \tag{4.16}$$

Pokazat ćemo da za dovoljno velike uzorke $\boldsymbol{b} \sim AN(\boldsymbol{\beta}, \mathcal{I}^{-1})$. Naime, kako je $\boldsymbol{b}$ procjenitelj maksimalne vjerodostojnosti parametra $\boldsymbol{\beta}$, onda vrijedi da je $\boldsymbol{U}(\boldsymbol{b}) = 0$ pa je iz (4.16)

$$\boldsymbol{b} - \boldsymbol{\beta} \approx (\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{U}(\boldsymbol{\beta}). \tag{4.17}$$

Sada iz svojstava matematičkog očekivanja primijenjenog na konstante $\boldsymbol{\beta}$ i $\mathcal{I}(\boldsymbol{b}))^{-1}$ te činjenice da je $\mathbb{E}[\boldsymbol{U}(\boldsymbol{\beta})] = \boldsymbol{0}$ slijedi

$$\begin{aligned}\mathbb{E}[\boldsymbol{b} - \boldsymbol{\beta}] = \mathbb{E}[(\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{U}(\boldsymbol{\beta})] \iff \mathbb{E}[\boldsymbol{b}] - \boldsymbol{\beta} &= (\mathcal{I}(\boldsymbol{b}))^{-1}\mathbb{E}[\boldsymbol{U}(\boldsymbol{\beta})] \\ &= (\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{0} \\ &= \boldsymbol{0}\end{aligned}$$

pa je $\mathbb{E}[\boldsymbol{b}] = \boldsymbol{\beta}$, što nam govori da je procjenitelj $\boldsymbol{b}$ nepristran procjenitelj parametra $\boldsymbol{\beta}$, tj. procjenitelj u prosjeku daje točnu vrijednost pravog parametra. Nadalje, iz definicije kovarijacijske matrice slijedi

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{b}) &\overset{(1.13)}{=} \mathbb{E}\left[(\boldsymbol{b} - \boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})^T\right] \\ &\overset{(4.17)}{=} \mathbb{E}\left[[(\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{U}(\boldsymbol{\beta})][(\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{U}(\boldsymbol{\beta})]^T\right] \\ &= \mathbb{E}\left[(\mathcal{I}(\boldsymbol{b}))^{-1}\boldsymbol{U}(\boldsymbol{\beta})[\boldsymbol{U}(\boldsymbol{\beta})]^T(\mathcal{I}(\boldsymbol{b}))^{-1}\right] \\ &= (\mathcal{I}(\boldsymbol{b}))^{-1}\mathbb{E}\left[\boldsymbol{U}(\boldsymbol{\beta})[\boldsymbol{U}(\boldsymbol{\beta})]^T\right](\mathcal{I}(\boldsymbol{b}))^{-1} \\ &= (\mathcal{I}(\boldsymbol{b}))^{-1}\mathrm{Cov}(\boldsymbol{U}(\boldsymbol{\beta}))(\mathcal{I}(\boldsymbol{b}))^{-1} \\ &= (\mathcal{I}(\boldsymbol{b}))^{-1}\mathcal{I}(\boldsymbol{\beta})(\mathcal{I}(\boldsymbol{b}))^{-1} \\ &\approx (\mathcal{I}(\boldsymbol{b}))^{-1}\mathcal{I}(\boldsymbol{b})(\mathcal{I}(\boldsymbol{b}))^{-1} \\ &= (\mathcal{I}(\boldsymbol{b}))^{-1}.\end{aligned}$$

Dakle, za dovoljno velike uzorke je $\boldsymbol{b} \sim AN(\boldsymbol{\beta}, (\mathcal{I}(\boldsymbol{b}))^{-1})$ pa je

$$(\boldsymbol{b} - \boldsymbol{\beta})^T\mathcal{I}(\boldsymbol{b})(\boldsymbol{b} - \boldsymbol{\beta}) \sim A\chi^2(p+1). \tag{4.18}$$

Devijanca

U kontekstu testiranja statističkih hipoteza, jedan od načina za procjenu prikladnosti modela od interesa je uspoređivanje s općenitijim, tzv. *saturiranim modelom*. Saturirani model je model koji ima maksimalni broj parametara koje možemo procijeniti za dani skup podataka. Drugim riječima, ovaj model može savršeno pristajati podacima, jer ima dovoljno slobode da opiše svaki pojedinačni podatak točno.

Ako imamo n opažanja y_i gdje $i = 1, \ldots, n$ takvih da sva opažanja imaju različite vrijednosti linearnog predviditelja $\boldsymbol{x}_i^T\boldsymbol{\beta}$, saturirani će model imati n parametara. Ovu vrstu modela nazivamo i **maksimalnim** ili **punim modelom**.

Međutim, ako neka opažanja dijele iste vrijednosti linearnog predviditelja (odnosno, ako dvije ili više opaženih vrijednosti ishoda y_i odgovaraju istoj kombinaciji varijabli $\boldsymbol{x}_i^T\boldsymbol{\beta}$), tada nije moguće svakom opažanju pridružiti zaseban parametar. To se događa jer se ista kombinacija nezavisnih varijabli pojavljuje više puta, što znači da nemamo dovoljno informacija za procjenu različitih parametara za svako pojedinačno opažanje. Zbog toga je maksimalni broj parametara koje možemo procijeniti u saturiranom modelu manji ili jednak n i ovisi o broju različitih kombinacija nezavisnih varijabli (tj. različitih linearnih predviditelja).

Neka $m + 1 \leq n$ predstavlja najveći broj parametara koji se mogu procijeniti u saturiranom modelu s vektorom parametara $\boldsymbol{\beta}_{\max}$ te neka je $\boldsymbol{b}_{\max}$ procjenitelj maksimalne vjerodostojnosti koji maksimizira funkciju vjerodostojnosti. S drugu stranu, pretpostavimo da naš model od interesa ima $p + 1 < m + 1$ parametara $\boldsymbol{\beta}$ te neka je $\boldsymbol{b}$ odgovarajući procjenitelj maksimalne vjerodostojnosti.

Budući da saturirani model uzima u obzir sve moguće parametre, njegova funkcija vjerodostojnosti $L(\boldsymbol{b}_{\max}; \boldsymbol{y})$ uvijek će biti veća ili jednaka od $L(\boldsymbol{b}; \boldsymbol{y})$, kao i svih drugih mogućih kombinacija parametara u manje složenim modelima.

Usporedbom maksimalnih vrijednosti funkcija vjerodostojnosti za oba modela, možemo izračunati omjer vjerodostojnosti

$$\lambda = \frac{L(\boldsymbol{b}_{\max}; \boldsymbol{y})}{L(\boldsymbol{b}; \boldsymbol{y})}.$$

U praksi, za procjenu prikladnosti modela koristi se dvostruki logaritam omjera funkcija vjerodostojnosti, poznat kao **devijanca**. Definira se kao

$$D = 2 \log \lambda = 2[l(\boldsymbol{b}_{\max}; \boldsymbol{y}) - l(\boldsymbol{b}; \boldsymbol{y})]. \tag{4.19}$$

Visoke vrijednosti devijance ukazuju na loše prilagođavanje modela od interesa podacima u odnosu na saturirani model, što sugerira da model od interesa možda nije adekvatan za opisivanje skupa podataka. Sljedeći je korak određivanje uzoračke distribucije devijance.

Taylorov razvoj drugog reda funkcije log-vjerodostojnosti oko procjene $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ koja je dovoljno blizu pravoj vrijednosti parametra $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ dan je s

$$\begin{aligned} l(\boldsymbol{\beta}) &\approx l(\boldsymbol{b}) + (\boldsymbol{\beta} - \boldsymbol{b})^T \boldsymbol{U}(\boldsymbol{b}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b})^T \mathcal{J}(\boldsymbol{b})(\boldsymbol{\beta} - \boldsymbol{b}) \\ &\overset{(3.14)}{\approx} l(\boldsymbol{b}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b})^T \mathcal{I}(\boldsymbol{b})(\boldsymbol{\beta} - \boldsymbol{b}) \end{aligned}$$

gdje smo u drugoj jednakosti koristili činjenicu da je $\boldsymbol{b}$ procjenitelj maksimalne vjerodostojnosti parametra $\boldsymbol{\beta}$ pa je $\boldsymbol{U}(\boldsymbol{b}) = 0$ te smo aproksimirali Jacobijevu matricu očekivanjem, tj. $\mathcal{J}(\boldsymbol{b}) \approx \mathbb{E}[\mathcal{J}(\boldsymbol{b})] = -\mathcal{I}(\boldsymbol{b})$. Dakle, imamo

$$l(\boldsymbol{b}) - l(\boldsymbol{\beta}) \approx \frac{1}{2}(\boldsymbol{b} - \boldsymbol{\beta})^T \mathcal{I}(\boldsymbol{b})(\boldsymbol{b} - \boldsymbol{\beta}),$$

odakle za dovoljno velike uzorke slijedi

$$2[l(\boldsymbol{b}; \boldsymbol{y}) - l(\boldsymbol{\beta}; \boldsymbol{y})] \approx (\boldsymbol{b} - \boldsymbol{\beta})^T \mathcal{I}(\boldsymbol{b})(\boldsymbol{b} - \boldsymbol{\beta}) \sim A\chi^2(p + 1). \tag{4.20}$$

Sada raspisivanjem (4.19) dobivamo

$$\begin{aligned} D &= 2[l(\boldsymbol{b}_{\mathbf{max}}; \boldsymbol{y}) - l(\boldsymbol{b}; \boldsymbol{y})] \\ &= 2[l(\boldsymbol{b}_{\mathbf{max}}; \boldsymbol{y}) - l(\boldsymbol{\beta}_{\mathbf{max}}; \boldsymbol{y})] - 2[l(\boldsymbol{b}; \boldsymbol{y}) - l(\boldsymbol{\beta}; \boldsymbol{y})] + 2[l(\boldsymbol{\beta}_{\mathbf{max}}; \boldsymbol{y}) - l(\boldsymbol{\beta}; \boldsymbol{y})], \end{aligned}$$

odakle iz (4.20) slijedi

(i) $2[l(\boldsymbol{b}_{\mathbf{max}}; \boldsymbol{y}) - l(\boldsymbol{\beta}_{\mathbf{max}}; \boldsymbol{y})] \sim A\chi^2(m + 1)$
(ii) $2[l(\boldsymbol{b}; \boldsymbol{y}) - l(\boldsymbol{\beta}; \boldsymbol{y})] \sim A\chi^2(p + 1)$
(iii) $\nu = 2[l(\boldsymbol{\beta}_{\mathbf{max}}; \boldsymbol{y}) - l(\boldsymbol{\beta}; \boldsymbol{y})]$ je pozitivna konstanta koja je blizu 0 ako model od interesa pruža jednako dobru prilagodbu podacima kao saturirani model.

Prema propoziciji 4.1.2 slijedi

$$D \sim \chi^2(m - p, \nu). \tag{4.21}$$

Primjer 4.1.1. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *vektor nezavisnih slučajnih varijabli takvih da* $Y_i \sim P(\lambda_i), \lambda_i > 0$ *za* $i = 1, \ldots, n$. *Funkcija log-vjerodostojnosti parametra* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ *s obzirom na opaženi uzorak* $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ *dana je s*

$$l(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \log y_i!$$

Za saturirani su model svi λ_i različiti pa je procjenitelj maksimalne vjerodostojnosti parametra $\boldsymbol{\beta}_{max} = (\lambda_1, \ldots, \lambda_n)^T$ vektor $\boldsymbol{b}_{max} = (y_1, \ldots, y_n)^T$, odakle slijedi

$$l(\boldsymbol{b}_{max}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log y_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \log y_i!$$

Pretpostavimo sada da naš model od interesa ima $p < n$ parametara te neka je $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ procjenitelj maksimalne vjerodostojnosti parametra $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. Nadalje, neka je $\hat{y}_i = \hat{\lambda}_i = g^{-1}(\boldsymbol{x}_i^T \boldsymbol{b})$, gdje su $x_1, \ldots, x_p$ prediktori te g odgovarajuća funkcija poveznica. Tada je

$$l(\boldsymbol{b}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \hat{y}_i - \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} \log y_i!$$

Sada je devijanca dana s

$$\begin{aligned} D &\overset{(4.19)}{=} 2[l(\boldsymbol{b}_{max}; \boldsymbol{y}) - l(\boldsymbol{b}; \boldsymbol{y})] \\ &= 2\left[\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - \sum_{i=1}^{n} (y_i - \hat{y}_i)\right]. \end{aligned} \tag{4.22}$$

Primjer 4.1.2. *Vratimo se na primjer 3.2.1 gdje smo modelirati broj kupaca u danu pomoću prosječne dnevne temperature i broja kampanji u danu pomoću Poissonove regresije.*

Funkcija log-vjerodostojnosti parametra $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ s obzirom na opaženi uzorak $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ dana je s

$$l(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \log y_i!$$

Za saturirani su model svi λ_i različiti pa je procjenitelj maksimalne vjerodostojnosti parametra $\boldsymbol{\beta}_{max} = (\lambda_1, \ldots, \lambda_n)^T$ vektor $\boldsymbol{b}_{max} = (y_1, \ldots, y_n)^T$, odakle slijedi

$$l(\boldsymbol{b}_{max}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log y_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \log y_i!$$

Za model s nepoznatim parametrom $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ procjenitelj maksimalne vjerodostojnosti je $\boldsymbol{b} = (b_0, b_1, b_2)^T = (0.71650, 0.27802, 0.15880)^T$, odakle iz

$$\hat{y}_i = \hat{\lambda}_i = \exp[0.71650 + 0.27802 x_{i1} + 0.15880 x_{i2}]$$

slijedi

$$l(\boldsymbol{b}; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \hat{y}_i - \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} \log y_i!$$

Za model s nepoznatim parametrom $\beta = \lambda$ procjenitelj maksimalne vjerodostojnosti je $b = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{\boldsymbol{y}}$, odakle slijedi

$$l(b; \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \bar{\boldsymbol{y}} - \sum_{i=1}^{n} \bar{\boldsymbol{y}} - \sum_{i=1}^{n} \log y_i!$$

Sada možemo izračunati

$$\begin{aligned} D_0 &= 2[l(\boldsymbol{b}_{max}; \boldsymbol{y}) - l(b; \boldsymbol{y})] \\ &= 2\left[\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\bar{\boldsymbol{y}}}\right) - \sum_{i=1}^{n}(y_i - \bar{\boldsymbol{y}})\right] \approx 174.631 \\ D_{res} &= 2[l(\boldsymbol{b}_{max}; \boldsymbol{y}) - l(\boldsymbol{b}; \boldsymbol{y})] \\ &= 2\left[\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - \sum_{i=1}^{n}(y_i - \hat{y}_i)\right] \approx 26.413 \end{aligned}$$

Vrijednost D_0 u literaturi ćemo često naći pod nazivom **nul devijanca***, koja mjeri odstupanje između modela koji predviđa samo prosječnu vrijednost zavisne varijable i stvarnih opažanja, odnosno devijancu modela koji koristi samo presjek bez prediktora.U našem slučaju visoka vrijednost devijance D_0 pokazuje veliko odstupanje prosjeka od opaženih vrijednosti, što nam sugerira da broj kupaca u trgovini vjerojatno ovisi o nekim nezavisnim varijablama koje bi trebalo uključiti u model.*

S drugu stranu, vrijednost D_{res} nazivamo **rezidualna devijanca** *i ona mjeri odstupanje između modela od interesa i stvarnih opažanja. Vidimo da je vrijednost rezidualne devijance znatno manja od nul devijance što ukazuje da model s prediktorima objašnjava puno više varijacije u podacima nego model bez prediktora.*

4.2 Testiranje statističkih hipoteza

Testiranje statističkih hipoteza ključno je za analizu veza između varijabli, testiranje specifičnih pretpostavki te pružanje informacija o preciznosti naših procjena. Osnovni koraci u formiranju statističkog testa uključuju *definiranje nulte* i *alternativne hipoteze*, *odabir testne statistike*, *određivanje kritičnog područja* te *izračun p-vrijednosti*. U nastavku slijedi kratak opis svakog od ovih koraka prema [8].

1. Definicija nulte i alternativne hipoteze. Nulta hipoteza H_0 je pretpostavka koju želimo testirati. Ona obično predstavlja stanje ”nema efekta” ili ”nema razlike”. S drugu stranu, alternativna hipoteza H_1 je hipoteza koju prihvaćamo ako odbacimo nultu hipotezu. Ona predstavlja stanje ”ima efekta” ili ”ima razlike”.

2. Odabir testne statistike. Na temelju nulte i alternativne hipoteze, odabiremo testnu statistiku T za koju pretpostavljamo da poznajemo distribuciju f u uvjetima nulte hipoteze, tj. $T \overset{H_0}{\sim} f$. Statistika T predstavlja mjeru koja se koristi za procjenu hipoteza.

3. Određivanje kritičnog područja. Za zadanu razinu značajnosti α (najčešće 0.01, 0.05 ili 0.1), određujemo kritično područje C. Ovo područje predstavlja skup vrijednosti testne statistike T za koje bismo odbacili nultu hipotezu. Kritično se područje definira kao

$$\mathbb{P}(T \in C|H_0) = \alpha.$$

To znači da, ako je H_0 istinita, vjerojatnost da ćemo dobiti vrijednosti testne statistike T unutar kritičnog područja C ili da ćemo odbaciti nultu hipotezu, je jednaka α.

4. Izračun p-vrijednosti. Nakon što odredimo realizaciju t testne statistike T, izračunavamo p-vrijednost koja predstavlja vjerojatnost dobivanja rezultata koji su ekstremniji od onog što smo dobili, pod pretpostavkom da je nulta hipoteza istinita, tj.

$$p\text{-vrijednost} = \mathbb{P}(t \text{ na rubu kritičnog područja } |H_0).$$

Na kraju, odluku o odbacivanju ili ne odbacivanju nulte hipoteze H_0 na zadanoj razini značajnosti α možemo donijeti na osnovi

(i) kritičnog područja C:

$t \in C \implies$ odbacujemo nultu hipotezu H_0,
$t \notin C \implies$ ne odbacujemo nultu hipotezu H_0,

(ii) p-vrijednosti:

p-vrijednost $\leq \alpha \implies$ odbacujemo H_0 na razini značajnosti α,
p-vrijednost $> \alpha \implies$ ne odbacujemo H_0 na razini značajnosti α.

Testiranje parametara

Nulta hipoteza H_0 često postavlja okvire za testiranje značajnosti pojedinih parametara u modelu, obično navodeći da parametar nije značajan (tj. da je koeficijent jednak nuli). Nasuprot njoj, alternativna hipoteza H_1 sugerira da je parametar značajan te da se koeficijent razlikuje od nule. Ova analiza omogućuje donošenje informiranih odluka o korištenju određenih varijabli u modelu.

Općenito, za $j = 0, 1, \ldots, p$ testiramo hipoteze

$$H_0 : \beta_j = a_j$$
$$H_1 : \beta_j \neq a_j,$$

gdje je $a_j \in \mathbb{R}$ neka konstanta. Za testiranje ovih hipoteza uzimamo jednu od testnih statistika procjenitelja maksimalne vjerodostojnosti, tj. za dovoljno velike uzorke je

$$Z_j = \frac{b_j - a_j}{\sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})}} \overset{H_0}{\sim} AN(0,1) \Longleftrightarrow Z_j^2 = \frac{(b_j - a_j)^2}{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})} \overset{H_0}{\sim} A\chi^2(1)$$

U slučaju kada je hipoteza H_0 istinita, očekujemo da će testna statistika Z_j biti mala, jer će procijenjena vrijednost parametra b_j biti blizu stvarne vrijednosti $\beta_j = a_j$. Ako je Z_j velika, to sugerira da je H_0 vjerojatno pogrešna, te ćemo u tom slučaju odbaciti H_0 na razini značajnosti α. Ovisno o vrsti statistike koju koristimo, možemo primijeniti test na temelju (aproksimativne) normalne distribucije $N(0,1)$ ili (aproksimativne) $\chi^2(1)$ distribucije. Konačno, na temelju dobivene realizacije testne statistike, možemo izračunati odgovarajuću p-vrijednost i donijeti zaključak o pravoj vrijednosti parametra β_j.

Primjer 4.2.1. *U primjeru 3.2.1 smo iterativnom težinskom metodom najmanjih kvadrata smo došli do procjene parametara* $\boldsymbol{\beta} = (\beta_0, \beta_1, b_2)^T$ *koji iznose*

$$\boldsymbol{b} = (b_0, b_1, b_2)^T = (0.71650, 0.27802, 0.15880)^T.$$

Kao što smo već tada sugerirali, iz desnog grafikona na slici 3.5 nije jasno utječe li broj kampanja na broj kupaca u danu. Sada to želimo testirati. Na razini značajnosti od $\alpha = 0.05$ *želimo testirati hipotezu o značajnosti parametra* β_2*, tj. testiramo hipoteze*

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0.$$

Koristit ćemo testnu statistiku

$$Z = \frac{b_2}{\sqrt{\mathcal{I}^{-1}_{22}(\boldsymbol{b})}} \overset{H_0}{\sim} AN(0,1),$$

gdje je

$$\mathcal{I}(\boldsymbol{b}) \overset{(3.37)}{=} \begin{pmatrix} 454.000 & 2678.146 & 1497.000 \\ 2678.146 & 17372.661 & 8719.750 \\ 1497.000 & 8719.750 & 6151.332 \end{pmatrix} \tag{4.23}$$

pa je

$$(\mathcal{I}(\boldsymbol{b}))^{-1} = \begin{pmatrix} 0.035730 & -0.00396 & -0.00308 \\ -0.00396 & 0.00064 & 0.00006 \\ -0.00308 & 0.00006 & 0.00083 \end{pmatrix}. \tag{4.24}$$

Realizacija testne statistike iznosi

$$z = \frac{0.15880}{\sqrt{0.00083}} \approx 5.518.$$

Za zadanu razinu značajnosti $\alpha = 0.05$*, kritično područje je oblika*

$$C = \langle -\infty, -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}, +\infty\rangle \approx \langle -\infty, -1.96] \cup [1.96, +\infty\rangle.$$

Kako realizacija testne statistike Z upada u kritično područje C zaključujemo da odbacujemo nultu hipotezu $H_0 : \beta_2 = 0$ *u korist alternativne* $H_1 : \beta_2 \neq 0$ *na razini značajnosti od* $\alpha = 0.05$.
Nadalje, vrijedi

$$p\text{-vrijednost} = 2[1 - \mathbb{P}(Z \leq |z|)] = 2[1 - \Phi(|z|)] = 2[1 - \Phi(5.518)] \approx 3.43 \cdot 10^{-8}.$$

Kako je p-vrijednost $\approx 3.43 \cdot 10^{-8} \leq 0.05$ *ponovno donosimo odluku o odbacivanju nulte hipoteze na razini značajnosti od* $\alpha = 0.05$*. Dakle, ovime smo potvrdili da broj kampanja u danu statistički značajno utječe na dnevni broj kupaca u trgovini. Rezultati testiranja značajnosti svih parametara prikazani su u tablici 4.1.*

Prediktor	b_j	$[\mathcal{I}(\boldsymbol{b})]^{-1}$	z	p-vrijednost
1	0.71650	0.18902	3.791	$1.5 \cdot 10^{-4}$
x_1	0.27802	0.02529	10.995	$< 2 \cdot 10^{-16}$
x_2	0.15880	0.02878	5.518	$3.43 \cdot 10^{-8}$

Tablica 4.1: Rezultati testiranja značajnosti parametara modela

Na osnovi prikazanih rezultata vidimo da su svi parametri modela statistički značajni.

Testiranje modela

Osim što možemo testirati svaki pojedinačni parametar u modelu, postoji mogućnost da istovremeno testiramo sve parametre zajedno.
Testiramo hipoteze

$$H_0 : \boldsymbol{\beta} = \boldsymbol{a}$$
$$H_1 : \boldsymbol{\beta} \neq \boldsymbol{a},$$

gdje je $\boldsymbol{a} = (a_0, a_1, \ldots, a_p)^T \in \mathbb{R}^{p+1}$ konstantan vektor.

Definiramo testnu statistiku X^2 s

$$X^2 = (\boldsymbol{b} - \boldsymbol{a})^T \mathcal{I}(\boldsymbol{b})(\boldsymbol{b} - \boldsymbol{a}) \overset{H_0}{\sim} A\chi^2(p+1).$$

Na temelju testne statistike odredimo kritično područje i p-vrijednost kako bismo donijeli odluku o hipotezi od interesa.

Primjer 4.2.2. *Za naš model iz primjera 3.2.1 na razini značajnosti od* $\alpha = 0.05$ *testiramo hipoteze*

$$H_0 : \boldsymbol{\beta} = \mathbf{0}$$
$$H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Testna statistiku X^2 *dana je s*

$$X^2 = \boldsymbol{b}^T \mathcal{I}(\boldsymbol{b})\boldsymbol{b} \overset{H_0}{\sim} A\chi^2(3).$$

Realizacija testne statistike iznosi

$$x^2 = \begin{pmatrix} 0.71650 & 0.27802 & 0.15880 \end{pmatrix} \begin{pmatrix} 0.035730 & -0.00396 & -0.00308 \\ -0.00396 & 0.00064 & 0.00006 \\ -0.00308 & 0.00006 & 0.00083 \end{pmatrix} \begin{pmatrix} 0.71650 \\ 0.27802 \\ 0.15880 \end{pmatrix}$$
$$\approx 3908.529.$$

Već na temelju visoke vrijednosti X^2 *statistike možemo naslutiti da ćemo odbaciti nultu hipotezu u korist alternativne, ali ćemo to potvrditi određivanjem kritičnog područja ili p-vrijednosti.*
Kritično područje dano je s

$$C = [\chi^2_{1-\alpha}(3), +\infty\rangle \approx [7.81, +\infty\rangle.$$

Sada vidimo da $x^2 = 3908.529 \in C$ *pa odbacujemo nultu hipotezu u korist alternativne na razini značajnosti* $\alpha = 0.05$.
Također, p-vrijednost

$$p\text{-}vrijednost = 1 - \mathbb{P}(X^2 \leq x^2) = 1 - \mathbb{P}(X^2 \leq 3908.529) \approx 0,$$

odakle ponovno zaključujemo da odbacujemo nultu hipotezu u korist alternativne. Dakle, naš je model statistički značajan.

Usporedba modela

Osim testiranja značajnosti pojedinih parametara, hipoteze se mogu koristiti i za usporedbu složenijih i jednostavnijih modela. U tom kontekstu, složeniji model M_1 može uključivati dodatne varijable ili interakcije koje ne postoje u jednostavnijem modelu M_0. Ključno je da jednostavniji model bude specijalni slučaj složenijeg modela, što znači da oba modela imaju istu vjerojatnosnu razdiobu i istu funkciju poveznicu, ali je linearni predviditelj jednostavnijeg modela poseban slučaj linearnog predviditelja složenijeg modela. Tada nulta hipoteza može postaviti pretpostavku da složeniji model ne donosi značajno poboljšanje u odnosu na jednostavniji model, dok bi alternativna hipoteza sugerirala da složeniji model značajno poboljšava izvedbu modela.

Općenito, testiramo hipotezu

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = (\beta_0, \beta_1, \ldots, \beta_q)^T$$

koja odgovara jednostavnijem modelu M_0 nasuprot alternativne

$$H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = (\beta_0, \beta_1, \ldots, \beta_p)^T,$$

koja odgovara složenijem modelu M_1, za $q < p < m$, pri čemu je $m + 1$ maksimalan broj parametara koje koji se mogu procijeniti u saturiranom modelu.

Definiramo testnu statistiku

$$\begin{aligned} \Delta D = D_0 - D_1 &\overset{(4.19)}{=} 2[l(\boldsymbol{b}_{\max}; \boldsymbol{y}) - l(\boldsymbol{b}_0; \boldsymbol{y})] - 2[l(\boldsymbol{b}_{\max}; \boldsymbol{y}) - l(\boldsymbol{b}_1; \boldsymbol{y})] \\ &= 2[l(\boldsymbol{b}_1; \boldsymbol{y}) - l(\boldsymbol{b}_0; \boldsymbol{y})] \end{aligned}$$

Ako oba modela dobro opisuju podatke, onda iz (4.21) za dovoljno velike uzorke slijedi $D_0 \sim A\chi^2(m - q)$ te $D_1 \sim A\chi^2(m - p)$ pa iz propozicije 4.1.2 slijedi

$$\Delta D = 2[l(\boldsymbol{b}_1; \boldsymbol{y}) - l(\boldsymbol{b}_0; \boldsymbol{y})] \sim A\chi^2(p - q).$$

Primjer 4.2.3. *U primjeru 3.2.1 na razini značajnosti od $\alpha = 0.05$ želimo testirati hipotezu*

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = (\beta_0, \beta_1)^T \tag{4.25}$$

koja odgovara modelu $\lambda_i = \exp[\beta_0 + \beta_1 x_{i1}]$ nasuprot alternativne

$$H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = (\beta_0, \beta_1, \beta_2)^T \tag{4.26}$$

koja odgovara modelu $\lambda_i = \exp[\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}]$, tj. testiramo je li dodavanje varijable o dnevnom broju kampanji statistički značajno za poboljšanje modela.

Definiramo testnu statistiku

$$\Delta D = D_0 - D_1 = 2[l(\boldsymbol{b}_1; \boldsymbol{y}) - l(\boldsymbol{b}_0; \boldsymbol{y})] \sim A\chi^2(1).$$

Za model s jednim prediktorom x_{i1} i nepoznatim parametrom $\boldsymbol{\beta}_0 = (\beta_0, \beta_1)^T$ može se pokazati da je

$$\begin{aligned} \boldsymbol{b}_0 = (1.28105, 0.26568)^T \Longrightarrow [\hat{y}_i]_0 &= [\hat{\lambda}_i]_0 \\ &= \exp[1.28105 + 0.26568 x_{i1}]. \end{aligned}$$

Za model s dva prediktora x_{i1}, x_{i2} i nepoznatim parametrom $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \beta_2)^T$ pokazali smo

$$\begin{aligned} \boldsymbol{b}_1 = (0.71650, 0.27802, 0.15880)^T \Longrightarrow [\hat{y}_i]_1 &= [\hat{\lambda}_i]_1 \\ &= \exp[0.71650 + 0.27802 x_{i1} + 0.15880 x_{i2}]. \end{aligned}$$

Sada, slično kao u primjeru 4.1.1 dobivamo

$$\begin{aligned} \Delta D &= 2[l(\boldsymbol{b}_1; \boldsymbol{y}) - l(\boldsymbol{b}_0; \boldsymbol{y})] \\ &= 2\left[\sum_{i=1}^{n} y_i \log\left(\frac{[\hat{y}_i]_1}{[\hat{y}_i]_0}\right) - \sum_{i=1}^{n}([\hat{y}_i]_1 - [\hat{y}_i]_0)\right] \sim A\chi^2(1). \end{aligned}$$

Realizacija testne statistike iznosi

$$\Delta d \approx 28.909.$$

Kritično područje dano je s

$$C = [\chi^2_{1-\alpha}(1), +\infty\rangle \approx [3.84, +\infty\rangle.$$

Sada vidimo da $\Delta d = 28.909 \in C$ pa odbacujemo nultu hipotezu u korist alternativne na razini značajnosti $\alpha = 0.05$.
Također, p-vrijednost

$$\textit{p-vrijednost} = 1 - \mathbb{P}(\Delta D \leq \Delta d) = 1 - \mathbb{P}(\Delta D \leq 3.841) \approx 7.59 \cdot 10^{-8},$$

odakle ponovno zaključujemo da odbacujemo nultu hipotezu u korist alternativne. Dakle, model s prosječnom dnevnom temperaturom i brojem kampanja u danu bolje opisuje dnevni broj kupaca u odnosu na model s prosječnom dnevnom temperaturom.

4.3 Izračun pouzdanih intervala

Izračun pouzdanih intervala pruža pregled raspona vrijednosti unutar kojih se procijenjeni parametar vjerojatno nalazi. Pouzdani se intervali često smatraju korisnijima od testova hipoteza jer širina pouzdanih intervala pruža mjeru preciznosti procjena, što olakšava razumijevanje mogućih varijacija u populaciji na način koji je jednostavniji od samog statističkog testa. U nastavku dajemo općenitu definiciju pouzdanog intervala slijedeći [5], a zatim ćemo vidjeti kako se ovi rezultati primjenjuju na generalizirane linearne modele.

Definicija 4.3.1. *Neka je* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ *slučajni uzorak duljine n te neka su* $L = l(Y_1, \ldots, Y_n) : \Omega \to \mathbb{R}$ *i* $U = u(Y_1, \ldots, Y_n) : \Omega \to \mathbb{R}$ *statistike. Slučajni interval* $[L, U]$ *je* $(1 - \alpha) \cdot 100\%$ **pouzdani interval parametra** θ *ako vrijedi*

$$\mathbb{P}(\theta \in [L, U]) = \mathbb{P}(L \leq \theta \leq U) = 1 - \alpha, \quad \alpha \in \langle 0, 1 \rangle.$$

U praksi se najčešće koriste 95% pouzdani intervali za parametar θ, tj.

$$\mathbb{P}(L \leq \theta \leq U) = 0.95,$$

što implicira da će u 95% svih realizacija ovog intervala prava vrijednost parametra θ biti unutar granica intervala, dok će u 5% slučajeva prava vrijednost parametra θ biti izvan tog intervala.

Kontrukcija pouzdanih intervala

Pouzdani se intervali u većini slučajeva konstruiraju *pivotnom metodom* koja pretpostavlja da postoji funkcija $h(\boldsymbol{Y}, \theta)$ strogo monotona u parametru θ s poznatom razdiobom. Kako je razdioba $h(\boldsymbol{Y}, \theta)$ poznata, tražimo h_1, h_2 takve da

$$\mathbb{P}(h(\boldsymbol{Y}; \theta) \in [h_1, h_2]) = \mathbb{P}(h_1 \leq h(\boldsymbol{Y}, \theta) \leq h_2) = 1 - \alpha. \tag{4.27}$$

Dakle, vjerojatnost da se statistika $h(\boldsymbol{Y}, \theta)$ nalazi u intervalu $[h_1, h_2]$ iznosi $1 - \alpha$. To znači da je vjerojatnost da statistika $h(\boldsymbol{Y}, \theta) \notin [h_1, h_2]$, tj. $h(\boldsymbol{Y}, \theta) \in \langle -\infty, h_1] \cup [h_2, +\infty \rangle$ iznosi α pa zbog disjunktnosti skupova vrijedi

$$\mathbb{P}(h(\boldsymbol{Y}, \theta) \leq h_1) + \mathbb{P}(h(\boldsymbol{Y}, \theta) \geq h_2) = \alpha.$$

Razlikujemo jednostrane i dvostrane intervale, koji procjenjuju raspon vrijednosti statistike u jednom ($h(\boldsymbol{Y}, \theta) \leq h_1$ ili $h(\boldsymbol{Y}, \theta) \geq h_2$) ili oba smjera ($h(\boldsymbol{Y}, \theta) \leq h_1$ i $h(\boldsymbol{Y}, \theta) \geq h_2$) te simetrične intervale ($h_1 = -h_2$) i intervale najkraće duljine. Za računanje pouzdanih intervala najčešće se koristi normalna distribucija (ili njena aproksimacija za dovoljno velike

uzorke) koja je simetrična i u tom su nam slučaju simetrični pouzdani intervali ujedno i intervali najkraće duljine. Dakle, tada vrijedi $h_1 = -h_2$ te

$$\mathbb{P}(h(\boldsymbol{Y}, \theta) \leq -h_2) = \mathbb{P}(h(\boldsymbol{Y}, \theta) \geq h_2) = \frac{\alpha}{2}.$$

Ako je $h(\boldsymbol{Y}, \theta) \sim AN(0, 1)$, onda je

$$\mathbb{P}(h(\boldsymbol{Y}, \theta) \leq -h_2) = \Phi(-h_2) = 1 - \Phi(h_2) = \frac{\alpha}{2}$$

pa su

$$h_2 = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$
$$h_1 = -h_2.$$

Dakle, rubovi intervala su kvantili razdiobe statistike $h(\boldsymbol{Y}, \theta)$ pa se često označavaju sukladno tome, na primjer za $h(\boldsymbol{Y}, \theta) \sim AN(0, 1)$ koristimo oznaku $h_2 = z_{1-\alpha/2}$ te za $h(\boldsymbol{Y}, \theta) \sim A\chi^2(\cdot)$ koristimo $h_2 = \chi^2_{1-\alpha/2}(\cdot)$.

Sada, zbog pretpostavke o strogoj monotonosti funkcije h u parametru θ postoje statistike L, U takve da $L \leq \theta \leq U$, odakle slijedi

$$\mathbb{P}(L \leq \theta \leq U) = 1 - \alpha.$$

Važno je naglasiti da funkcija $h(\boldsymbol{Y}, \theta)$ nije jedinstvena, tj. pouzdani se intervali mogu konstruirati na različite načine koristeći uzoračke distribucije statistika koje smo u ovome poglavlju opisali.

Primjer 4.3.1. *Iz primjera 3.2.1 želimo konstruirati dvostrani simetrični* 95% *pouzdani interval parametara* β_j *za* $j = 0, 1, 2$ *korištenjem statistike*

$$Z_j = \frac{b_j - \beta_j}{\sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})}} \sim AN(0, 1), \quad \textit{za dovoljno velike uzorke}. \tag{4.28}$$

Za $\alpha = 0.05$ *je* $z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) \approx 1.96$ *pa je*

$$\mathbb{P}(-z_{0.975} \leq Z_j \leq z_{0.975}) = \mathbb{P}(-1.96 \leq Z_j \leq 1.96) = 0.95.$$

Uvrštavanjem Z_j *dobivamo*

$$\begin{aligned} 0.95 &= \mathbb{P}(-1.96 \leq Z_j \leq 1.96) \\ &= \mathbb{P}\left(-1.96 \leq \frac{b_j - \beta_j}{\sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})}} \leq 1.96\right) \\ &= \mathbb{P}\left(b_j - 1.96\sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})} \leq \beta_j \leq b_j + 1.96\sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})}\right) \end{aligned}$$

pa je 95% *pouzdani interval parametra* β_j *za* $j = 0, 1, 2$ *dan s*

$$\beta_j \in \left[b_j - 1.96 \sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})}, b_j + 1.96 \sqrt{\mathcal{I}^{-1}_{jj}(\boldsymbol{b})} \right].$$

Tablica 4.2 prikazuje procjene parametara β_j *zajedno s njihovim 95% pouzdanim intervalima.*

Parametar	**Procjena**	**Donja granica**	**Gornja granica**
β_0	0.71650	0.34603	1.08697
β_1	0.27802	0.22846	0.32758
β_2	0.15880	0.10240	0.21520

Tablica 4.2: Procjene parametara s 95% pouzdanim intervalima

Vidjeli smo kako nam generalizirani linearni modeli omogućuju razumijevanje složenijih odnosa te prepoznavanje obrazaca kao što su trendovi i veze između varijabli. Nakon što smo procijenili parametre i provjerili preciznost naših procjena pomoću statističkih inferencija, dobili smo model koji se može koristiti za daljnje analize i donošenje odluka. Ove nam metode osiguravaju preciznost i jačaju pouzdanost naših zaključaka, što pomaže u boljem donošenju odluka.

Bibliografija

[1] A. J. Dobson i A. G. Barnett, *An Introduction to Generalized Linear Models, Second Edition*, Chapman & Hall/CRC Texts in Statistical Science, CRC Press, 2002.

[2] P. K. Dunn i G. K. Smyth, *Chapter 1: Statistical Models*, str. 1–30, Springer New York, 2018, https://doi.org/10.1007/978-1-4419-0118-7_1, (lipanj, 2024.).

[3] B. Efron, *Exponential Families in Theory and Practice*, Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2022.

[4] L. Fahrmeir i H. Kaufmann, *Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models*, The Annals of Statistics **13** (1985), br. 1, 342–368.

[5] M. Huzak, *Vjerojatnost i matematička statistika*, 2006., http://aktuari.math.pmf.unizg.hr/docs/vms.pdf, (srpanj, 2024.).

[6] ———, *Matematička statistika*, pogl. 1: Slučajne varijable i vektori, 2020./2021., https://web.math.pmf.unizg.hr/nastava/ms/files/ms1v9.pdf, (lipanj, 2024.).

[7] ———, *Matematička statistika*, pogl. 4: Statistička procjena, 2020./2021., https://web.math.pmf.unizg.hr/nastava/ms/files/ms4v5.pdf, (kolovoz, 2024.).

[8] T. Kralj i P. Lazić, *Statistički praktikum 1*, pogl. 5: Testiranje statističkih hipoteza, https://web.math.pmf.unizg.hr/nastava/statpr/files/sp1_vjezbe5_novo.pdf, (kolovoz, 2024.).

[9] N. Sarapa, *Teorija vjerojatnosti*, Udžbenici Sveučilišta u Zagrebu, Školska knjiga, 2002.

[10] A. S. Wahed, *BIOST 2083: Linear Models*, pogl. 3: Random Vectors and Multivariate Normal Distributions, 2007., https://sites.pitt.edu/~wahed/teaching/2083/fall07/Lecture3.pdf, (kolovoz, 2024.).

Sažetak

Za vektor međusobno nezavisnih slučajnih varijabli $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ za koje pretpostavljamo da ovise o vrijednostima $x_1, \ldots, x_p$ generalizirane linearne modele definiramo kao

$$\begin{cases} Y_i \sim EFD(\theta_i) \\ \mathbb{E}[Y_i] = \mu_i = q^{-1}(\boldsymbol{x}_i^T \boldsymbol{\beta}), \end{cases}$$

za $i = 1, \ldots, n$. Slučajne varijable $Y_i \sim EFD(\theta_i)$ dolaze iz eksponencijalne familije distribucija u standardnoj formi, čija gustoća ovisi o parametru θ_i. Ta familija uključuje brojne poznate statističke distribucije, uključujući binomnu, normalnu, Poissonovu i gama distribuciju te mnoge druge. Na ovaj način generalizirani linearni modeli omogućuju modeliranje zavisne varijable $\boldsymbol{Y}$ koja pripada i drugim distribucijama, ne samo normalnoj.

Nadalje, funkcija g iz gornjeg zapisa je monotono diferencijabila funkcija koju nazivamo funkcija poveznica. Ona povezuje distribuciju zavisne varijable, njeno očekivanje i varijancu, s linearnom kombinacijom nezavisnih varijabli $\boldsymbol{x}_i^T \boldsymbol{\beta}$. Na ovaj nam način generalizirani linearni modeli omogućuju modeliranje i nelinearnih veza.

Nepoznate parametre $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ procjenjujemo metodom najveće vjerodostojnosti, tražeći maksimum $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ funkcije log-vjerodostojnosti na temelju uzorka $\boldsymbol{y}$. Maksimizacija se svodi na traženje nultočaka parcijalnih derivacija log-vjerodostojnosti. Ovisno o složenosti, nultočke možemo tražiti analitički ili iterativnom težinskom metodom najmanjih kvadrata koja koristi Fisherov algoritam za poboljšanje procjena parametara putem težinske matrice i pseudo-odgovora.

Nakon što smo procijenili parametre generaliziranog linearnog modela i dobili jednadžbu modela, želimo provjeriti preciznost modela statističkim inferencijama koje uključuju testiranje statističkih hipoteza o značajnosti parametara, modela te usporedba modela korištenjem asimptotske $N(0, 1)$ i χ^2 distribucije, kao i izračunavanje pouzdanih intervala. Nakon što potvrdimo da smo dobili precizan model, možemo ga koristiti za donošenje odluka.

Summary

For a vector of independent random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, where we assume that they depend on the values $x_1, \ldots, x_p$, generalized linear models are defined as

$$\begin{cases} Y_i \sim EFD(\theta_i) \\ \mathbb{E}[Y_i] = \mu_i = q^{-1}(\boldsymbol{x}_i^T \boldsymbol{\beta}), \end{cases}$$

for $i = 1, \ldots, n$. The random variables $Y_i \sim EFD(\theta_i)$ come from the exponential family of distributions in standard form, with the density depending on the parameter θ_i. This family includes numerous well-known statistical distributions, such as the binomial, normal, Poisson, and gamma distributions, among many others. In this way, generalized linear models allow modeling of the dependent variable $\boldsymbol{Y}$ that belongs to distributions other than just the normal distribution.

Furthermore, the function g in the above expression is a monotonic differentiable function called the link function. It connects the distribution of the dependent variable, its expectation, and its variance with the linear combination of the independent variables $\boldsymbol{x}_i^T \boldsymbol{\beta}$. This way, generalized linear models enable the modeling of nonlinear relationships as well.

The unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ are estimated using the method of maximum likelihood by finding the maximum $\boldsymbol{b} = (b_0, b_1, \ldots, b_p)^T$ of the log-likelihood function based on the sample $\boldsymbol{y}$. Maximization reduces to finding the zeroes of the partial derivatives of the log-likelihood. Depending on the complexity, the zeroes can be found either analytically or by an iterative weighted least squares method, which uses Fisher's scoring algorithm to improve parameter estimates via a weight matrix and pseudo-responses.

Once the parameters of the generalized linear model have been estimated and the model equation obtained, we want to assess the accuracy of the model through statistical inferences, which include hypothesis testing for the significance of parameters, testing the overall model, and model comparison using the asymptotic $N(0, 1)$ and χ^2 distributions, as well as the computation of confidence intervals. Once we confirm the precision of the model, it can be used for decision-making.

Životopis

Zovem se Antonija Andrijević, rođena sam 17.09.1997. godine. Nakon što sam s odličnim uspjehom završila srednjoškolsko obrazovanje u X. gimnaziji Ivan Supek, odlučila sam nastaviti svoje školovanje u području matematike, te sam upisala preddiplomski studij Matematike, nastavničkog smjera, na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu. Tijekom preddiplomskog studija, koji sam uspješno završila 2020. godine, stekla sam temeljna znanja i vještine iz različitih matematičkih disciplina, s posebnim naglaskom na njihovu primjenu u obrazovanju.

Nakon završetka preddiplomskog studija, odlučila sam produbiti svoje znanje upisom diplomskog studija Matematičke statistike na istom fakultetu. Tijekom tog studija sve sam više otkrivala svoj interes za statistiku i njene primjene u različitim područjima. Posebno su me privukla istraživanja vezana uz analizu podataka, modeliranje i statističku interpretaciju, što je dodatno učvrstilo moj entuzijazam za daljnje usavršavanje i rad u ovoj dinamičnoj znanstvenoj disciplini.