UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 1. stopnja

Urh Peček

Bayesova linearna regresija

Delo diplomskega seminarja

Mentor: izred. prof. dr. Jaka Smrekar

Ljubljana, 2020

Kazalo

Bayesova linearna regresija

Povzetek

Bayesova linearna regresija je dobila ime po angleškem statistiku Thomasu Bayesu, ki je živel v prvi polovici 18. stoletja. Namen naloge je predstaviti temeljne ideje Bayesovega linearnega modeliranja, začenši s teorijo, ki stoji za Bayesovo statistiko, pa tudi nekatere praktične primere Bayesove linearne regresije. Na začetku je opisana regresijska analiza in njena uporaba. Podrobneje je obravnavana Bayesova statistika in izpeljava obrazca, ki temelji na Bayesovem izreku in je osnova, na kateri temelji Bayesovo sklepanje. Vključena je tudi praktična uporaba Bayesovega izreka ter posodabljanja na primeru testiranja prisotnosti bolezenskega stanja pri pacientu. Opisana in na poučnem primeru meta kovanca je predstavljena ocena parametrov z Bayesovim pristopom in njene razlike ter prednosti v primerjavi s frekventističnim pristopom ocenjevanja parametrov. Podan je normalni model linearne regresije. Na njem je predstavljena klasična linearna regresija, kjer je za oceno parametrov uporabljena metoda največjega verjetja. Kot glavni del naloge, je na normalnem modelu linearne regresije opisana Bayesova linearna regresija, kjer za oceno parametrov uporabimo Bayesov pristop. Izpeljana je skupna aposteriorna porazdelitev parametrov normalnega modela prek konjugiranih družin. Za konec sta Bayesovo posodabljanje in aposteriorna porazdelitev parametrov predstavljena tudi na primeru vpliva statističnih podatkov košarkašev lige NBA na njihovo plačo.

Bayesian Linear Regression

Abstract

The Bayesian linear regression was named after the English statistician Thomas Bayes, who lived in the first half of the 18th century. The purpose of this bachelors' thesis is to present the basic ideas of Bayesian linear modeling, starting with the theory behind Bayesian statistics, as well as some practical examples of Bayesian linear regression. Regression analysis and its application are described at the beginning. In more detail is discussed the Bayesian statistics and the derivation of the form, which is based on the Bayesian theorem and is the basis on which Bayes' reasoning is based. Also included is the practical application of Bayesian theorem and updating in the case of testing for the presence of a disease state in a patient. Described and on an instructive example of a coin flip is presented the estimation of parameters with the Bayesian approach and its differences and advantages in comparison with the frequency approach of parameter estimation. A normal linear regression model is given. A classical linear regression is presented on it, where the maximum probability method is used to estimate the parameters. As the main part of the thesis, the Bayesian linear regression is described on a normal linear regression model, where we use the Bayesian approach to estimate the parameters. A joint a posteriori distribution of the parameters of the normal model over conjugated families is derived. Finally, Bayes 'update and a posteriori distribution of parameters are also presented in the case of the impact of NBA basketball players' statistics on their salary.

Math. Subj. Class. (2010): 62J05

Ključne besede: linearna regresija, Bayesova statistika, Bayesov izrek, apriorna in aposteriorna porazdelitev, Bayesovo sklepanje, frekventistično sklepanje, Bayesovo posodabljanje

Keywords: linear regression, Bayesian statistics, Bayesian theorem, a priori and a posteriori distribution, Bayesian inference, frequency inference, Bayesian updating

1. Uvod

Regresijska analiza je vrsta napovedne analize, oziroma statistična metoda, ki nam omogoča preučevanje odnosa med dvema ali več spremenljivkami, ki nas zanimajo. Enostavneje, z njo preverjamo vpliv neodvisnih spremenljivk na odvisno spremenljivko. Neodvisnim spremenljivkam pravimo tudi prediktorji oziroma napovedovalci, medtem ko odvisno spremenljivko lahko poimenujemo tudi odzivna ali izhodna spremenljivka.

V splošnem z regresijo preverjamo dvoje:

(1) kako dobro deluje nabor neodvisnih spremenljivk pri napovedovanju odvisne spremenljivke,
(2) katere spremenljivke so zlasti pomembni napovedovalci odvisne spremenljivke in kako močno vplivajo na njene končne vrednosti.

Dobljene regresijske ocene uporabljamo za razlago razmerja med odvisnimi in neodvisnimi spremenljivkami. Te ocene lahko uporabljamo tudi za:

(1) določanje jakosti prediktorjev (primer v praksi: »Kakšna je moč povezave med odmerkom in učinkom?«),
(2) napovedovanje učinka in vpliva sprememb, torej, kako se spreminja odvisna spremenljivka, s spremembo ene ali več neodvisnih spremenljivk (primer v praksi: »Koliko dodatnih prihodkov od prodaje dobim za vsakih dodatnih tisoč evrov, porabljenih za trženje?«),
(3) napovedovanje trendov in prihodnjih vrednosti (primer v praksi: »Kakšna bo cena zlata v šestih mesecih?«).

Pri raziskovanju imamo na voljo več vrst linearnih regresijskih analiz. Pri izbiri modela za analizo je potrebno upoštevati primernost modela. Dodajanje neodvisnih spremenljivk regresijskemu modelu bo vedno povečalo pojasnjeno variabilnost odvisne spremenljivke, vendar pa z dodajanjem prekomernega števila neodvisnih spremenljivk zmanjšujemo splošnost modela, saj bi bile lahko nekatere spremenljivke pomembne samo zaradi naključja.

2. Bayesova statistika

Bayesova statistika je statistična veja, ki nam nudi matematična orodja za racionalno posodobitev naših subjektivnih prepričanj o proučevanem parametru glede na nove podatke. Bayesova statistika za opis negotovosti v svoje modele vključuje pogojno verjetnost, za izračun katere uporablja Bayesovo pravilo.

Bayesov model je osnovan na tem, da ima človek določena predhodna oziroma apriorna prepričanja o proučevani količini ali karakteristiki, ki jih z uporabo verjetnostnih modelov posodablja s pomočjo novo pridobljenih podatkov, z namenom dobiti "končno" oziroma aposteriorno prepričanje, ki se lahko uporabi kot podlaga za kasnejše odločitve.

Vključevanje subjektivnih informacij v postopek sklepanja je v nasprotju z drugo obliko statističnega sklepanja, znano kot klasična ali frekventistična statistika. Tudi iz klasičnega stališča se sklep izvede z uporabo verjetnostnih modelov, vendar se za razliko od Bayesovega modela upoštevajo samo informacije iz znanih podatkov, subjektivne informacije pa niso vključene v postopek odločanja. Ta predvideva, da so verjetnosti pogostost določenih naključnih dogodkov, ki se pojavljajo v dolgotrajnem ponavljajočem se preskušanju.

2.1. **Bayesovo pravilo.**

Osnova za obrazec, na katerem temelji Bayesovo sklepanje je Bayesovo pravilo. Ta izraža pogojno verjetnost dogodka na podlagi obrnjene pogojne verjetnosti ter brezpogojnih verjetnosti dveh dogodkov. Pogojna verjetnost nam daje pravilo za določitev verjetnosti dogodka A, glede na pojav drugega dogodka B, njena matematična definicija pa je naslednja.

Definicija 2.1. Pogojna verjetnost dogodka A, glede na dogodek B, za katerega predpostavimo P(B) $>$ 0, je

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

To preprosto pove, da je verjetnost dogodka A glede na to, da se je B zgodil, enaka razmerju verjetnosti, da sta se zgodila oba in verjetnosti, da se je zgodil samo dogodek B.

Če formulo pogojne verjetnosti uporabimo še v števcu, pridemo do Bayesovega pravila:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}. \tag{2}$$

Za njegovo kasnejšo uporabo je koristna menjava člena P(B), na desni strani zgornjega razmerja, v obliko P(B|H). Ustrezen zapis omogoča formula o popolni verjetnosti.

Trditev 2.2. *Imamo poskus, ki ga opravimo v dveh fazah. V prvi fazi se zgodi dogodek iz popolnega sistema končno ali števno mnogo dogodkov $H \in \Lambda$. V drugi fazi nas pa zanima dogodek B, ki je odvisen od realizacije prve faze. Formula o popolni verjetnosti nam izračuna verjetnost dogodka iz 2. faze in je enaka*

$$P(B) = \sum_{H \in \Lambda} P(B|H) \times P(H) \tag{3}$$

Izrek 2.3. *Bayesovo pravilo v diskretni obliki, kjer Λ predstavlja popoln sistem dogodkov prve faze, ki vsebuje dogodek A, zapišemo kot*

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} = \frac{P(B|A) \times P(A)}{\sum_{H \in \Lambda} P(B|H) \times P(H)} \tag{4}$$

Verjetnost dogodka A iz popolnega sistema dogodkov Λ prve faze poskusa, pogojno na opažen dogodek B druge faze poskusa.

Ker nam Bayesovo pravilo omogoča izračun pogojne verjetnosti dogodka prve faze poskusa glede na dogodek druge faze, se njegov koncept uporablja tudi pri testiranju v medicini, kjer se pogosto opazi pojav lažnih pozitivnih in lažnih negativnih rezultatov. Primer je testiranje prisotnosti bolezenskega stanja pri bolniku, kar predstavlja dogodek prve faze, glede na rezultat izvedbe testa, ki je dogodek druge faze poskusa.

3. Primer - Testiranje virusa Covid - 19

Da bi bolje razumeli pogojne verjetnosti, njihov pomen in posodabljanje našega prepričanja, razmislimo o primeru, ki vključuje virus Covid - 19. Za virus Covid - 19 ni poznanega zdravila in lažni pozitivni ter lažni negativni rezultati testiranja so zaradi družbene zaznamovanosti oziroma nadaljnjega širjenja virusa zelo neželeni. Lažen pozitiven rezultat pomeni, da nekomu, ki nima virusa, test pokaže, da je okužen. Lažen negativen rezultat pa pomeni, da nekomu, ki je okužen, sporočijo, da je zdrav.
Verjetnost lažnega pozitivnega rezultata se imenuje lažna pozitivna stopnja. Podobno se lažna negativna stopnja imenuje verjetnost lažnega negativnega rezultata. Obe stopnji sta pogojni verjetnosti.

Poimenujmo test bolezni C19. Želeli bi vedeti, kakšna je verjetnost, da ima nekdo virus Covid - 19 v primeru, da C19 pokaže pozitivno,
P(oseba je okužena | C19 je pozitiven).

Potrebujemo resnično pozitivno stopnjo C19, ki ji rečemo tudi občutljivost testa, ocenjena je kot P(C19 je pozitiven | oseba je okužena) = 80% = 0,80.
Resnična negativna stopnja, znana tudi kot specifičnost testa, je enaka
P(C19 je negativen | oseba ni okužena) = 98% = 0,98.
Potrebovali bomo tudi razširjenost virusa Covid - 19 v celotni populaciji. Recimo, da ga ima 5 od vsakih 1000 prebivalcev. Tako dobimo prvotno apriorno verjetnost:
P(oseba je okužena) = 5 / 1000 = 0,5% = 0,005.
Seveda sta občutljivost in specifičnost testa zgolj predvidevanji, za naš namen pa ju bomo obravnavali kot natančni vrednosti.

Čeprav v nobenem od zgornjih izrazov nismo določili vrednosti pogojno na rezultat C19, nam Bayesovo pravilo dovoljuje uporabo zgornjih številk, za izračun verjetnosti, ki jo iščemo. Bayesovo pravilo pravi:
P(oseba je okužena | C19 je pozitiven) =
= P(C19 pozitiven | oseba je okužena) × P(oseba je okužena) / P(C19 je pozitiven).

S pomočjo formule popolne verjetnosti izračunamo
P(C19 je pozitiven) = P(C19 pozitiven | oseba je okužena) × P(oseba je okužena)
+ P(C19 je pozitiven | oseba ni okužena) × P(oseba ni okužena) =
= 0,8 × 0,005 + (1 - 0,98) × (1 - 0,005) = 0,004 + 0,0199.

Končno dobimo aposteriorno verjetnost, da je oseba okužena,
P(oseba je okužena | C19 je pozitiven) = 0,004 / (0.004 + 0.0199) ≈ 16,73%.

Torej tudi, ko je test C19 pozitiven, je verjetnost da imamo virus Covid - 19 le 16,73%. Razlog zakaj je ta številka tako nizka, je posledica razširjenosti virusa Covid - 19. Pred testiranjem je bila verjetnost, da je oseba okužena le 0,5%, tako da pozitivni test to verjetnost drastično spremeni, vendar je še vedno precej nizka.

Kot smo videli, samo resnične pozitivne in resnične negativne stopnje testa ne povedo celotne zgodbe, veliko vlogo igra tudi razširjenost bolezni. Bayesovo pravilo

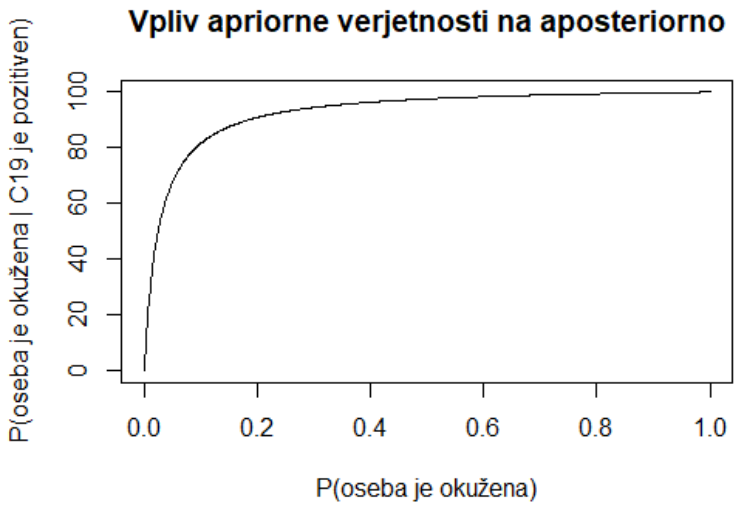
je orodje za prevajanje takšnih številk, v verjetnost bolezni po rezultatu testa.

Sedaj pa se vprašajmo, kako se spremeni P(oseba je okužena | C19 je pozitiven), v primeru, da je pri posamezniku verjetnost okužbe večja, kot pri naključno vzorčeni osebi iz populacije, P(oseba je okužena) $>$ 0,005 . Če ima oseba apriori večjo verjetnost okužbe in testira pozitivno, potem mora biti verjetnost, da je okužena z virusom pri rezultatu pozitivnega testa višja kot pri naključni osebi iz populacije, torej P(oseba je okužena | C19 je pozitiven) $>$ 16.73%.

3.1. **Bayesovo posodabljanje.** V prejšnjem razdelku smo videli, da pozitiven test C19 prinaša verjetnost okužbe z virusom Covid - 19 16,73%. Da bi dobili bolj prepričljiv odgovor na verjetnost okužbe, bi morda želeli narediti še drugi test C19, potem, ko je prvi pozitiven. Zanima nas torej, kakšna je verjetnost, da je testirana oseba okužena pod pogojem, da sta tako prvi kot drugi test pozitivna. Upoštevali bomo, da na pravilnost drugega testa prvi ne vpliva, torej je od njega neodvisen. (Čeprav to v praksi ne drži popolnoma.)

V primeru prvega testiranja, smo za izračun željene verjetnosti apriorno upoštevali P(oseba je okužena) = 0,0015. Sedaj bomo test ponovili v primeru, da je C19 pokazal pozitivno vrednost. Tako je posodobljena apriorna verjetnost enaka
P(oseba je okužena) = 0,1673.

Novo aposteriorno verjetnost po dveh testiranjih, dobimo po enakem postopku kot prej, z uporabo enakih resničnih pozitivnih in negativnih stopenj in je enaka
P(oseba je okužena | C19 je pozitiven dvakrat) = 0,889378 $\approx$ 89%. Vidimo, da po dveh pozitivnih testih lahko veliko bolj verjamemo v prisotnost virusa, kot pa v primeru zgolj enega pozitivnega testa. Če analogno izračunamo še verjetnost okužbe z virusom Covid – 19, po 3 pozitivnih testih, pri čemer upoštevamo
P(oseba je okužena) = 0,89, za rezultat dobimo
P(oseba je okužena | C19 je pozitiven trikrat) $\approx$ 99,7% (Clyde idr., 2020).



Slika 1. Vpliv apriorne verjetnosti okužbe, na aposteriorno verjetnost po pozitivnem testu.

4. Bayesova ocena parametrov

Poglejmo si, kako bi ocenili parametre modela z Bayesovim pristopom, ki temelji na že opisani Bayesovi statistiki, katere glavna lastnost je vključevanje subjektivnih informacij v postopek sklepanja.

Recimo, da je $\theta = (\theta_0, \theta_1, \ldots, \theta_{p-1})^T$ neznan vektor parametrov, ki bi ga radi ocenili. Naj bo $y = (y_1, \ldots, y_n)^T$ vektor, za katerega privzamemo, da je realizacija slučajne spremenljivke Y, katere porazdelitev pogojno na θ, je dana z gostoto $f(y|\theta)$. Predstavljamo si lahko, da realiziramo slučajen vektor $Y = (Y_1, Y_2, \ldots, Y_n)$.

Gostoto $f(y|\theta)$ imenujemo vzorčna gostota in jo lahko razumemo tudi kot funkcijo verjetja $L(y|\theta) = f(y|\theta)$. Pripomnimo, da če so slučajne spremenljivke Y_i neodvisne pogojno na θ, velja, da je funkcija verjetja realizacij $y = (y_1, \ldots, y_n)^T$ dana z $L(y|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta)$.

Če imata Y in θ družno gostoto $f(y, \theta)$, je $f(y|\theta) = \frac{f(y,\theta)}{f(\theta)}$ in $f(\theta|y) = \frac{f(y,\theta)}{f(y)}$, v kolikor sta $f(y)$ in $f(\theta)$ različni od 0, sicer v praksi vzamemo gostoto poljubne fiksne porazdelitve v spremenljivki Y oziroma θ.

Cilj Bayesovega ocenjevanja parametrov ni najti ene same »najboljše« vrednosti parametrov modela, temveč določiti aposteriorno porazdelitev vektorja parametrov θ. Aposteriorno porazdelitev θ definira Bayesov izrek, pri čemer je fikcija, da θ v prvi fazi poskusa dobimo z vzorčenjem iz neke, v našem primeru zvezne, porazdelitve. Aposteriorna porazdelitev θ pogojno na y, je dana z gostoto

$$f(\theta|y) = \frac{L(y|\theta)\, f(\theta)}{f(y)} = \frac{L(y|\theta)\, f(\theta)}{\int_{t\in\Theta} f(y|t)\, f(t)\, dt}. \tag{5}$$

Y in θ predstavljata slučajni spremenljivki oziroma vektorja, z vnaprej danima verjetnostnima porazdelitvama. Θ predstavlja parametrični prostor za θ in $f(\theta)$ gostoto njegove apriorne porazdelitve, ki temelji na znanju pred poskusom in v katero so vključene subjektivne informacije. Z apriorno porazdelitvijo v kombinaciji z vzorčno gostoto verjetnosti $L(y|\theta)$ vplivamo na končno, aposteriorno porazdelitev verjetnosti $f(\theta|y)$, ki vsebuje vse informacije, ki jih imamo o θ potem, ko opazimo vzorec y.

Bayesov izrek je mogoče razumeti kot način posodabljanja naše negotovosti, z novimi informacijami, kjer za izražanje negotovosti uporabljamo verjetnostne porazdelitve, model pa uporabljamo kot sredstvo za matematično kodiranje našega dogodka. Bayesovo sklepanje je naravni podaljšek naše intuicije. Pogosto imamo začetno oceno parametrov predstavljeno z apriorno porazdelitvijo in s tem, ko zbiramo nove podatke, spreminjamo svoje mišljenje o porazdelitvi parametrov. Aposteriorne porazdelitve so s postopkom zbiranja podatkov posodobljene apriorne porazdelitve.

4.1. **Konjugirane apriorne porazdelitve.**

Ključnega pomena za Bayesovo statistiko je apriorna porazdelitev parametrov, ki jih ocenjujemo in je vedno sporna, razen če obstaja fizični mehanizem vzorčenja, ki bi opravičil izbiro njihove porazdelitve. V praksi moramo Bayesove statistične analize upravičiti tretjim osebam, zato morajo apriorne porazdelitve, kolikor je to mogoče, temeljiti na prepričljivih zunanjih dokazih, oziroma moramo zagotoviti, da so šibko informativne.

Ena od možnosti za določanje apriornih porazdelitev je iskanje objektivnih ali neinformativnih apriornih porazdelitev, ki temeljijo na vzorčni porazdelitvi in jih uporabimo v primeru, ko naj bi bil prispevek subjektivnih stališč čim manjši. Na primer v znanstvenih objavah ali, ko o objektu analize nimamo dodatnih informacij. Alternativa pa je popolnoma subjektivistično stališče, kjer vse parametre modela določimo popolnoma prek presoje posameznika. Če je vzorec zelo velik, apriorna porazdelitev pri določitvi aposteriorne porazdelitve ne igra pomembnejše vloge in je skorajda vseeno za katero od zgoraj naštetih vrst apriornih porazdelitev se odločimo.

Pri določanju apriornih porazdelitev, se večinoma odločamo za konjugirane apriorne porazdelitve, katerih izbira je velikokrat povezana z matematičnimi in računskimi praktičnostmi, saj za vsako apriorno porazdelitev morda ni eksplicitne formule za gostoto aposteriorne porazdelitve.

Definicija 4.1. Družina apriornih porazdelitev $\mathcal{P}$ je konjugirana k družini vzorčnih porazdelitev, če vse pridružene aposteriorne porazdelitve ravno tako pripadajo $\mathcal{P}$.

5. Primer - met kovanca

Izvedli bomo poskus meta, ne nujno poštenega, kovanca. Radi bi ocenili verjetnost cifre, torej parametra p. Proučujemo Bernoullijev model s parametrom θ = p in parametričnim prostorom Θ = (0,1).

Apriorna porazdelitev je formalizacija predhodnega prepričanja o vrednosti parametra p. V primeru, da nimamo predhodnega mnenja, bomo za apriorno porazdelitev p privzeli enakomerno porazdelitev na intervalu (0,1). Torej je f(p) = 1 za $p \in (0,1)$. Lahko pa imamo mnenje o parametru p in smo prepričani, da se vrednost parametra p giblje okoli 1/2. V vsakem primeru za zvezno apriorno porazdelitev p velja $P(p \in (a,b)) = \int_a^b f(p)dp$.

Privzemimo, da smo v prvi fazi poskusa iz Θ izvlekli p s porazdelitveno gostoto f(p) in nato kovanec vržemo n-krat. Vsako ponovitev meta označimo z X_i.
Velja $P(X_i = 1 \mid p) = p$.
Za brezpogojno verjetnost s pomočjo formule o popolni verjetnosti velja
$P(X_i = 1) = \int_0^1 P(X_i = 1|\, p)\, f(p)\, dp = \int_0^1 p\, f(p)\, dp$.
V primeru neinformativne apriorne porazdelitve, kjer je f(p) = 1, dobimo
$P(X_i = 1) = \int_0^1 p\, dp = \frac{1}{2}$.

Pri n neodvisnih metih kovanca, pogojno na p tvorimo proučevano slučajno spremenljivko $X = X_1 + X_2 + \ldots + X_n$, ki nam pove število cifer pri n metih kovanca. Velja $P(X = k|p) = \binom{n}{k} p^k\,(1-p)^{n-k} = f(k|p)$ in $P(X = k) = \int_0^1 \binom{n}{k} p^k\,(1-p)^{n-k}\, f(p)\, dp$.

Ker parametra p ne poznamo in ga želimo oceniti, nas zanimajo aposteriorne verjetnosti, posodobljene z novimi opažanji.

V primeru, da smo v n metih opazili k cifer, za aposteriorno verjetnost velja:

$$P(p \in (a,b)|X = k) = \frac{P(p \in (a,b) \cap X = k)}{P(X = k)},$$
$$= \int_a^b \frac{P(X = k|p)\, f(p)}{P(X = k)}\, dp.$$

Pogojna porazdelitvena funkcija aposteriorne porazdelitve v točki p_0, pa je enaka $P(p \in (-\infty, p_0]|X = k) = \frac{1}{P(X=k)} \int_{-\infty}^{p_0} P(X = k|p)\, f(p)\, dp$.

Po osnovnem izreku analize izpeljemo poseben primer Bayesove formule za primer Bernoullijevega modela in tako aposteriorne gostote parametra p.

$$f(p_0|X = k) = \frac{d}{dp_0}\left(\frac{1}{P(X = k)} \int_{-\infty}^{p_0} P(X = k|p)\, f(p)\, dp\right),$$
$$= \frac{P(X = k|p_0)\, f(p_0)}{P(X = k)}.$$

Za cenilko parametra p, pogojno na opažen vzorec bi lahko vzeli pričakovano vrednost aposteriorne porazdelitve, $\hat{p} = \mathbb{E}[p|X = k] = \int_0^1 p\, f(p|X = k)\, dp$.

Konjugirana družina apriornih porazdelitev za binomsko vzorčno porazdelitev je družina beta porazdelitev s parametroma a, b $\in (0, \infty)$. Spodaj je naštetih nekaj njenih lastnosti.

$f_{Beta(a,b)}(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}$,
$B(a,b) = \int_0^1 p^{a-1}(1-p)^{b-1} dp$,
$\mathbb{E}[Beta(a,b)] = \frac{a}{a+b}$ in
$Var(Beta(a,b)) = \frac{ab}{(a+b)^2(a+b+1)}$.
Omeniti velja tudi da je Beta(1,1) = U(0,1), torej neinformativna apriorna porazdelitev.

Aposteriorna gostota je tako enaka

$$\begin{aligned} f(p|X=k) &= \frac{\binom{n}{k} p^k (1-p)^{n-k} f(p)}{\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p)\, dp}, \\ &= \frac{p^k (1-p)^{n-k} \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}}{\int_0^1 p^k (1-p)^{n-k} \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\, dp}, \\ &= \frac{\frac{1}{B(a,b)} p^{a+k-1}(1-p)^{b+n-k-1}}{\int_0^1 \frac{1}{B(a,b)} p^{a+k-1}(1-p)^{b+n-k-1}\, dp}, \\ &= \frac{p^{a+k-1}(1-p)^{b+n-k-1}}{B(a+k, b+n-k)}. \end{aligned}$$

Tu prepoznamo beta porazdelitev Beta(a+k, b+n-k). Aposteriorna porazdelitev p|X=k je torej Beta(a+k, b+n-k) in njena pričakovana vrednost je $\hat{p} = \frac{a+k}{a+b+n}$.

Za praktično uporabo zgornjih izračunov si oglejmo primer kjer kovanec denimo 12-krat neodvisno vržemo in opazimo 8 cifer. Radi bi ocenili parameter p. Velja P(cifra) = p in P(grb) = 1 - p za p $\in$ (0,1). Za oceno parametra p bomo upoštevali 2 pristopa, frekventistični in Bayesov pristop.

Oglejmo si najprej frekventistično oceno parametra p. Kovanec smo vrgli 12-krat, torej je n = 12, cifra pa se je pokazala 8-krat, torej je k = 8. Cenilko za p bomo poiskali z metodo največjega verjetja, ki pravi, da je cenilka tista vrednost p, v kateri funkcija verjetja L, kot funkcija parametra p, doseže maksimum.

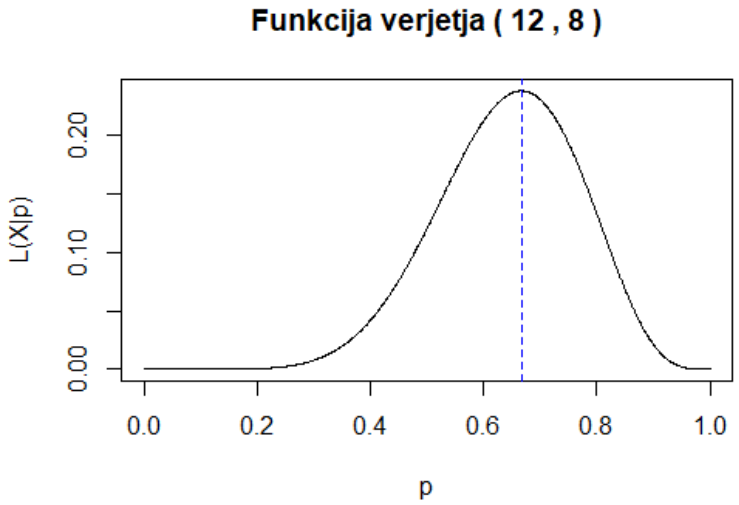
Verjetnost, da pri n neodvisnih metih kovanca dobimo k enic, je
$L(X=k|p) = \binom{n}{k} p^k (1-p)^{n-k}$.

$$\begin{aligned} L'(p) &= \binom{n}{k} [k p^{k-1}(1-p)^{n-k} - p^k (n-k)(1-p)^{n-k-1}], \\ &= \binom{n}{k} p^{k-1}(1-p)^{n-k-1}[k(1-p) - (n-k)p]. \end{aligned}$$

Njeno stacionarno točko dobimo z k(1 - p) - (n - k)p = 0 in s tem p = $\frac{k}{n}$, kar je tudi maksimum funkcije L, $\hat{p} = \frac{k}{n}$.

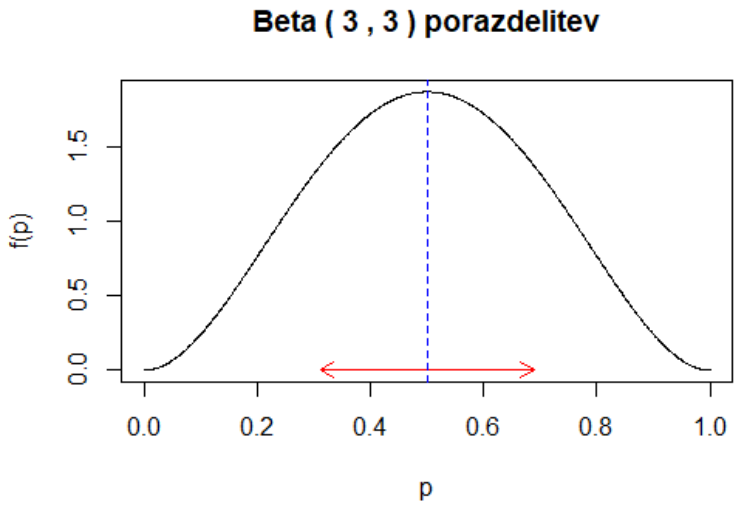
Tako smo s frekventističnim pristopom, prek metode največjega verjetja dobili cenilko za p, $\hat{p} = \frac{8}{12} \approx 0{,}67$. Ocena njene standardne napake pa je
$\hat{SE}_p = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0{,}136$.

Sedaj pa si oglejmo, kako bi parameter p ocenili z Bayesovim pristopom.



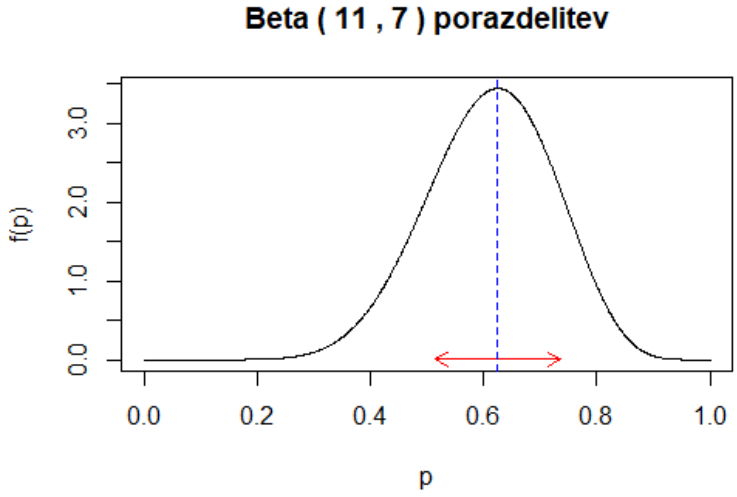
Slika 2. Graf funkcije verjetja, pri 12 metih kovanca in 8 cifrah.

Za apriorno porazdelitev parametra p upoštevamo beta porazdelitev Beta(a,b).
V kolikor menimo, da je kovanec pošten, mora veljati a = b. Če smo v poštenost kovanca trdno prepričani, moramo določiti dovolj majhen standardni odklon.
Recimo, da menimo, da je kovanec pošten, vendar o tem nismo trdno prepričani. Izberimo a = b = 3 in s tem apriorno porazdelitev Beta(3,3). Standardni odklon je enak SD(p) $= \sqrt{Var(Beta(3,3))} \approx 0{,}1889$.



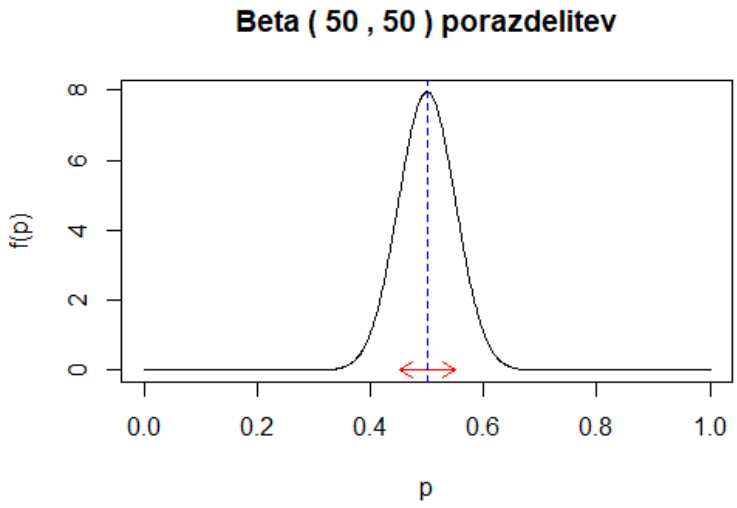
SLIKA 3. Graf gostote verjetnosti in standardnega odklona apriorne Beta(3,3) porazdelitve.

Aposteriorna porazdelitev parametra p je p|k$\sim$ Beta(11,7).
Za predstavo si oglejmo njeno matematično upanje, ki je enako $\mathbb{E}(p|k) = \frac{11}{11+7} \approx$ 0,6111 in standardni odklon SD(p|k) $= \sqrt{Var(Beta(11,7))} = \sqrt{\frac{11\times 7}{(11+7)^2(11+7+1)}} \approx$ 0,1118.



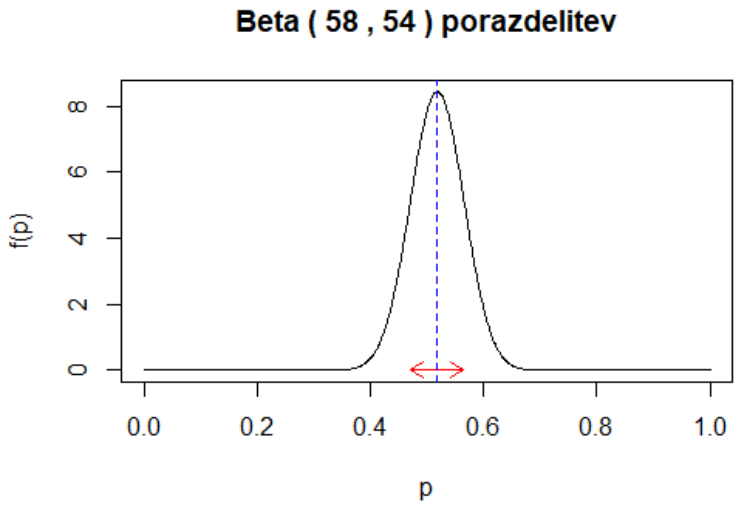
SLIKA 4. Graf gostote verjetnosti in standardnega odklona aposteriorne Beta(11,7) porazdelitve.

Kaj pa se zgodi, ko smo v poštenost kovanca prepričani trdneje? Ocene parametra p se lotimo kot zgoraj, vendar tokrat za apriorno porazdelitev p izberemo beta porazdelitev z manjšim standardnim odklonom. Določiti moramo večja a in b, kot v prvem primeru. Recimo, da določimo a = b = 50 in s tem apriorno Beta(50, 50) porazdelitev. Za standardni odklon sedaj velja SD(p) $\approx$ 0,00498.



SLIKA 5. Graf gostote verjetnosti in standardnega odklona apriorne Beta(50,50) porazdelitve.

Aposteriorna porazdelitev parametra p je p|k$\sim$ Beta(58,54).
Za predstavo je njeno matematično upanje enako $\mathbb{E}(p|k) = \frac{58}{58+54} \approx 0{,}5179$ in standardni odklon enak SD(p|k) $\approx$ 0,0470.

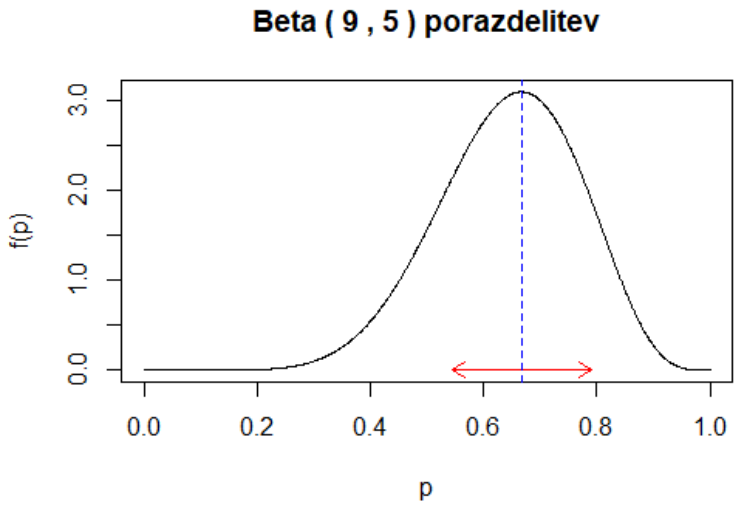


SLIKA 6. Graf gostote verjetnosti in standardnega odklona aposteriorne Beta(58,54) porazdelitve.

Oglejmo si še Bayesovo oceno parametra p z neinformativno apriorno porazdelitvijo. To kot rečeno uporabimo v primeru, ko o parametru nimamo izoblikovanih predhodnih mnenj in so edini vir informacij vnaprej dani podatki. Izberemo torej Beta(1,1) porazdelitev, ki je ekvivalentna enakomerni porazdelitvi na intervalu (0,1).

Ker za apriorno porazdelitev velja f(p) $= p^{1-1}(1-p)^{1-1}$, je aposteriorna porazdelitev enaka funkciji verjetja $f(p|k) \propto L(k|p)p(p) = L(k|p) \propto p^8(1-p)^4$, kar je

Beta(9,5) porazdelitev. Njena matematično upanje in standardni odklon sta enaka $\mathbb{E}(p|k) \approx 0{,}6429$ in SD(p|k) $\approx 0{,}1237$.



SLIKA 7. Graf gostote verjetnosti in standardnega odklona aposteriorne Beta(9,5) porazdelitve.

Če pa bi verjeli v to, da kovanec ni pošten, bi za apriorno porazdelitev uporabili nesimetrično beta porazdelitev in ponovili zgornje postopke za izračun aposteriorne porazdelitve parametra.

TABELA 1. Rezultati ocenjevanja parametra p

Pristop	Frekventistično	Bayes blago	Bayes prepričani	Bayes neinf.
Ocena p	0,6700	0,6111	0,5179	0,6429
Std. odklon	0,1360	0,1118	0,0470	0,1237

Opazimo, da če imamo predhodno znanje o parametrih modela oziroma o njih ugibamo, to lahko vključimo v naš model. Tudi, če o parametrih nimamo nobenih ocen, jih lahko vključimo v model, pri čemer uporabimo neinformativne apriorne porazdelitve. To je v nasprotju s frekventističnim pristopom, ki predpostavlja, da vse znanje o parametrih modela izhaja iz že znanih podatkov. Poleg tega lahko prek apriorne, v našem primeru Beta(a,b), porazdelitve določimo, kako veliko vlogo pri končni porazdelitvi parametrov, bo igralo našo predhodno znanje, oziroma prepričanje o parametru. To je opazno tudi iz pričakovane vrednosti aposteriorne porazdelitve $\hat{p} = \frac{a+k}{a+b+n} = \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{k}{n}$. Vidimo, da ob fiksni vrednosti parametrov a in b z rastjo velikosti podatkov, torej vrednosti n in k, v limiti velja $\hat{p} = \frac{k}{n}$, oziroma lahko ob končnih vrednostih n in k izberemo zelo velika a in b ter tako v limiti dobimo $\hat{p} = \frac{a}{a+b}$.

Vključevanje apriorne porazdelitve parametrov nam omogoča tudi, da količinsko določimo našo negotovost o modelu. Manj kot imamo podatkov in bolj kot smo negotovi v pravilnost predhodnega znanja, večji standardni odklon ji bomo dodelili. S povečevanjem količine podatkov se vpliv apriorne porazdelitve manjša in funkcija

verjetja »izpere« predhodno znanje o parametrih.

V našem primeru velja, da čim večja sta parametra a in b, tem večjo vlogo igra subjektivna odločitev glede apriorne porazdelitve. V aposteriorni porazdelitvi parametra p vidimo tudi, da z rastjo velikosti podatkov, kar predstavljata n in k, raste pomen vzorca in pada pomen subjektivnega prepričanja. V prvem primeru smo določili Beta(3,3) porazdelitev, kar je ekvivalentno temu, da smo v 4 metih dobili 2 cifri, v drugem primeru pa smo izbrali Beta(50,50) porazdelitev, kar je ekvivalentno 49 cifram v 98 metih. V kombinaciji z apriorno porazdelitvijo, v prvem primeru to pomeni 10 cifer v 16 metih, v drugem primeru pa 57 cifer v 110 metih, kar veliko bolj nakazuje poštenost (Toman 2012).

6. Normalni model linearne regresije

Kot že rečeno, je regresija pristop za modeliranje razmerja med odvisno spremenljivko in eno ali več pojasnjevalnimi spremenljivkami. V linearni regresiji se ta razmerja modelirajo s funkcijami prediktorjev in parametrov, ki so linearne v parametrih. Tako kot vse oblike regresijske analize, se tudi linearna regresija osredotoča na pogojno porazdelitev odziva glede na vrednosti parametrov, ki napovedujejo izhodno spremenljivko.

Enostavni model linearne regresije predvideva, da je odzivna spremenljivka (Y) definirana kot linearna kombinacija koeficientov (β), pomnoženih z naborom pojasnjevalnih spremenljivk (X).

Definicija 6.1. Naivni model enostavne linearne regresije je načeloma podan z enačbo

$$Y = \beta_0 + \beta_1 X + \epsilon. \tag{6}$$

Da iz tega naredimo parametričen model, moramo obravnavati n različno porazdeljenih slučajnih spremenljivk, v vlogi odvisne spremenljivke, pri čemer vsaki pripada svoja vrednost "neodvisne" spremenljivke in zapišemo kot

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_1 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_2 + \epsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 x_n + \epsilon_n \end{aligned}$$

Tu je X napovedna spremenljivka, katere vrednosti vplivajo na odvisno spremenljivko Y. Koeficienta β predstavljata parametra modela. Začetno vrednost odvisne spremenljivke predstavlja β_0, medtem ko je β_1 koeficient naklona, ki nam pove, kako močno v povprečju sprememba X vpliva na spremembo Y. Napako izračunane vrednosti Y predstavlja ϵ, ki ga imenujemo naključni vzorčni hrup in zajema vse druge dejavnike, ki vplivajo na vrednost Y in jih v model nismo vključili.

Model lahko posplošimo na poljubno število pojasnjevalnih spremenljivk in ga napišemo kot $Y = X\beta + \epsilon$, kjer velja:

- $\{Y_i\}_{i=1,2,\dots,n}$ so proučevane slučajne spremenljivke,
- $\{y_i\}_{i=1,2,\dots,n}$ so njihove opažene vrednosti, ki jih zapišemo v vektor $y = (y_1, y_2, \dots, y_n)^T$,
- $\{x_{ij}\}_{j=1,2,\dots,p}$ je j-ta pojasnjevalna spremenljivka za odziv y_i,
- vrednosti pojasnjevalnih spremenljivk zapišemo v matriko $X = (1, x_{i1}, \dots, x_{ip})_{i=1,\dots,n}$, ki vsebuje vrednosti vseh spremenljivk, za katere menimo, da vplivajo na odziv spremenljivke Y,
- $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ je vektor iskanih parametrov velikosti p + 1,
- $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ je vektor napak velikosti n.

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Za opažen odziv y_i ima linearni regresijski model obliko
$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i$, i = 1, ..., n.

Vektor opaženih vrednosti y lahko imenujemo tudi regresand. Vrstične vektorje matrike X, ki predstavljajo vhodne podatke, ki v kombinaciji s koeficienti vektorja β določajo pričakovane vrednosti slučajnih spremenljivk Y_i, imenujemo tudi regresorji. Običajno je kot eden od regresorjev vključena regresijska konstanta, na primer kot $x_{i0} = 1$ za vsak i=1, ..., n. Elementom vektorja β rečemo tudi regresijski koeficienti.

Definicija 6.2. Model za normalno linearno regresijo z odzivom Y, vzorčenim iz normalne porazdelitve, je podan z $Y = X\beta + \epsilon$, kjer je $\epsilon \sim N_n(0, \sigma^2 I_n)$. Predpostavimo, da so slučajne spremenljivke ϵ_i neodvisne. Za porazdelitev slučajnega vektorja Y tako velja: $Y_n \sim N(X\beta, \sigma^2 I_n)$.

Izhod Y je torej ustvarjen iz večrazsežne normalne porazdelitve, kjer srednjo vrednost predstavlja vektor parametrov, pomnožen z matriko prediktorjev, varianca pa je predstavljena kot kvadrat standardnega odklona, pomnoženega z matriko identitete. Parametre modela predstavljajo vektor koeficientov β in varianca σ^2. Okrog njih se osredotoča naša statistična ocena in sklepanje. Njihove aposteriorne porazdelitve naj bi najbolje razložile opažene vrednosti $\{y_i\}_{i=1,\ldots,n}$.

7. Klasična linearna regresija

Predpostavimo normalni model linearne regresije, $Y \sim N_n(X\beta, \sigma^2 I_n)$. Eden izmed načinov, kako oceniti parametre modela, torej koeficiente vektorja β in varianco σ^2 je frekventistični pristop z metodo največjega verjetja.

Definicija 7.1. Funkcija verjetja odziva Y, glede na vhodne podatke X, parametre, ki se skrivajo v vektorju β in vektor napake ϵ z varianco $\sigma^2 I_n$, L(y|β, σ^2), je ob predpostavki, da so odzivi Y_i neodvisne slučajne spremenljivke, kar produkt posameznih gostot verjetnosti:

$$\begin{aligned} L(y|\beta,\sigma^2) &= f_{(Y_1,Y_2,\ldots,Y_n)}(y_1,\ldots,y_n|\beta,\sigma^2), \\ &= \prod_{i=1}^{n} f_{Y_i}(y_i|\beta,\sigma^2), \\ &= (2\pi\sigma^2)^{-n/2} exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\}. \end{aligned} \tag{7}$$

Kot funkcija dejanskega vzorca $(y_1, \ldots, y_n)$ je to seveda gostota verjetnosti slučajnega vektorja Y, z neodvisnimi komponentami, izračunana v točkah $(y_1, \ldots, y_n)$. Tu gledamo na »L« kot funkcijo parametra β.

Ocena za β, σ^2 po metodi največjega verjetja je tak vektor parametrov $\widehat{(\beta, \sigma^2)}$, pri katerem ima funkcija verjetja L, kot funkcija parametrov β in σ^2, maksimum. Ker je logaritem naraščajoča funkcija in so vrednosti parametrov, ki maksimizirajo originalno funkcijo in njen naravni logaritem, enake, lahko za matematično udobje maksimiziramo logaritemsko funkcijo verjetja, dano z

$$\ln L(y|\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta). \tag{8}$$

Cenilko za (β, σ^2), $\widehat{(\beta, \sigma^2)}$, izračunano na opaženem vektorju y, po metodi največjega verjetja, dobimo z odvajanjem logaritemske funkcije verjetja, po parametru β in enačenjem z 0 ter hkratnem odvajanjem po parametru σ^2 in enačenjem z nič. Torej hkrati $\frac{\partial \ln L(y|\beta,\sigma^2)}{\partial \beta} = 0$ in $\frac{\partial \ln L(y|\beta,\sigma^2)}{\partial \sigma^2} = 0$.

8 Upoštevajoč, da je $y^T X\beta = \beta^T X^T y = (y^T X\beta)^T$ skalar, iz odvajanja po β dobimo

$$\frac{\partial \ln L(y|\beta,\sigma^2)}{\partial \beta} = \frac{1}{2\sigma^2}(2X^T y - 2X^T X\hat{\beta}) = 0.$$

Ocena največjega verjetja $\hat{\beta}(Y)$ za β je tako

$$\hat{\beta} = (X^T X)^{-1} X^T y \; oziroma \; \hat{\beta}(Y) = (X^T X)^{-1} X^T Y, \tag{9}$$

kjer privzamemo, da je matrika $X^T X$ obrnljiva.

Z odvajanjem logaritemske funkcije verjetja, po parametru σ^2 in enačenjem z 0, pa dobimo oceno največjega verjetja $\sigma^2(Y)$ za σ^2,

$$\begin{aligned}\hat{\sigma^2}(Y) &= \frac{1}{n}(Y - X\hat{\beta})^T(y - X\hat{\beta}),\\ &= \frac{1}{n}(Y - X(X^TX)^{-1}X^TY)^T(Y - X(X^TX)^{-1}X^TY),\\ &= \frac{1}{n}\left\|Y - X(X^TX)^{-1}X^TY\right\|^2.\end{aligned} \tag{10}$$

Cenilka za β, σ^2 po metodi največjega verjetja je tako
$\widehat{(\beta, \sigma^2)}(Y) = (X^TX)^{-1}X^TY,\ \frac{1}{n}\left\|Y - X(X^TX)^{-1}X^TY\right\|^2)$.

Oglejmo si matematično upanje in varianco naše cenilke $\hat{\beta} = \hat{\beta}(Y)$:
$\mathbb{E}[\hat{\beta}] = \mathbb{E}[\hat{\beta}(Y)]$,
$= (X^TX)^{-1}X^T\,\mathbb{E}[Y]$,
$= (X^TX)^{-1}X^TX\beta$
$= \beta$.
Zgoraj upoštevamo: $(X^TX)^{-1}X^TX = \mathrm{I}$.
Ker velja $\mathbb{E}[\hat{\beta}] = \beta$, je cenilka $\hat{\beta}$ nepristranska.

Njena varianca je enaka:
$Var(\hat{\beta}) = Var((X^TX)^{-1}X^TY)$,
$= (X^TX)^{-1}X^TVar(Y)((X^TX)^{-1}X^T)^T$,
$= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}$,
$= \sigma^2(X^TX)^{-1}$.

Kot neizrojena linearna transformacija slučajnega vektorja Y, je porazdelitev cenilke $\hat{\beta}$ enaka: $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$.

Za izračun porazdelitve cenilke $\hat{\sigma}^2(Y)$ si lahko pomagamo s preoblikovanjem le te.
Matriko $X \in \mathbb{R}^{n\times(p+1)}$ zapišemo kot $X = S\,U$, kjer je $S \in \mathbb{R}^{n\times(p+1)}$ matrika, z ortonormiranimi stolpci in $U \in \mathbb{R}^{(p+1)\times(p+1)}$ zgornje trikotna matrika, s pozitivnimi vrednostmi na diagonali.
Matriko S dopolnimo do ortonormirane baze velikosti n, $Q = [S\ S'] \in O(n)$, kjer je S'$\in \mathbb{R}^{n\times(n-(p+1))}$ matrika z ortonormiranimi stolpci.
Naprej velja $X^TX = (SU)^T(SU) = U^TS^TSU = U^TU$, kjer smo upoštevali ortonormiranost stolpcev matrike S.
Sledi, $X(X^TX)^{-1}X^T = SU(U^TU)^{-1}(SU)^T = SUU^{-1}(U^T)^{-1}U^TS^T = SS^T \in \mathbb{R}^{n\times n}$.
Tako je $\hat{\sigma}^2 = \frac{1}{n}\left\|Y - SS^TY\right\|^2 = \frac{1}{n}\left\|(I - SS^T)Y\right\|^2$.

Identično matriko $I \in \mathbb{R}^{n\times n}$ zapišemo kot $I = [S\ S']\,[S\ S']^T = [S\ S']\left[\begin{smallmatrix}S^T\\ S'^T\end{smallmatrix}\right] = SS^T + S'S'^T$.
Ker je $Q = [S\ S']$ ortogonalna matrika, katere ohranjajo dolžino, lahko zapišemo

$$\begin{aligned}
\left\|(I-SS^T)Y\right\|^2 &= \left\|Q^T(I-SS^T)Y\right\|^2, \\
&= \left\|(\left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right] - \left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right]SS^T)Y\right\|^2, \\
&= \left\|(\left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right] - \left[\begin{smallmatrix} S^TSS^T \\ S'^TSS^T \end{smallmatrix}\right])Y\right\|^2, \\
&= \left\|(\left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right] - \left[\begin{smallmatrix} S^T \\ 0 \end{smallmatrix}\right])Y\right\|^2, \\
&= \left\|(\left[\begin{smallmatrix} 0 \\ S'^T \end{smallmatrix}\right])Y\right\|^2 = \left\|S'^TY\right\|^2
\end{aligned}$$

Iz $Y \sim N_n(X\beta, \sigma^2 I)$ sledi $Q^TY = \left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right]Y \sim N(\left[\begin{smallmatrix} S^T \\ S'^T \end{smallmatrix}\right]X\beta, \sigma^2 Q^TQ)$ in
$\left[\begin{smallmatrix} S^TY \\ S'^TY \end{smallmatrix}\right] \sim N(\left[\begin{smallmatrix} S^TSU\beta \\ 0 \end{smallmatrix}\right], \sigma^2 I)$, zato je
$S'^TY \sim N_{n-(p+1)}(0, \sigma^2 I)$ in $\frac{1}{\sigma}S'^TY \sim N_{n-(p+1)}(0, I)$

Končno je kot vsota n-(p+1) kvadratov neodvisnih standardno normalno porazdeljenih slučajnih spremenljivk,
$\left\|\frac{1}{\sigma}S'^TY\right\|^2 \sim \chi^2_{n-(p+1)}$ in $\frac{1}{\sigma^2}\left\|S'^TY\right\|^2 \sim \chi^2_{n-(p+1)}$.

8. Bayesova linearna regresija

V klasični linearni regresiji smo upoštevali frekventistični pristop ocenjevanja želenih parametrov. Sedaj pa si poglejmo še Bayesov pristop ocenjevanja parametrov.

Ponovno bomo obravnavali normalni model linearne regresije
$(Y|\beta,\sigma^2) \sim N(X\beta,\sigma^2 I_n)$. Iskana aposteriorna porazdelitev parametrov modela β,σ^2 je odvisna od vhodnih podatkov, ki jih predstavlja matrika X in izhoda, ki je dan z vektorjem y. Kot smo povedali v razdelku 4.2 jo izračunamo s pomočjo Bayesovega pravila in je dana z

$$f(\beta,\sigma^2|y) = \frac{L(y|\beta,\sigma^2)f(\beta,\sigma^2)}{f(y)} \tag{11}$$

Gostoto aposteriorne porazdelitve $f(\beta,\sigma^2|y,X)$, glede na že znane vhodne in izhodne vrednosti, dobimo kot produkt funkcije verjetja dejanskega vzorca $L(y|\beta,\sigma^2)$, pomnožene z apriorno gostoto verjetnosti parametrov modela $f(\beta,\sigma^2)$, kjer vse skupaj delimo še z normalizacijsko konstanto $f(y)$.

Poglejmo si naprej Bayesov model linearne regresije z znano varianco σ^2.
Imamo torej $(Y|\beta) \sim N_n(X\beta,\sigma^2 I_n)$, kjer je σ^2 poznan in X $\in \mathbb{R}^{n\times d}$ fiksna matrika.
Gostota Y$|\beta$ je enaka $f(y|\beta) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\}$.
Predpostavimo, da je $\beta \sim N_d(m,\sigma^2 V)$, kjer je m $\in \mathbb{R}^d$ in V $\in \mathbb{R}^{d\times d}$ simetrična, pozitivno definitna matrika.
Torej velja $f(\beta) = (2\pi\sigma^2)^{-\frac{d}{2}} det(V)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(\beta-m)^T V^{-1}(\beta-m)\}$

Izračunajmo aposteriorno porazdelitev $\beta|Y$.

$$f(\beta|y) = \frac{f(y|\beta)f(\beta)}{f(y)} \propto f(y|\beta)f(\beta),$$

$$\propto exp\{-\frac{1}{2\sigma^2}((y-X\beta)^T(y-X\beta) + (\beta-m)^T V^{-1}(\beta-m))\}$$

Če zdaj v namen poenostavitve pogledamo oklepaj znotraj eksponentne funkcije
$(y-X\beta)^T(y-X\beta) + (\beta-m)^T V^{-1}(\beta-m) =$
$= \beta^T(X^TX+V^{-1})\beta - \beta^T(X^Ty+V^{-1}m) + (m^TV^{-1}m + y^Ty) - (y^TX + m^TV^{-1})\beta =$

Za udobje označimo $\Lambda = (X^TX+V^{-1})^{-1}$, matrika velikosti d×d in
$\mu = (X^TX+V^{-1})^{-1}(X^Ty+V^{-1}m)$, vektor dolžine d.

Nadaljujemo z zgornjim,
$= \beta^T\Lambda^{-1}\beta - \beta^T\Lambda^{-1}\mu - \mu^T\Lambda^{-1}\beta + m^TV^{-1}m + y^Ty \propto$
$\propto \beta^T\Lambda^{-1}\beta - \beta^T\Lambda^{-1}\mu - \mu^T\Lambda^{-1}\beta =$
$= (\beta-\mu)^T\Lambda^{-1}(\beta-\mu) - \mu^T\Lambda^{-1}\mu \propto$
$\propto (\beta-\mu)^T\Lambda^{-1}(\beta-\mu).$

Prepoznamo normalno, d - dimenzionalno porazdelitev, $N_d(\mu,\sigma^2\Lambda)$.
S tem smo dokazali, da je za model $(Y|\beta) \sim N_n(X\beta,\sigma^2 I_n)$, kjer je σ^2 poznan, $N(m,\sigma^2 V)$ konjugirana družina porazdelitev, kjer sta $m \in \mathbb{R}^d$ in $V \in \mathbb{R}^{d\times d}$ pozitivno definitna matrika.

Aposteriorne porazdelitve $\beta|y$ so torej oblike $N_d(\mu, \sigma^2\Lambda)$;
$\Lambda = (X^TX + V^{-1})^{-1}$, matrika velikosti d×d in $\mu = (X^TX + V^{-1})^{-1}(X^Ty + V^{-1}m)$.
Velja $f(\beta|y) = (2\pi\sigma^2)^{-\frac{d}{2}}(det\Lambda)^{-\frac{1}{2}}exp\{-\frac{1}{2\sigma^2} < \Lambda^{-1}(\beta-\mu), \beta-\mu >\}$,
torej $f(\beta|y) \propto exp\{-\frac{1}{2\sigma^2} < \Lambda^{-1}(\beta-\mu), \beta-\mu >\}$.

Sedaj si podobno oglejmo Bayesov model linearne regresije, vendar tokrat z znanim β.
Imamo torej $(Y|\sigma^2) \sim N_n(X\beta, \sigma^2 I_n)$, kjer je $\beta \in \mathbb{R}^d$ poznan in X $\in \mathbb{R}^{n\times d}$ fiksna matrika. Gostota Y$|\sigma^2$ je enaka $f(y|\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\}$.
Predpostavimo, da je apriorna porazdelitev σ^2 inverzna gama porazdelitev,
$\sigma^2 \sim IG(a,b)$; a,b > 0. Torej je $f(\sigma^2) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1}exp\{-\frac{b}{\sigma^2}\}$

Izračunajmo aposteriorno porazdelitev $\sigma^2|Y$.

$$f(\sigma^2|y) = \frac{f(y|\sigma^2)f(\sigma^2)}{f(y)} \propto f(y|\sigma^2)f(\sigma^2),$$
$$\propto (\sigma^2)^{-\frac{n}{2}}exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\}(\sigma^2)^{-a-1}exp\{-\frac{b}{\sigma^2}\},$$
$$= (\sigma^2)^{-\frac{n}{2}-a-1}exp\{-\frac{1}{\sigma^2}(\frac{1}{2}(y-X\beta)^T(y-X\beta)+b)\}$$

Prepoznamo inverzno gama porazdelitev, IG$(\frac{n}{2}+a, \frac{1}{2}(y-X\beta)^T(y-X\beta)+b)$. S tem smo dokazali, da je za model $(Y|\sigma^2) \sim N_n(X\beta, \sigma^2 I_n)$, kjer je $\beta \in \mathbb{R}^d$ poznan, konjugirana družina porazdelitev družina IG(a,b); a,b > 0.

Aposteriorne porazdelitve $\sigma^2|Y$ so torej oblike IG$(\frac{n}{2}+a, \frac{1}{2}(y-X\beta)^T(y-X\beta)+b)$.
Velja $f(\sigma^2|y) \propto (\sigma^2)^{-\frac{n}{2}-a-1}exp\{-\frac{1}{\sigma^2}(\frac{1}{2}(y-X\beta)^T(y-X\beta)+b)\}$.

Nazadnje pa si oglejmo še Bayesov model linearne regresije, kjer sta oba parametra β in σ^2 neznana.
Velja $(Y|\beta,\sigma^2) \sim N_n(X\beta, \sigma^2 I_n)$; X $\in \mathbb{R}^{n\times d}$ fiksna matrika, $\beta \in \mathbb{R}^d$ neznan in $\sigma^2 \in (0,\infty)$ neznan. Vzemimo $\beta|\sigma^2 \sim N_d(m, \sigma^2 V)$ in $\sigma^2 \sim$ IG(a, b).

Skupna apriorna porazdelitev β, σ^2 je podana z
$f(\beta,\sigma^2) = f(\beta|\sigma^2)f(\sigma^2)$,

$$= (2\pi\sigma^2)^{-\frac{d}{2}}(detV)^{-\frac{1}{2}}exp\{-\frac{1}{2\sigma^2}(\beta-m)^TV^{-1}(\beta-m)\}\times$$
$$\times \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1}exp\{-\frac{b}{\sigma^2}\},$$
$$\propto (\sigma^2)^{-(\frac{d}{2}+a)-1}exp\{-\frac{1}{\sigma^2}(\frac{1}{2}(\beta-m)^TV^{-1}(\beta-m)+b)\}.$$

Zanima mas skupna aposteriorna porazdelitev $\beta, \sigma^2|Y$.
$f(\beta,\sigma^2|y) = \frac{f(y|\beta,\sigma^2)f(\beta,\sigma^2)}{f(y)} = \frac{f(y|\beta,\sigma^2)f(\beta|\sigma^2)}{f(y|\sigma^2)}\frac{f(y|\sigma^2)f(\sigma^2)}{f(y)} = f(\beta|\sigma^2,y)\ f(\sigma^2|y)$.
Posebej bomo obravnavali prvi ulomek, (1) in drugi ulomek, (2).

(1) Iz modela z znano varianco σ^2 vemo,
$f(\beta|\sigma^2,y) = \frac{f(y|\beta,\sigma^2)f(\beta|\sigma^2)}{f(y|\sigma^2)} = f_{N_d(\mu,\sigma^2\Lambda)}(\beta)$

(2) Najprej si oglejmo izraz za $f(y|\sigma^2)$.

$$f(y|\sigma^2) = \int_\beta f(y,\beta|\sigma^2)d\beta = \int_\beta f(y|\beta,\sigma^2)f(\beta|\sigma^2)d\beta =$$
$$= \int_\beta \frac{f(y|\beta,\sigma^2)f(\beta|\sigma^2)}{f_{N_d(\mu,\sigma^2\Lambda)}(\beta)} f_{N_d(\mu,\sigma^2\Lambda)}(\beta)d\beta$$

$\frac{f(y|\beta,\sigma^2)f(\beta|\sigma^2)}{f_{N_d(\mu,\sigma^2\Lambda)}(\beta)} =$

$= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\}(2\pi\sigma^2)^{-\frac{d}{2}}(detV)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(\beta-m)^TV^{-1}(\beta-m)\}}{(2\pi\sigma^2)^{-\frac{d}{2}}(det\Lambda)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(\beta-\mu)^T\Lambda^{-1}(\beta-\mu)\}} =$

$= \frac{(2\pi\sigma^2)^{-\frac{n}{2}}(detV)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}((y-X\beta)^T(y-X\beta)+(\beta-m)^TV^{-1}(\beta-m))\}}{(det\Lambda)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(\beta-\mu)^T\Lambda^{-1}(\beta-\mu)\}} =$

$= \frac{(2\pi\sigma^2)^{-\frac{n}{2}}(detV)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}((\beta-\mu)^T\Lambda^{-}1(\beta-\mu)-\mu^T\Lambda^{-1}\mu+m^TV^{-1}m+y^Ty)\}}{(det\Lambda)^{-\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(\beta-\mu)^T\Lambda^{-1}(\beta-\mu)\}} =$

$= (2\pi\sigma^2)^{-\frac{n}{2}}(detV)^{-\frac{1}{2}}(det\Lambda)^{\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(-\mu^T\Lambda^{-1}\mu + m^TV^{-1}m + y^Ty)\},$

kar je neodvisno od β in lahko nesemo pred integral, ostane $\int_\beta f_{N_d(\mu,\sigma^2\Lambda)}(\beta)d\beta = 1$

Torej, $f(y|\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}(detV)^{-\frac{1}{2}}(det\Lambda)^{\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(-\mu^T\Lambda^{-1}\mu+m^TV^{-1}m+y^Ty)\}$.

Tako dobimo,

$\frac{f(y|\sigma^2)f(\sigma^2)}{f(y)} \propto f(y|\sigma^2)f(\sigma^2) =$

$= (2\pi\sigma^2)^{-\frac{n}{2}}(detV)^{-\frac{1}{2}}(det\Lambda)^{\frac{1}{2}} exp\{-\frac{1}{2\sigma^2}(-\mu^T\Lambda^{-1}\mu + m^TV^{-1}m + y^Ty)\}$

$\times \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} exp\{-\frac{b}{\sigma^2}\}$

$\propto (\sigma^2)^{-\frac{n}{2}-a-1} exp\{-\frac{1}{\sigma^2}\frac{1}{2}(-\mu^T\Lambda^{-1}\mu + m^TV^{-1}m + y^Ty + 2b)\}.$

Prepoznamo inverzno gama porazdelitev,
IG$(a+\frac{n}{2}, \frac{1}{2}(-\mu^T\Lambda^{-1}\mu + m^TV^{-1}m + y^Ty + 2b))$.

Ugotovili smo, da je gostota skupne aposteriorne porazdelitve $\beta, \sigma^2|Y$, produkt gostote normalne porazdelitve, $N_d(\mu, \sigma^2\Lambda)$ v spremenljivki β in gostote inverzne gama porazdelitve, IG$(a+\frac{n}{2}, \frac{1}{2}(-\mu^T\Lambda^{-1}\mu+m^TV^{-1}m+y^Ty+2b))$ v spremenljivki σ^2, torej

$$f(\beta,\sigma^2|y) = f_{N_d(\mu,\sigma^2\Lambda)}(\beta) f_{IG(a+\frac{n}{2},\frac{1}{2}(-\mu^T\Lambda^{-1}\mu+m^TV^{-1}m+y^Ty+2b)}(\sigma^2). \quad (12)$$

9. Primer - vpliv statističnih podatkov na plačo košarkašev lige NBA

Nazadnje si oglejmo še kako poteka zgoraj opisano Bayesovo posodabljanje v praksi, pri čemer si bomo pomagali s programom R.

Obravnavali bomo vpliv statističnih podatkov karakteristik igre košarkašev lige NBA na njihovo plačo. Za odvisno spremenljivko bomo vzeli plačo, izraženo v milijonih ameriških dolarjev (Salary), za pojasnjevalne spremenljivke pa si bomo izbrali odstotek zadetih metov za tri točke (ThreePT), dve točki (TwoPT) in prostih metov (FT) ter število skokov (TRB), podaj (AST), ukradenih žog (STL), blokad (BLK), izgubljenih žog (TOV), osebnih napak (PF) ter točk (PTS), vse izraženo kot povprečno število na odigrano minuto. Model je torej oblike
$\text{Salary} = \beta_0 + \beta_1\text{ThreePT} + \beta_2\text{TwoPt} + \beta_3\text{FT} + \beta_4\text{TRB} + \beta_5\text{AST} + \beta_6\text{STL} + \\ + \beta_7\text{BLK} + \beta_8\text{TOV} + \beta_9\text{PF} + \beta_{10}\text{PTS} + \epsilon.$

Za določitev apriornih porazdelitev β in σ^2 sem vzel zgoraj navedene statistične podatke košarkašev iz sezone 2006/07, ki predstavljajo matriko X in njihove plače v sezoni 2007/08, kar je realizacija y. Košarkaše, ki v teh dveh zaporednih sezonah niso nastopili, sem izključil iz obravnave. Začetni model iz katerega sem črpal potrebne podatke za apriorni porazdelitvi sem izračunal s pomočjo funkcije 'lm' v R, ki poda začetne ocene koeficientov $\hat{\beta}$ in nekatere statistične lastnosti. Za apriorno porazdelitev β pri znanem σ^2, sem vzel normalno porazdelitev z matematičnim upanjem in varianco, kot jih dobimo v klasični linearni regresiji z metodo največjega verjetja, pri čemer sem varianco prilagodil tako, da sem jo delil še z velikostjo vzorca, torej $\beta \sim N(\hat{\beta}, \sigma^2(X^TX)^{-1}/n)$. Za apriorno porazdelitev σ^2 pa sem vzel inverzno gama porazdelitev IG(a,b), pri čemer sem parametra ocenil prek metode momentov kot $a = \frac{\mu^2}{v} + 2$ in $b = \mu(\frac{\mu^2}{v} + 1)$, kjer je μ povprečna vrednost kvadratov residualov in $v = \sum (residuali - \mu)^2/(n-1)$.

Skupno aposteriorno porazdelitev sem izračunal, kot je opisano v teoriji zgoraj, pri čemer sem za nove podatke vzel statistične podatke košarkašev iz sezone 2018/19 in njihove plače v sezoni 2019/20.

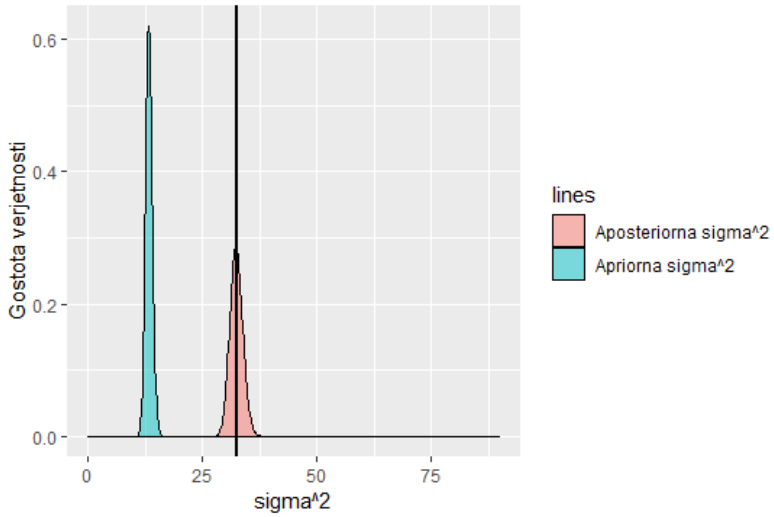
Za začetek si oglejmo osnovni model, kjer so s pomočjo statističnih podatkov sezone 2006/07 in njihovega vpliva na plačo, izračunane ocene zgoraj navedenih parametrov.

TABELA 2. Ocene parametrov $\hat{\beta} = (X^T X)^{-1} X^T y$ za X in y iz sezone 2006/07 oziroma 2007/08.

Parameter	Vrednost parametra	Pr(>|t|)
Reg. konst.	1,7780	0.3642
ThreePT	0,1272	0,9316
TwoPt	-3,4224	0,2316
FT	-2,5356	0,0983
TRB	17,7784	$2{,}58 \cdot 10^{-6}$
AST	19,1415	$1{,}12 \cdot 10^{-5}$
STL	-27,2258	0,0416
BLK	38,6669	0,0011
TOV	- 0,7700	0,9101
PF	-39,0053	$1{,}15 \cdot 10^{-9}$
PTS	16,6870	$9{,}69 \cdot 10^{-14}$

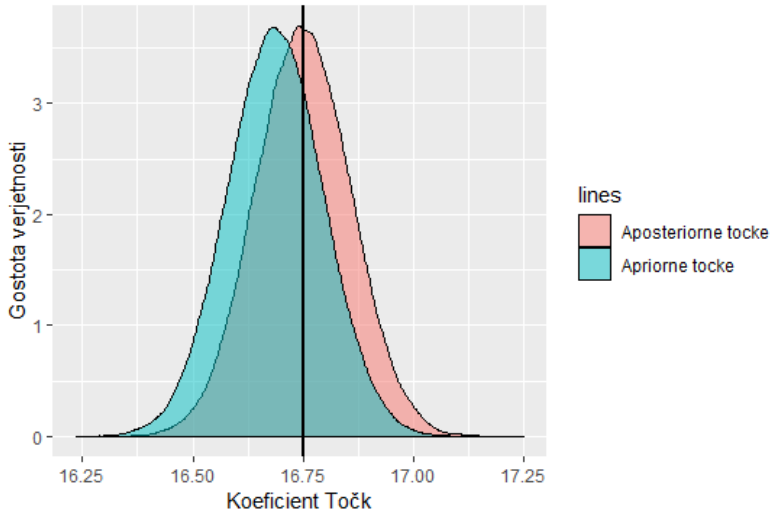
Kot bi lahko pričakovali, so močno statistično značilni predvsem parametri pri številu skokov, podaj ter točk. Zanimiva je statistična značilnost parametra pri osebnih napakah, ki ima negativno vrednost. To je posledica dejstva, da visoko plačani igralci ne delajo osebnih napak, temveč jih delajo predvsem njihovi soigralci, ki so za ekipo manj pomembni in imajo posledično nižjo plačo. Na prvi pogled bi nas lahko zmotila tudi statistična neznačilnost parametrov pri metih za 2 in 3 točke ter negativna vrednost pri metu za dve točki. Ti dve pa sta posledici dejstva, da imajo visoko plačani igralci veliko več svobode v igri in tako veliko mečejo na koš ter prevzemajo odgovornost ob koncu napadov, ko so meti neizdelani. Tako imajo ob velikem številu težkih metov tudi nižji odstotek uspešnih.

Spodaj so grafično predstavljene apriorne in aposteriorne vrednosti σ^2 ter statistično značilnih ocenjevanih parametrov. Za njihov prikaz sem iz apriornih in aposteriornih porazdelitev vzorčil milijon vrednosti in njihovo simulirano gostoto upodobil na grafu. Komentirane bodo tudi statistične značilnosti posameznih koeficientov, ki povedo pojasnjevalno moč posamezne pojasnjevalne spremenljivke, pri pojasnjevanju variabilnosti odvisne spremenljivke.



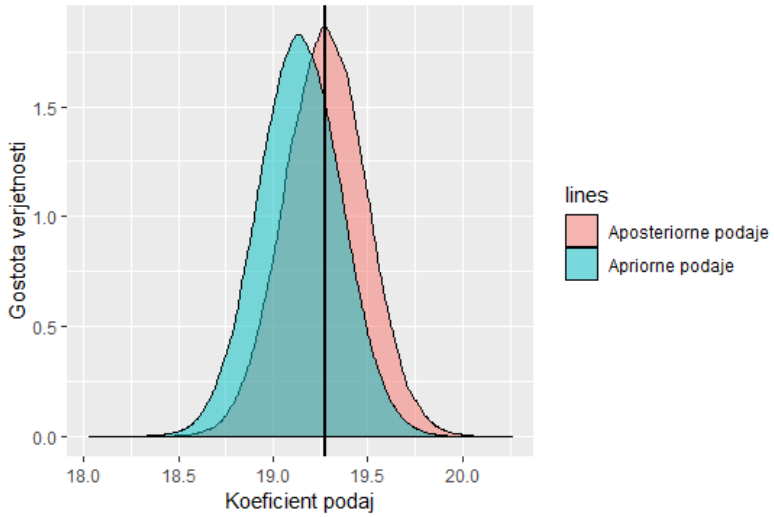
SLIKA 8. Gostota verjetnosti σ^2.

Vidimo, da je apriorna pričakovana vrednost σ^2 nižja od aposteriorne in znaša 13,3, v primerjavi z aposteriorno vrednostjo, ki je enaka 32,4. Prav tako je standardni odklon σ^2 višji pri aposteriorni porazdelitvi, kjer je enak 1,41, pri apriorni porazdelitvi pa znaša 0,72.



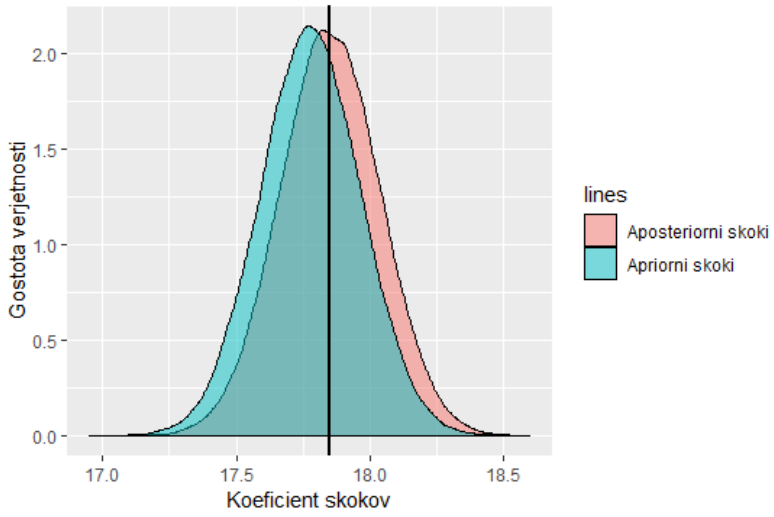
SLIKA 9. Gostota verjetnosti parametra pri točkah.

V grafu, ki prikazuje porazdelitev parametra točk je razvidno, da je aposteriorna porazdelitev v primerjavi z apriorno, predstavljena nekoliko bolj desno. Povprečna apriorna vrednost koeficienta točk znaša 16,7, medtem ko je povprečna aposteriorna vrednost enaka približno 16,74. Standardni odklon pa je v obeh primerih enak približno 0,115. Statistična značilnost koeficienta točk je enaka $9{,}69 \cdot 10^{-14}$ in točke igrajo največjo vlogo pri pojasnjevanju variabilnosti plače.



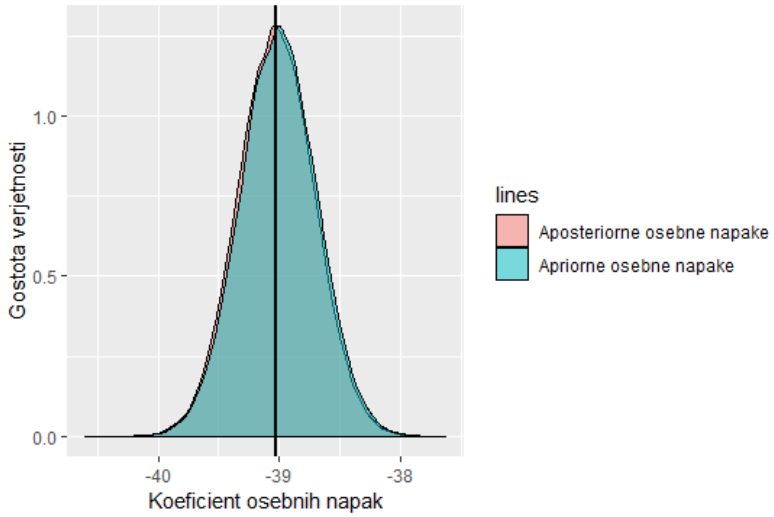
SLIKA 10. Gostota verjetnosti parametra pri podajah.

Tudi pri parametru podaj je podobno odstopanje v desno kot pri točkah. Apriorna pričakovana vrednost je približno 19,1, aposteriorno povprečje pa 19,3. Standardni odklon je v obeh primerih približno 0,23. Statistična značilnost koeficienta podaj je $1{,}12 \cdot 10^{-5}$ in podaje igrajo tretjo največjo vlogo pri pojasnjevanju variabilnosti plače.



SLIKA 11. Gostota verjetnosti parametra pri skokih.

Prav tako kot pri točkah in podajah je aposteriorna porazdelitev parametra skokov v primerjavi z apriorno zamaknjena nekoliko bolj v desno. Približek apriornega vzorčnega povprečja je 17,78 in aposteriornega vzorčnega povprečja 17,85 medtem ko je standardni odklon v obeh primerih enak približno 0,199. Statistična značilnost koeficienta skokov je $2{,}58 \cdot 10^{-6}$ in skoki igrajo četrto največjo vlogo pri pojasnjevanju variabilnosti plače.



Slika 12. Gostota verjetnosti parametra pri osebnih napakah.

Za konec pa je prikazana še razlika med apriorno in aposteriorno porazdelitvijo parametra osebnih napak, ki pa je v bistvu ni. Pričakovana vrednost znaša približno -39,0, standardni odklon pa je enak približno 0,33. Statistična značilnost koeficienta osebnih napak znaša $1{,}15 \cdot 10^{-9}$ in osebne napake igrajo drugo največjo vlogo pri pojasnjevanju variabilnosti plače. Kot pa je že rečeno, je to potrebno jemati precej z rezervo, saj gre povezava v drugi smeri, kot pa je bilo preučevano skozi celotno nalogo.

Pri porazdelitvah ostalih parametrov ne prihaja do večjih razlik in so si bolj kot ne identične. Njihove vrednosti so podane v spodnji tabeli.

Tabela 3. Primerjava apriornih in aposteriornih porazdelitev parametrov.

Parameter	Pričakovana vrednost	Standardni odklon
ThreePT	0,13	0,079
TwoPt	-3,4	0,15
FT	-2,5	0,08
STL	-27,2	0,71
BLK	38,7	0,63
TOV	-0,77	0,68

Na podlagi zgornje analize opazimo, da se je višina plač v ligi NBA iz leta 2007 na leto 2020 povečala. Na višino plač in posledično njihovo povečanje imajo največji vpliv število točk, podaj, skokov. Vendar pa razlike v velikosti med apriornimi in aposteriornimi vrednostmi parametrov niso prav velike, zato bi lahko sklepali, da so razlike v plačah majhne. Če pa upoštevamo to, da je bilo povprečno število točk na tekmo v sezoni 2006/07 98,7, v sezoni 2018/19 pa 111,2, torej 12,5 več, pa pridemo do zaključka, da so se plače povečale več, kot bi pričakovali zgolj iz ocen vrednosti parametrov. Plače so se v resnici povišale celo za več kot 100 odstotkov, vendar pa zaradi precejšne velikosti vzorca, iz katerega smo črpali apriorno znanje, aposteriorne vrednosti parametrov tega ne kažejo popolnoma.

10. Zaključek

Na začetku naloge smo spoznali osnove regresijske analize, njeno uporabo in temeljne pojme povezane z njo. Poleg tega, da dobljene regresijske ocene uporabljamo za razlago razmerja med eno ali več odvisnimi in neodvisnimi spremenljivkami, so bili navedeni tudi nekateri drugi primeri uporabe regresijskih ocen. Predstavljene so bile glavne ideje Bayesove statistike in Bayesovega sklepanja, ki temelji na uporabi Bayesovega pravila. Glavna lastnost Bayesove statistike je vključevanje subjektivnih informacij v postopek sklepanja, Bayesovo pravilo pa omogoča izračun pogojne verjetnosti dogodka prve faze poskusa glede na dogodek druge faze poskusa in je dan z $P(A|B) = \frac{P(B|A)\times P(A)}{P(B)} = \frac{P(B|A)\times P(A)}{\sum_{H\in\Lambda} P(B|H)\times P(H)}$. Ogledali smo si tudi zgled testiranja prisotnosti virusa Covid - 19, pri katerem smo videli uporabo elemenentarnega Bayesovega pravila in Bayesovo posodabljanje v praksi.

Opisana je bila ocena parametrov z Bayesovim pristopom, pri čemer je cilj ocenjevanja parametrov določiti njihovo aposteriorno porazdelitev, ki najbolje razloži opažene vrednosti y_i. Definira jo Bayesov izrek, tokrat zapisan v zvezni obliki, ki pravi da je aposteriorna porazdelitev θ pogojno na y dana z gostoto
$f(\theta|y) = \frac{L(y|\theta)\, f(\theta)}{f(y)} = \frac{L(y|\theta)\, f(\theta)}{\int_{t\in\Theta} f(y|t)\, f(t)\, dt}$.
Ključnega pomena pri izračunu aposteriorne porazdelitve je določitev apriornih porazdelitev ocenjevanih parametrov. Apriorne porazdelitve so lahko neinformativne, v primeru, ko naj bi bil prispevek subjektivnih stališč čim manjši, oziroma popolnoma subjektivne , kjer so vsi parametri modela določeni preko presoje posameznika. V praksi se največkrat poslužujemo konjugiranih družin porazdelitev. Velja, da je družina apriornih porazdelitev $\mathcal{P}$ konjugirana k družini vzorčnih porazdelitev, če vse pridružene aposteriorne porazdelitve ravno tako pripadajo $\mathcal{P}$.

Razliko med frekventističnem ocenjevanjem parametrov in Bayesovim pristopom smo si ogledali na primeru meta kovanca. Razvidna je bila vloga vključitve subjektivnih informacij v postopek sklepanja ter prilagajanje njihovega vpliva na aposteriorno porazdelitev prek izbire vrednosti parametrov v konjugirani družini apriornih porazdelitev.

V namen nadaljevanja naloge je bil predstavljen normalni model linearne regresije. Naivni model linearne regresije je načeloma podan z enačbo $Y = \beta_0 + \beta_1 X + \epsilon$. Iz naivnega modela lahko naredimo parametričen model, tega pa nadalje posplošimo na poljubno število pojasnjevalnih spremenljivk. Za opažen odziv y_i ima linearni regresijski model obliko $Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i = x_i^T\beta + \epsilon_i$, i = 1, . . . , n. Tako je model za normalno linearno regresijo z odzivom Y, vzorčenim iz normalne porazdelitve podan z $Y = X\beta + \epsilon$, kjer je $\epsilon \sim N_n(0, \sigma^2 I_n)$. Parametre modela predstavljajo vektor koeficientov β in varianca σ^2.
Izpeljana je bila ocena parametrov normalnega modela linearne regresije s frekventističnim pristopom, natančneje metodo največjega verjetja. Cenilka za β, σ^2 po metodi največjega verjetja je $\widehat{(\beta, \sigma^2)}(Y) = (X^TX)^{-1}X^TY, \frac{1}{n}\left\|Y - X(X^TX)^{-1}X^TY\right\|^2)$. Porazdelitev cenilke $\hat{\beta}$ je $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$, porazdelitev cenilke $\hat{\sigma^2}$ pa je dana z $\frac{1}{\sigma^2}\left\|S'^TY\right\|^2 \sim \chi^2_{n-(p+1)}$.

Nazadnje je bila izpeljana in na primeru vpliva statističnih podatkov košarkašev lige NBA na njihovo plačo, predstavljena skupna aposteriorna porazdelitev β in σ^2 za normalni model linearne regresije $(Y|\beta,\sigma^2) \sim N_n(X\beta,\sigma^2 I_n)$. Videli smo, da je v modelu z znano varianco σ^2, $\{\mathrm{N(m},\sigma^2\mathrm{V}) \mid \mathrm{m} \in \mathbb{R}^d$ in $\mathrm{V} \in \mathbb{R}^{d\times d}$ sim. poz. def. matrika$\}$ konjugirana družina porazdelitev. V modelu z znanim β, pa je konjugirana družina porazdelitev inverzna gama, $\{\mathrm{IG(a,b)} \mid \mathrm{a} > 0, \mathrm{b} > 0\}$. V kolikor pa sta oba parametra β in σ^2 neznana, z apriornima porazdelitvama $\beta|\sigma^2 \sim N_d(m,\sigma^2 V)$ in $\sigma^2 \sim \mathrm{IG(a,\ b)}$, pa je gostota skupne aposteriorne porazdelitve $\beta,\sigma^2|Y$ produkt gostote normalne porazdelitve $N_d(\mu,\sigma^2\Lambda)$ izračunane v spremenljivki β in gostote inverzne gama porazdelitve $\mathrm{IG}(a+\frac{n}{2}, \frac{1}{2}(-\mu^T\Lambda^{-1}\mu+m^TV^{-1}m+y^Ty+2b))$ izračunane v spremenljivki σ^2, kjer je $\Lambda = (X^TX+V^{-1})^{-1}$ in $\mu = (X^TX+V^{-1})^{-1}(X^Ty+V^{-1}m)$.

Uporabo Bayesovega pristopa ocenjevanja parametrov smo tako pokazali na praktičnih primerih in spoznali razlike v primerjavi s klasičnim pristopom. Opažen je bil širok spekter uporabe Bayesovega posodabljanja ter njegov velik plus v primeru predhodnih znanj o opazovani količini. Za nadaljnje raziskave bi nas lahko zanimala tudi prediktivna oziroma napovedna porazdelitev, ki bi se uporabila za napovedovanje neopaženih vrednosti odziva y. Iskali bi lahko tudi aposteriorne porazdelitve v primeru nekonjugiranih družin porazdelitev.

11. Priloga

Koda v programu r; vpliv statističnih podatkov na plačo košarkašev lige NBA.

```
# Uvoz potrebnih knjižnic.
library(tidyr)
library(readxl)
library(data.table)
library(dplyr)
library(readr)
library(ggplot2)
library(knitr)
library(rvest)
library(gsubfn)
library(reshape2)
library(shiny)
sl <- locale("sl", decimal_mark=",", grouping_mark=".")

# UVOZ PODATKOV

# Uvoz statitistike sezone 2018/19; basketball-reference.com
uvozi.Statistiko1 <- function() {
  data <- read_csv("Stat19.txt", locale=locale(encoding="Windows-1250"))
}

# Ureditev podatkov
Stat1 <- uvozi.Statistiko1()
Stat1 <- Stat1[,-1]
Stat1$Player = gsub("^(.*)\\\\.*", "\\1", Stat1$Player)
names(Stat1)[4] <- "FG"
names(Stat1)[5] <- "ThreePT"
names(Stat1)[6] <- "TwoPT"
names(Stat1)[7] <- "FT"
Stat1$FG[is.na(Stat1$FG)]<- 0
Stat1$ThreePT[is.na(Stat1$ThreePT)]<- 0
Stat1$TwoPT[is.na(Stat1$TwoPT)]<- 0
Stat1$FT[is.na(Stat1$FT)]<- 0
Stat1 <- Stat1 %>% distinct(Player, .keep_all = TRUE)
# Izbrišem stolpce, ki se mi ne zdijo potrebni, bom jih pa morda še potreboval.
Stat1 <- Stat1[,-c(2,4,10)]
# Posodobim stolpce da bodo kazali statistiko glede na odigrane minute.
Stat1 <- Stat1 %>%
  mutate_at(vars(ORB: PTS), funs(./Stat1$MP )) %>%
  mutate_at(6:13, funs(round(., 4))) %>%
  select(-MP)

# Uvoz statistike sezone 2006/07; basketball-reference.com
uvozi.Statistiko0 <- function() {
  data <- read_csv("Stat07.txt", locale=locale(encoding="Windows-1250"))
}

# Ureditev podatkov.
Stat0 <- uvozi.Statistiko0()
Stat0 <- Stat0[,-1]
Stat0$Player = gsub("^(.*)\\\\.*", "\\1", Stat0$Player)
names(Stat0)[4] <- "FG"
names(Stat0)[5] <- "ThreePT"
names(Stat0)[6] <- "TwoPT"
names(Stat0)[7] <- "FT"
Stat0$FG[is.na(Stat0$FG)]<- 0
```

```
Stat0$ThreePT[is.na(Stat0$ThreePT)]<- 0
Stat0$TwoPT[is.na(Stat0$TwoPT)]<- 0
Stat0$FT[is.na(Stat0$FT)]<- 0
Stat0 <- Stat0 %>% distinct(Player, .keep_all = TRUE)
# Izbrišem stolpce, ki se mi ne zdijo potrebni, bom jih pa morda še potreboval.
Stat0 <- Stat0[,-c(2,4,10)]
# Sedaj posodobim stolpce da bodo kazali statistiko glede na odigrane minute.
Stat0 <- Stat0 %>%
  mutate_at(vars(ORB: PTS), funs(./Stat0$MP )) %>%
  mutate_at(6:13, funs(round(., 4))) %>%
  select(-MP)

# Uvoz plač za sezono 2007/08; hoopshype.com za
uvozi.Place0 <- function(){
  link <- "https://hoopshype.com/salaries/players/2007-2008/"
  stran <- html_session(link) %>% read_html()
  Place0 <- stran %>% html_nodes(xpath="//table[@class='hh-salaries-ranking-table
      hh-salaries-table-sortable responsive']") %>%
    .[[1]] %>% html_table()
}

# Ureditev podatkov.
Place0 <- uvozi.Place0()
colnames(Place0) <- Place0[1,]
Place0 <- Place0[-1, -1 ]
names(Place0)[2] <- "Salary"
Place0 <- Place0 %>% mutate(Salary=parse_number(Salary,
                                                locale=locale(grouping_mark=","))
                            )

Place0 <- Place0[,-3]
Place0$Salary <- signif(Place0$Salary, digits=5)

# Uvoz plač za sezono 2019/20; hoopshype.com
uvozi.Place1 <- function(){
  link <- "https://hoopshype.com/salaries/players/"
  stran <- html_session(link) %>% read_html()
  Place1 <- stran %>% html_nodes(xpath="//table[@class='hh-salaries-ranking-table
      hh-salaries-table-sortable responsive']") %>%
    .[[1]] %>% html_table()
}

# Ureditev podatkov.
Place1 <- uvozi.Place1()
colnames(Place1) <- Place1[1,]
Place1 <- Place1[-1, -c(1,4:8) ]
names(Place1)[2] <- "Salary"
Place1 <- Place1 %>% mutate(Salary=parse_number(Salary,
                                                locale=locale(grouping_mark=","))
                            )

Place1 <- Place1[,-3]
Place1$Salary <- signif(Place1$Salary, digits=5)

# Združitev tabel, igralci, ki niso nastopali v upoštevanih dveh zaporednih sezonah
    so izpuščeni.
Place0Stat0 <- inner_join(Place0, Stat0, by="Player")
Place1Stat1 <- inner_join(Place1, Stat1, by="Player")

Place0Stat0 <- transform(Place0Stat0, SalaryMUSD = Salary / 1000000)
Place0Stat0 <- Place0Stat0[,-2]
Place0Stat0 <- Place0Stat0[c(1,13,2:12)]

Place1Stat1 <- inner_join(Place0, Stat0, by="Player")
```

```
Place1Stat1 <- inner_join(Place1, Stat1, by="Player")

Place1Stat1 <- transform(Place1Stat1, SalaryMUSD = Salary / 1000000)
Place1Stat1 <- Place1Stat1[,-2]
Place1Stat1 <- Place1Stat1[c(1,13,2:12)]


Place0Stat0 <- Place0Stat0[,-1]
Place1Stat1 <- Place1Stat1[,-1]

Place0Stat0 <- transform(Place0Stat0, TRB = ORB+DRB)
Place0Stat0 <- Place0Stat0[,-c(5,6)]
Place0Stat0 <- Place0Stat0[c(1:4,11,5:10)]

Place1Stat1 <- transform(Place1Stat1, TRB = ORB+DRB)
Place1Stat1 <- Place1Stat1[,-c(5,6)]
Place1Stat1 <- Place1Stat1[c(1:4,11,5:10)]

#
    ------------------------------------------------------------------------------------------------------------


library(MASS)
library(tidyverse)
library(invgamma)
library(ggplot2)
library(base)

# REGRESIJA

# Osnovni model - predhodno znanje.
ModelOsnovni <- lm(SalaryMUSD ~ ThreePT + TwoPT + FT + TRB + AST + STL + BLK + TOV
    + PF + PTS, data = Place0Stat0)
summary(ModelOsnovni)

#Klasični model prek katerega pridobimo apriorne porazdelitve.
X0 <- model.matrix(~ ThreePT + TwoPT + FT + TRB + AST + STL + BLK + TOV + PF + PTS,
     data = Place0Stat0)
k <- ncol(X0)
n0 <- nrow(X0)
y0 <- Place0Stat0$SalaryMUSD
v0 <- solve(t(X0) %*% X0)
# Ocena za beta0 po metodi največjega verjetja -> beta = (X'X)^(-1) * X'y.
beta_hat0 <- v0 %*% t(X0) %*% y0
# Vzorec ocen realizacij iz ocenjenih parametrov tj. y_hat = X * beta_hat.
 y_hat0 <- X0 %*% beta_hat0
# Ostanki 2019.
 Place0Stat0$residuals <- Place0Stat0$SalaryMUSD - y_hat0

# Določitev apriornih porazdelitev.
# Črpanje znanja iz vzorca statistike 06/07 in plač 07/08.

 # Apriorna porazdelitev sigma^2, vzorčim sigma^2.
muIG0 <- mean(Place0Stat0$residuals^2)
varIG0 <- sum((Place0Stat0$residuals - muIG0)^2) / (nrow(Place0Stat0) - 1) / n0
aHat <- muIG0^2 / varIG0 + 2
bHat <- muIG0 * (muIG0^2 / varIG0 + 1)
sigma2 <- 1/rgamma(n = 1, shape = aHat, rate = bHat)
# Naključni vzorčni hrup, epsilon ~ N(0, sigma^2*I).
epsilon0 <- mvrnorm(n = 1, mu = rep(0, n0), Sigma = sigma2 * diag(n0))

# Apriorna porazdelitev beta|sigma^2.
```

```
# beta|sigma^2 ~ N(m, sigma^2 * V), dosedanje znanje nam da metoda največjega
    verjetja.
# Tako menimo, da je beta ~ N(beta_hat0, sigma^2 * (X'X)^-1) = (m, sigma^2 * V).
m <- beta_hat0
V <- v0 / n0
betasigma2 <- mvrnorm(n = 1, mu = m, Sigma = V * sigma2)

# Izračun aposteriornih porazdelitev.
# Nov vzorec s katerim bomo posodobili prejšnji porazdelitvi.
X <- model.matrix(~ ThreePT + TwoPT + FT + TRB + AST + STL + BLK + TOV + PF + PTS,
    data = Place1Stat1)
y <- Place1Stat1$SalaryMUSD

# Skupna aposteriorna porazdelitev beta, sigma^2 je sestavljena iz produkta
    normalne in inverzne gama porazdelitve.
# Normalno - aposteriorna beta, inverzna gama - aposteriorna sigma^2.
Lambda <- solve(t(X) %*% X + solve(V))
mu <- solve(t(X) %*% X + solve(V)) %*% (t(X) %*% Place1Stat1$SalaryMUSD + solve(V)
    %*% m)
# Aposteriorna porazdelitev sigma^2.
a <- aHat + nrow(X)/2
b <- 1/2 * (-t(mu)%*%solve(Lambda)%*%mu + t(m)%*%solve(V)%*%m + t(y)%*%y + 2*bHat)
sigma2Post <- 1/rgamma(n = 1, shape = a, rate = b)
betaPost <- mvrnorm(n = 1, mu = mu, Sigma = Lambda * sigma2Post)
# Aposteriorni epsilon, dobljen iz aposteriornega sigma^2.
epsilonPost <- mvrnorm(n=1, mu = rep(0, nrow(Place1Stat1)), Sigma = sigma2Post *
    diag(nrow(Place1Stat1)))

# Primerjava apriornih in aposteriornih parametrov.

# Razlikovanje apriorne in aposteriorne porazdelitve sigma^2.
GraficnaApriornaSigma2 <- 1/rgamma(n = 100000, shape = aHat, rate =bHat)
GraficnaAposteriornaSigma2 <- 1/rgamma(n = 100000, shape = a, rate = b)
Sigma2Graf <- data.frame(Sigma2Plot =c(GraficnaApriornaSigma2, Posterior =
    GraficnaAposteriornaSigma2),
                         lines = rep(c("Apriorna sigma^2", "Aposteriorna sigma^2"),
                              each =100000))

ggplot(Sigma2Graf, aes(x = Sigma2Plot , fill = lines)) + xlab("sigma^2") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaSigma2), size=1) +
  xlim(0,90)

#Razlikovanje apriorne in aposteriorne porazdelitve beta.
GraficnaApriornaBetaSigma2 <- mvrnorm(n = 100000, mu = m, Sigma = V * sigma2)
GraficnaApriornaBetaSigma2 <- data.frame(GraficnaApriornaBetaSigma2)
GraficnaAposteriornaBeta <- mvrnorm(n = 100000, mu = mu, Sigma = Lambda * sigma2)
GraficnaAposteriornaBeta <- data.frame(GraficnaAposteriornaBeta)

# Točke
Points <- data.frame(Tocke =c(GraficnaApriornaBetaSigma2$PTS, Posterior =
    GraficnaAposteriornaBeta$PTS),
                     lines = rep(c("Apriorne tocke", "Aposteriorne tocke"), each
                         =100000))
ggplot(Points, aes(x = Tocke , fill = lines)) + xlab("Koeficient Točk") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$PTS), size=1)

# Podaje
Assists <- data.frame(Podaje =c(GraficnaApriornaBetaSigma2$AST, Posterior =
    GraficnaAposteriornaBeta$AST),
```

```
                      lines = rep(c("Apriorne podaje", "Aposteriorne podaje"), each
                          =100000))
ggplot(Assists, aes(x = Podaje , fill = lines)) + xlab("Koeficient podaj") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$AST), size=1)

# Skoki
Rebounds <- data.frame(Skoki =c(GraficnaApriornaBetaSigma2$TRB, Posterior =
    GraficnaAposteriornaBeta$TRB),
                       lines = rep(c("Apriorni skoki", "Aposteriorni skoki"), each
                          =100000))
ggplot(Rebounds, aes(x = Skoki , fill = lines)) + xlab("Koeficient skokov") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$TRB), size=1)

# Osebne napake
Fouls <- data.frame(Osebne =c(GraficnaApriornaBetaSigma2$PF, Posterior =
    GraficnaAposteriornaBeta$PF),
                    lines = rep(c("Apriorne osebne napake", "Aposteriorne osebne
                        napake"), each =100000))
ggplot(Fouls, aes(x = Osebne , fill = lines)) + xlab("Koeficient osebnih napak") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$PF), size=1)

# Izgubljene žoge
Turnovers <- data.frame(Izgubljene =c(GraficnaApriornaBetaSigma2$TOV, Posterior =
    GraficnaAposteriornaBeta$TOV),
                        lines = rep(c("Apriorne izgubljene žoge", "Aposteriorne
                            izgubljene žoge"), each =100000))
ggplot(Turnovers, aes(x = Izgubljene , fill = lines)) + xlab("Koeficient
    izgubljenih žog") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$TOV), size=1)

# Prosti meti
FreeThrows <- data.frame(Prosti =c(GraficnaApriornaBetaSigma2$FT, Posterior =
    GraficnaAposteriornaBeta$FT),
                         lines = rep(c("Apriorni prosti meti", "Aposteriorni prosti
                             meti"), each =100000))
ggplot(FreeThrows, aes(x = Prosti , fill = lines)) +
  xlab("Koeficient prostih metov") + ylab("Gostota verjetnosti") + geom_density(
      alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$FT), size=1)

# Met za 3
ThreePoints <- data.frame(Trojke =c(GraficnaApriornaBetaSigma2$ThreePT, Posterior =
     GraficnaAposteriornaBeta$ThreePT),
                          lines = rep(c("Apriorni met za 3 točke", "Aposteriorni
                              met za 3 točke"), each =100000))
ggplot(ThreePoints, aes(x = Trojke , fill = lines)) + xlab("Koeficient meta za 3 to
    čke") +
  ylab("Gostota verjetnosti") + geom_density(alpha = 0.5) +
  geom_vline(xintercept = mean(GraficnaAposteriornaBeta$ThreePT), size=1)
```

Slovar strokovnih izrazov

a posteriori distribution aposteriorna porazdelitev, končna porazdelitev dobljena s posodobitvijo apriorne porazdelitve

a priori distribution apriorna porazdelitev

conjugate distributions konjugirana porazdelitev, če je aposteriorna porazdelitev v isti družini porazdelitev verjetnosti kot apriorna porazdelitev, se apriorna in aposteriorna porazdelitev imenujeta konjugirana porazdelitev

conjugate prior konjugacijski prior, apriorna porazdelitev, ki ima isto parametrično obliko porazdelitvene funkcije kot jo ima funkcija največjega verjetja

inference (statistical) sklepanje (v statističnem smislu) **multiple linear regression model** multipli linearni regresijski model

sampling noise naključni vzorčni hrup, zajema vse druge dejavnike, ki vplivajo na vrednost odvisne spremenljivke in jih nismo vključili v model

statistical inference statistično sklepanje

stochastic simulation methods stohastične metode simulacije, s pomočjo naključnih ali kvazinaključnih števil in velikega števila izračunov in ponavljanj omogočajo predvidevanje obnašanja zapletenih matematičnih sistemov

Literatura

[1] A. Ali, A. N. Inglis, E. Prado in B. Wundervald, Bayesian Linear Regression, [ogled 11.11.2019], dostopno na https://brunaw.com/phd/bayes-regression/report.pdf?fbclid=IwAR3qVUBkwEbkfrSs5h9gCMwhPM2niGtL6VNE9dlJM48MlaMSL0VWLB4s68c

[2] Spiegelhalter, D., Rice, K., (2009) Bayesian statistics. Scholarpedia, 4(8):5230.

[3] Koehrsen, W., (2018). Introduction to Bayesian Linear Regression. Towards data science, [ogled 9.11. 2019], dostopno na https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7

[4] P. D. Hoff, A First Course in Bayesian Statistical Methods, Springer-Verlag. New York, 2009; dostopno na https://doi.org/10.1007/978-0-387-92407-6

[5] Linear regression, v Wikipedia: The Free Encyclopedia, [ogled 5.11. 2019], dostopno na https://en.wikipedia.org/wiki/Linear_regression

[6] Bayesian linear regression, v Wikipedia: The Free Encyclopedia, [ogled 5.11. 2019], dostopno na https://en.wikipedia.org/wiki/Bayesian_linear_regression

[7] Clyde, M., Cetinkaya-Rundel M., Rundel, C., Banks, D., Chai, C. in Huang, L. (2020). An Introduction to Bayesian Thinking, [ogled 28.12.2019], dostopno na https://statswithr.github.io/book/

[8] Halls-Moore L. Michael. Bayesian Statistics: A Beginner's Guide, [ogled 28.12.2019], dostopno na https://www.quantstart.com/articles/Bayesian-Statistics-A-Beginners-Guide/

[9] A. Llera in C. F. Beckmann, Estimating an Inverse Gamma distribution, v: arXiv, [2016], [ogled 4.7.2020], dostopno na https://arxiv.org/pdf/1605.01019.pdf?fbclid=IwAR2fwLeVN4TiXD4xoPAQjQ3LC8BB0GvO68mJVRDnQwqKZf835Vtgv62lpFc

[10] A. Toman, Zgledi uporabe statistike na različnih strokovnih področjih Bayesov pristop v statistiki, v: Seminar DMFA, [ogled 2.5.2020], dostopno na https://www.dmfa.si/Predavanja/Dokumenti/67/Toman.pdf?fbclid=IwAR126FCy31vQxn2iWUFmfXVy9c_c0whaOXZZ-DabzZz7i10ZbTnNfkeiP9A

[11] Basketball Stats and History, v: Basketball Reference, [ogled 1.7.2020], dostopno na https://www.basketball-reference.com/?fbclid=IwAR3UXbTVsYlQvyfPDQxzEnM2Kd3AKmn3mkDD8Q5SIyWhSxlvCUfxudobix0

[12] NBA Salaries, v: Hoops Hype, [ogled 1.7.2020], dostopno na https://hoopshype.com/salaries/?fbclid=IwAR3qIpWbv0yeGWfNWUVnHHSIi6QIu7AycQtDZ7m6g_VBitLtquMiTm0kaEs

[13] E. Inzaugarat, Linear and Bayesian modeling in R: Predicting movie popularity, v: Towards Data Science, [ogled 6.7.2020], dostopno na https://towardsdatascience.com/linear-and-bayesian-modelling-in-r-predicting-movie-popularity-6c8ef0a44184