

Bayesovska statistika

Uz R programski jezik

Ivan Papić

1/1/24

Sadržaj

Predgovor	3
1 Apriorna i posteriorna distribucija, vjerodostojnost	4
2 Apriorna gustoća	4
3 Posteriorna gustoća	5
4 Vjerodostojnost	5
5 Bayesova formula je temelj Bayesovske paradigme:	6
6 Oznaka $\propto$	6
7 Uočimo:	7
8 Apriorna prediktivna distribucija	7
9 Posteriorna prediktivna distribucija	8
9.1 Zadaci	8
11 Rješenje	9
13 Rješenje	10
15 Rješenje	11
17 Rješenje	12
19 Rješenje	13
20 Bayesov rizik i Bayesov procjenitelj	14
20.1 Bayesov rizik	14
21 Funkcija gubitka	14
22 Funkcija rizika	15

23 Minimax procjenitelj	15
24 Bayesov rizik	16
24.1 Bayesov procjenitelj	16
25 Bayesov procjenitelj	16
27 Očekivanje posteriorne distribucije kao procjena - kvadratni rizik	17
28 Medijan posteriorne distribucije kao procjena - apsolutni rizik	18
29 Bayesov procjenitelj uz konstantnu funkciju rizika - minimax procjenitelj	18
29.1 Zadaci	19
34 O izboru apriorne distribucije	23
34.1 Neinformativne apriorne distribucije	23
35 Fisherova informacija	24
36 Jeffreyjeva apriorna gustoća	24
36.1 Informativne apriorne distribucije	25
37 Konjugirana apriorna gustoća	25
38 Eksponecijalna familija distribucija	26
38.1 Konjugirani modeli	27
38.1.1 Gama-Poissonov model	27
38.1.2 Beta-Binomni model	28
38.1.3 Gama-Eksponecijalni model	29
38.1.4 Normalni model	30
38.2 Zadaci	32
40 Rješenje	33
42 Rješenje	34
44 Rješenje	35
46 Rješenje	38
48 Rješenje	42
50 Rješenje	48
52 Rješenje	53

53 Vjerodostojna područja (intervali)	56
54 α- vjerodostojno područje	56
55 područje najveće posteriorne gustoće (HPD područje)	56
56 Primjer	58
57 Zadaci	60
59 Rješenje	62
60 Monte Carlo metode	66
61 Monte Carlo aproksimacija	66
62 Primjer	67
62.1 Posteriorno zaključivanje za proizvoljne funkcije parametra	68
63 Primjer	68
63.1 Monte Carlo Markovljevi lanci	69
63.1.1 Metropolis-Hastings algoritam	69
64 Metropolis Hastings algoritam (Bayesovska verzija)	69
65 Konvergencija Metropolis Hastings algoritma	70
66 Pravilo sagorijevanja (“burn in period”)	71
67 Primjer - uzorkovanje standardne normalne distribucije	73
68 Primjer - normalni konjugirani model (poznata varijanca)	76
68.0.1 Gibbsovo uzorkovanje	78
69 Gibbsovo uzorkovanje (Bayesovska verzija) - algoritam	78
70 Ergodski teorem	79
71 Primjer - normalni konjugirani model (nepoznato očekivanje, nepoznata varijanca)	81
71.0.1 O brzini konvergencije MCMC metoda	87
72 efektivna veličina uzorka	89
72.0.1 (Ostale) MCMC metode i njihova implementacija	90
72.1 Zadaci	92
74 Rješenje	94

77 Rješenje	99
79 Rješenje	104
81 Rješenje	109
83 Rješenje	114
85 Rješenje	122
87 Rješenje	132
89 Rješenje	143
90 Bayesovski hijerarhijski modeli	149
91 Izmjenljivost (eng. Exchangeability)	149
92 j.s.u. i izmjenljivost	149
93 de Finetti	150
94 Bayesovski hijerarhijski model	151
94.1 Hijerarhijski normalni model (homogene varijance)	151
95 Bayesovski normalni hijerarhijski model	152
95.1 Hijerarhijski normalni model (različite varijance)	153
96 Bayesovski normalni hijerarhijski model (različite varijance)	153
97 Primjer - hijerarhijski normalni model (poznata varijanca)	156
99 Rješenje	165
101 Rješenje	171
102 Bayesov faktor, testiranje hipoteza i selekcija modela	177
103 Interpretacija Bayesovog faktora	177
104 Primjer	178
105 p-vrijednost	179
106 Regresija	180
106.1 Linearna regresija	180
106.1.1 O selekciji modela u linearnoj regresiji	182

106.2Poissonova regresija	183
107Poissonova regresija	183
108Pretpostavke u Poissonovoj regresiji	184
108.1Logistička regresija	184
109Logistička regresija	185
109.0.1 Klasifikacija u logističkom modelu	185
109.1Zadaci	186
111Rješenje	188
113Rješenje	195
114Vjerojatnosne distribucije	200
114.1Beta distribucija	200
115Beta distribucija	200
115.1Gama distribucija	201
116Gama distribucija	201
116.1Inverzna gama distribucija	201
117Inverzna gama distribucija	202
117.1Beta-binomna distribucija	202
118Beta-binomna distribucija	203
118.1Negativna-binomna distribucija	203
119Negativna-binomna distribucija	204
119.1Lomax distribucija	204
120Lomax distribucija	205
Reference	206

Predgovor

Ovaj materijal zamišljen je kao popratni sadržaj kolegiju Bayesovska statistika te služi kao teoretski i praktični materijal korištenjem R-a.

Korišteni paketi:

```
library("MASS")  
library("nimble")  
library("scatterplot3d")  
library("extraDistr")  
library("HDIInterval")  
library("Flury")  
library("fishmethods")  
library("coda")  
library("rstan")  
library("rstanarm")  
library("bayesplot")  
library("bayesrules")  
library("tidyverse")
```

Ukoliko želite reproducirati rezultate u primjerima i zadacima koji primijenjuju simulacije koristite

```
set.seed(934)
```

Većina materijala kreirana je na temelju strukture, primjera i ideja iz odlično napisanih knjiga Hoff (2009) i Johnson, Ott, and Dogucu (2022) (prva sadrži puno teoretskih detalja, dok je druga fokusirana na intuitivan i praktičan dio u R-u).

1 Apriorna i posteriorna distribucija, vjerodostojnost

Neka je $\mathbb{X} = (X_1, \dots, X_n)$ slučajni uzorak iz parametarskog statističkog modela $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$. Označimo njegovu realizaciju (podatke) s $\mathbf{x} = (x_1, \dots, x_n)$.

Obično pretpostavljamo da se radi o jednostavnom slučajnom uzorku pa tada za $\mathbb{X} \sim f(\mathbf{x}|\theta)$ imamo:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta),$$

pri čemu je $X_1 \sim f(x|\theta)$.

U Bayesovskoj paradigmi pretpostavljamo da parametar θ nije fiksna veličina, već slučajna varijabla s funkcijom gustoće $f(\theta)$.

Note

2 Apriorna gustoća

Gustoća parametra θ označava se $f(\theta)$ i naziva se **apriorna gustoća** (eng. **prior**).

3 Posteriorna gustoća

Gustoća parametra θ uvjetno na $\mathbb{X} = \mathbf{x}$ označava se $f(\theta|\mathbf{x})$ i naziva se **posteriorna gustoća** (eng. **posterior**).

4 Vjerodostojnost

Funkciju $\theta \mapsto f(\mathbf{x}|\theta)$, pri čemu $\mathbf{x} = (x_1, \dots, x_n)$ predstavlja podatke, zovemo **vjerodostojnost** (eng. **likelihood**).

Cilj nam je ažurirati početno znanje o parametru (koje karakteriziramo kroz apriornu distribuciju $f(\theta)$) temeljem podataka i doći do novog znanja o parametru (posteriorne distribucije $f(\theta|\mathbf{x})$).

Bayesova formula daje sljedeći odnos između definiranih funkcija:



Tip

5 Bayesova formula je temelj Bayesovske paradigme:

$$f(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) \cdot f(\theta)}{f(\mathbf{x})} = \frac{f(\mathbf{x}|\theta) \cdot f(\theta)}{\int_{\theta} f(\mathbf{x}|\theta) f(\theta) d\theta}. \quad (5.1)$$



Note

6 Oznaka $\propto$

Za funkcije f i g pišemo $f \propto g$ ukoliko postoji $c \in \mathbb{R}$ t.d. $f(x) = c \cdot g(x), \forall x \in \mathbb{R}$.

U tom kontekstu, možemo uočiti da formulu (5.1) možemo zapisati u obliku

$$f(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta) \cdot f(\theta)$$

jer je nazivnik konstanta. Bayesovu paradigmu riječima često pojednostavljeno pišemo:



Tip

7 Uočimo:

posteriorna gustoća $\propto$ vjerodostojnost $\cdot$ apriorna gustoća
(eng. **posterior $\propto$ likelihood $\cdot$ prior**).

U svrhu predikcije sljedeće realizacije x_{n+1} apriorno i posteriorno, odnosno bez uzimanja/uzimanjem u obzir postojećih podataka $(x_1, \dots, x_n)$ definiramo dva koncepta:



Note

8 Apriorna prediktivna distribucija

Funkciju f definiranu izrazom

$$f(x) = \int_{\Theta} f(x|\theta)f(\theta)d\theta$$

zovemo **apriorna prediktivna gustoća**.

9 Posteriorna prediktivna distribucija

Funkciju f definiranu izrazom

$$f(x_{n+1}|x_1, \dots, x_n) = \int_{\Theta} f(x_{n+1}|\theta) f(\theta|x_1, \dots, x_n) d\theta$$

zovemo **posteriorna prediktivna gustoća**.

9.1 Zadaci

 Danger

10

Zadatak 10.1. Neka je $\mathcal{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{P}(\lambda)$, $\lambda > 0$ distribucije. Za sljedeće apriorne distribucije odredite pripadne posteriorne distribucije.

- a) $f(\lambda) = I_{(0, \infty)}(\lambda)$
- b) $f(\lambda|\beta) = \beta e^{-\beta\lambda} I_{(0, \infty)}(\lambda)$
- c) $f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} I_{(0, \infty)}(\lambda)$

11 Rješenje

- a) $\lambda|\mathbb{X} \sim \Gamma\left(\sum_{i=1}^n x_i + 1, n\right)$
- b) $\lambda|\mathbb{X} \sim \Gamma\left(\sum_{i=1}^n x_i + 1, n + \beta\right)$
- c) $\lambda|\mathbb{X} \sim \Gamma\left(\sum_{i=1}^n x_i + \alpha, n + \beta\right)$

12

Zadatak 12.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{B}(m, \theta)$ distribucije, gdje je $m \in \mathbb{N}$ poznata konstanta, a $\theta \in \langle 0, 1 \rangle$ parametar. Za sljedeće apriorne distribucije odredite pripadne posteriorne distribucije.

- a) $f(\theta) = I_{\langle 0, 1 \rangle}(\theta)$
- b) $f(\theta) = I_{\{\theta=1/2\}}$
- c) $f(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{\langle 0, 1 \rangle}(\theta)$

13 Rješenje

- a) $\theta|\mathbb{X} \sim \text{Beta}\left(\sum_{i=1}^n x_i + 1, mn - \sum_{i=1}^n x_i + 1\right)$
- b) $\theta|\mathbb{X} \sim \theta$, tj. $\theta = 1/2$, g.s.
- c) $\theta|\mathbb{X} \sim \text{Beta}\left(\sum_{i=1}^n x_i + \alpha, mn - \sum_{i=1}^n x_i + \beta\right)$

14

Zadatak 14.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{Ber}(\theta)$ distribucije, $\theta \in \langle 0, 1 \rangle$ parametar. Za sljedeće apriorne distribucije odredite pripadne posteriorne distribucije.

- a) $f(\theta) = I_{\langle 0, 1 \rangle}(\theta)$
- b) $f(\theta) = I_{\{\theta=1/2\}}$
- c) $f(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{\langle 0, 1 \rangle}(\theta)$

15 Rješenje

- a) $\theta|\mathbb{X} \sim \text{Beta}\left(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1\right)$
- b) $\theta|\mathbb{X} \sim \theta$, tj. $\theta = 1/2$, g.s.
- c) $\theta|\mathbb{X} \sim \text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right)$

16

Zadatak 16.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{E}(\lambda)$ distribucije, $\lambda > 0$ parametar. Za sljedeće apriorne distribucije odredite pripadne posteriorne distribucije.

- a) $f(\lambda) = I_{(0,\infty)}(\lambda)$
- b) $f(\lambda|\beta) = \beta e^{-\beta\lambda} I_{(0,\infty)}(\lambda)$
- c) $f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} I_{(0,\infty)}(\lambda)$

 Danger

17 Rješenje

- a) $\lambda | \mathbb{X} \sim \Gamma \left(n + 1, \sum_{i=1}^n x_i \right)$
- b) $\lambda | \mathbb{X} \sim \Gamma \left(n + 1, \beta + \sum_{i=1}^n x_i \right)$
- c) $\lambda | \mathbb{X} \sim \Gamma \left(n + \alpha, \sum_{i=1}^n x_i + \beta \right)$

 Danger

18

Zadatak 18.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{N}(\mu, \sigma^2)$ distribucije, gdje je $\mu \in \mathbb{R}$ parametar, a $\sigma^2 > 0$ poznata veličina. Uz apriornu distribuciju $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $\mu_0 \in \mathbb{R}$, $\sigma_0^2 > 0$ odredite posteriornu distribuciju.

19 Rješenje

$\mu | \{\mathbb{X} = (x_1, x_2, \dots, x_n)\} \sim \mathcal{N}(\mu_n, \sigma_n^2)$, gdje je

$$\mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2} \bar{x}_n}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}.$$

20 Bayesov rizik i Bayesov procjenitelj

20.1 Bayesov rizik

U ovom poglavlju ilustriramo kako formalno uspoređivati frekvencionističke i Bayesovske procjenitelje, odnosno koje statistike posteriorne distribucije ima smisla koristiti kao točkovne procjene.

 Note

21 Funkcija gubitka

Neka je $T = t(X_1, \dots, X_n)$ procjenitelj parametra θ i $\mathbb{t} = t(x_1, \dots, x_n)$ njegova realizacija. **Funkcija gubitka** je bilo koja realna funkcija $L : \Theta \times \Theta \mapsto \mathbb{R}$ t.d. je - $L(t, \theta) \geq 0, \forall t, \forall \theta \in \Theta$ - $L(t, \theta) = 0 \iff \mathbb{t} = \theta$.

Tipične funkcije gubitka su:

- kvadratni gubitak: $L(t, \theta) = ||t - \theta||^2$
- apsolutna udaljenost: $L(t, \theta) = ||t - \theta||$

22 Funkcija rizika

Funkcija rizika je očekivani gubitak, tj. definiran je sljedećim izrazom

$$R_T(\theta) = E_\theta [L(t, \theta)]$$

Ideja je težiti procjeniteljima koji imaju manji rizik. Pri tome, uz kvadratni rizik kažemo da se radi o MSE (eng. mean square error).

U konzervativnom pristupu tražimo procjenitelj koji ima najmanji maksimalni rizik, a definiramo ga na sljedeći način.

23 Minimax procjenitelj

Procjenitelj T^* je **minimax procjenitelj** ako minimizira maksimalan rizik, tj.

$$\sup_{\theta \in \Theta} R_{T^*}(\theta) \leq \sup_{\theta \in \Theta} R_T(\theta), \quad \forall T.$$

Kako bismo definirali procjenitelj u Bayesovskoj paradigmi, treba definirati mjeru koja kvantificira rizik po svim mogućim realizacijama parametra kao slučajne varijable.

24 Bayesov rizik

Bayesov rizik procjenitelja T , relativno u odnosu na funkciju rizika $R_T(\theta)$ i apriornu gustoću $f(\theta)$, je očekivani rizik u odnosu na $f(\theta)$, tj.:

$$A_T = \int_{\Theta} R_T(\theta) f(\theta) d\theta.$$

Iz definicije Bayesovog rizika A_T slijedi da je on zapravo jednak srednjoj vrijednosti rizika po apriornoj gustoći, tj. $E_f[R_T(\theta)]$.

24.1 Bayesov procjenitelj

Ideja je pronaći onaj procjenitelj koji ima najmanji Bayesov rizik.

25 Bayesov procjenitelj

Bayesov procjenitelj T , relativno u odnosu na funkciju rizika $R_T(\theta)$ i apriornu gustoću $f(\theta)$, je procjenitelj s najmanjim Bayesovim rizikom, tj. za svaki drugi procjenitelj T vrijedi: očekivani rizik u odnosu na $f(\theta)$, tj.:

$$\int_{\Theta} R_{T^*}(\theta) f(\theta) d\theta \leq \int_{\Theta} R_T(\theta) f(\theta) d\theta, \quad \forall T.$$

Sljedeći rezultat nam daje ideju kako odrediti Bayesov procjenitelj.

! Important

26

Teorem 26.1. *Ako procjenitelj $T = t(\mathbb{X})$ za svaki $\mathbb{X} = \mathbf{x}$ minimizira očekivani posteriorni gubitak*

$$\int_{\Theta} L(t(\mathbf{x}), \theta) f(\theta|\mathbf{x}) d\theta$$

tada T minimizira i Bayesov rizik A_T .

Sljedeća dva teorema identificiraju kada ima smisla koristiti očekivanje i medijan posteriorne distribucije kao procjene.

! Important

27 Očekivanje posteriorne distribucije kao procjena - kvadratni rizik

Teorem 27.1. *Bayesov procjenitelj T za parametar θ , relativno u odnosu na kvadratni rizik ($L(a, b) = \|a - b\|^2$) i apriornu gustoću $f(\theta)$, je $T = E[\theta|\mathbb{X}]$*

! Important

28 Medijan posteriorne distribucije kao procjena - apsolutni rizik

Teorem 28.1. *Bayesov procjenitelj T za parametar θ , relativno u odnosu na apsolutni rizik ($L(a, b) = ||a - b||$) i apriornu gustoću $f(\theta)$, je $T = \text{medijan}[\theta|\mathbb{X}]$*

Sljedeći rezultat nam kaže kada će Bayesov procjenitelj biti ujedno (frekvencionistički) minimax procjenitelj.

! Important

29 Bayesov procjenitelj uz konstantnu funkciju rizika - minimax procjenitelj

Teorem 29.1. *Neka je T^* Bayesov procjenitelj uz danu funkciju rizika $R_T(\theta)$ i neka ima konstantnu funkciju rizika, tj. $R_{T^*}(\theta) = c, \forall \theta \in \Theta$. Tada je T^* minimax procjenitelj u odnosu na tu funkciju rizika.*

29.1 Zadaci



30

Zadatak 30.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{P}(\lambda)$, $\lambda > 0$ distribucije te neka je pripadna apriorna gustoća $f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} I_{(0, \infty)}(\lambda)$, $\alpha > 0, \beta > 0$.

- a) Odredite apriornu prediktivnu distribuciju
- b) Odredite posteriornu prediktivnu distribuciju
- c) Odredite Bayesov procjenitelj i pripadni Bayesov rizik uz funkciju gubitka $L(t, \lambda) = (t - \lambda)^2$.
- d) Odredite Bayesov procjenitelj uz funkciju gubitka $L(t, \lambda) = |t - \lambda|$.

31

Zadatak 31.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{P}(\lambda)$, $\lambda > 0$ distribucije te neka je pripadna apriorna gustoća $f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} I_{(0, \infty)}(\lambda)$, $\alpha > 0, \beta > 0$. Neka je dana kvadratna greška i dva procjenitelja nepoznatog parametra λ :

$$T_1 = \frac{n\bar{X}_n + \alpha}{n + \beta}, \quad T_2 = \bar{X}_n.$$

- Pokažite da se procjenitelj T_1 može zapisati kao težinska sredina prosjeka slučajnog uzorka i apriornog očekivanja. Komentirajte dobiveni rezultat.
- Izračunajte (Bayesove) rizike predloženih procjenitelja te ih usporedite. Komentirajte.

32

Zadatak 32.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{B}(m, \theta)$ distribucije, gdje je $m \in \mathbb{N}$ poznata konstanta, a $\theta \in \langle 0, 1 \rangle$ parametar. Uz zadanu apriornu gustoću $f(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta)$ odredite:

- a) apriornu prediktivnu distribuciju
- b) posteriornu prediktivnu distribuciju
- c) Bayesov procjenitelj i pripadni Bayesov rizik uz funkciju gubitka $L(t, \theta) = (t - \theta)^2$.
- d) Bayesov procjenitelj i pripadni Bayesov rizik uz funkciju gubitka $L(t, \theta) = \frac{(t-\theta)^2}{\theta^\alpha (1-\theta)^\beta}$.

Postoji li bolji procjenitelj prema kriteriju manjeg maksimalnog gubitka? Za koje vrijednosti parametara α i β postoji minimax procjenitelj?

33

Zadatak 33.1. Neka je $\mathbb{X} = (X_1, \dots, X_n)$ j.s.u. iz $\mathcal{B}(m, \theta)$ distribucije, gdje je $m \in \mathbb{N}$ poznata konstanta, a $\theta \in \langle 0, 1 \rangle$ parametar. Uz pripadnu apriornu gustoću $f(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta)$ i kvadratnu grešku dani su sljedeći procjenitelji nepoznatog parametra θ :

$$T_1 = \frac{n\bar{X}_n + \alpha}{nm + \alpha + \beta}, \quad T_2 = \frac{\bar{X}_n}{m}.$$

- Pokažite da se procjenitelj T_1 može zapisati kao težinska sredina T_2 i apriornog očekivanja. Komentirajte dobiveni rezultat.
- Izračunajte (Bayesove) rizike predloženih procjenitelja te ih usporedite. Komentirajte.

34 O izboru apriorne distribucije

Apriornu distribuciju shvaćamo kao subjektivno mišljenje o nepoznatom parametru koje kvantificiramo vjerojatnosno kroz distribuciju.

U tom kontekstu imamo dva pristupa:

- **neinformativne apriorne distribucije:** imaju minimalan utjecaj na posteriornu distribuciju, odnosno donesene zaključke u Bayesovskoj paradigmi
- **informativne apriorne distribucije:** imaju ugrađeno “znanje” o parametru na temelju prethodnih istraživanja, subjektivnog mišljenja ili iskustva.

34.1 Neinformativne apriorne distribucije

Grubo ih dijelimo na sljedeće skupine:

- **Uniformna apriorna distribucija**

Koristimo ukoliko nemamo nikakvo prethodno znanje o parametru (dodjeljujemo svim mogućim vrijednostima parametra jednake “težine”). U slučaju da je prostor parametara beskonačan (prebrojiv ili neprebrojiv) tada se radi o **nepravoj** apriornoj distribuciji jer to ona zapravo i nije (nije moguće normirati gustoću). Unatoč tome, možemo postaviti $f(\theta) \propto c$, $c \in \mathbb{R}$, ali je potrebno provjeriti je li posteriorna distribucija doista distribucija, tj. je li normirana.

- **Jeffreyjeva apriorna distribucija**

Motivirana je problemom reparametrizacije:

Ukoliko postavimo uniformnu apriornu distribuciju $f(\theta)$, hoće li reparametrizirani parametar $\phi = g(\theta)$ i dalje imati uniformnu apriornu distribuciju? Iz teorije vjerojatnosti znamo da uz g bijekciju, ϕ ima funkciju gustoće

$$f(\phi) = f(\theta) \left| \frac{d\theta}{d\phi} \right|$$

što ne mora imati isti oblik kao početna distribucija.

Taj problem rješava upravo Jeffreyjeva apriorna distribucija koja se definira na temelju Fisherove informacije.

35 Fisherova informacija

Za fiksni parametar θ Fisherova informacija definirana je izrazom

$$I(\theta) = E_{\theta} \left[\left(\frac{d}{d\theta} \ln p(x|\theta) \right)^2 \right] = -E_{\theta} \left[\frac{d^2}{d\theta^2} \ln p(x|\theta) \right],$$

gdje je $p(x|\theta)$ pripadna vjerodostojnost.

- Jeffreyjev pristup odabiru apriorne distribucije omogućuje konzistentnost obzirom na reparametrizaciju:

36 Jeffreyjeva apriorna gustoća

Apriornu gustoću za parametar θ definiranu sa $f(\theta) \propto \sqrt{I(\theta)}$, gdje je $I(\theta)$ Fisherova informacija, zovemo **Jeffreyjeva apriorna gustoća**.

Ovaj pristup (uz potrebne modifikacije) funkcionira i u multivarijatnom slučaju.

- **Apriorne distribucije maksimalne entropije**

Apriorne distribucije dobivene maksimiziranjem entropije kao mjere količine informacija u parametru (veća entropija podrazumijeva manje informacija).

- **Referentne apriorne distribucije**

Reflektiraju ideju traženja one apriorne distribucije koja je u srednjem smislu najudaljenija od posteriorne distribucije. Kao mjeru udaljenosti između distribucija koristi tzv. Kullback-Leiblerovu divergenciju. Zanimljivo je da za jednodimenzionalne parametre dobivamo istu

apriornu distribuciju kao i u Jeffreyjevom pristupu, ali je generalno bolji pristup u multivarijantnom slučaju.

36.1 Informativne apriorne distribucije

Grubo ih dijelimo na dvije skupine:

- one koje definiramo na temelju vlastitog subjektivnog mišljenja o parametru
- **konjugirane apriorne distribucije**: analitički pogodne za računske manipulacije

i Note

37 Konjugirana apriorna gustoća

Neka je $\mathcal{P}$ parametarska familija apriornih gustoća, a $\mathcal{F}$ parametarska familija uzoračkih distribucija (vjerodostojnosti). Za familiju $\mathcal{P}$ kažemo da je **konjugirana** u odnosu na familiju $\mathcal{F}$ ako

$$f(\theta|x) \in \mathcal{P}, \quad \forall f(\cdot|\theta) \in \mathcal{F} \text{ i } f(\cdot) \in \mathcal{P}.$$

- Prethodna definicija je vrlo općenita: ukoliko bismo za $\mathcal{P}$ postavili sve distribucije, svaka distribucija bi bila konjugirana apriorna. Međutim nas zanimaju prirodne konjugirane apriorne distribucije (gustoće iz $\mathcal{P}$ i $\mathcal{F}$ imaju istu funkcionalnu formu).
- Ukoliko koristimo konjugiranu apriornu distribuciju, posteriorna distribucija će dolaziti iz iste familije distribucija kao i apriorna.
- Konjugirane apriorne distribucije omogućavaju jednostavan pristup ažuriranju podacima kroz parametre zadane familije distribucija iz koje dolazi apriorna distribucija. Većina elemenata Bayesovskog modela se jednostavno računaju i imaju jasnu interpretaciju.
- Početne vrijednosti parametara apriorne distribucije obično zovemo **hiperparametri**, a definiramo ih na temelju znanja/mišljenja o primjerice očekivanoj vrijednosti i raspršenosti parametra.

U nastavku objašnjavamo vezu između konjugiranih apriornih distribucija i eksponencijalne familije distribucija.

38 Eksponencijalna familija distribucija

Za familiju distribucija $\mathcal{F}$ kažemo da je **eksponencijalna familija** ukoliko se svaki njen član može zapisati u formi

$$f(x|\theta) = h(x)g(\theta)e^{\Phi(\theta)^T \cdot u(x)},$$

gdje su $\Phi(\theta)$ i $u(x)$ vektori jednake dimenzije kao θ , pri čemu $\Phi(\theta)$ zovemo **prirodan parametar familije** $\mathcal{F}$.

U slučaju *j.s.u.* uzorka $(x_1, \dots, x_n)$ iz eksponencijalne familije imamo:

- vjerodostojnost:

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f(x_i|\theta) = \left[\prod_{i=1}^n h(x_i) \right] g(\theta)^n \exp\{\Phi(\theta)^T \sum_{i=1}^n u(x_i)\} \\ &\propto g(\theta)^n \exp\{\Phi^T(\theta)t(\mathbf{x})\}, \end{aligned}$$

gdje je $t(\mathbf{x}) = \sum_{i=1}^n u(x_i)$

- Izaberemo apriornu distribuciju u obliku koji prati funkcionalnu formu vjerodostojnosti

$$f(\theta) \propto g(\theta)^{\alpha} e^{\Phi(\theta)^T \beta},$$

pri čemu su α i β hiperparametri apriorne distribucije.

- Time dobivamo posteriornu distribuciju koja je istog oblika kao apriorna ali s ažuriranim parametrima temeljem podataka kroz zbroj vrijednosti podataka $\left(\sum_{i=1}^n x_i\right)$ i broj podataka (n) :

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)f(\theta) \\ &\propto g(\theta)^{n+\alpha} e^{\Phi(\theta)^T (\beta+t(\mathbf{x}))} \end{aligned}$$

38.1 Konjugirani modeli

Na temelju ideje konjugirane apriorne distribucije izdvajamo najosnovije i najčešće korištene konjugirane modele.

38.1.1 Gama-Poissonov model

Parametar od interesa	$\lambda > 0$ - očekivana vrijednost brojača
Vjerodostojnost	$f(x \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} I_{\mathbb{N}_0}(x)$
Apriorna distribucija	$\lambda \sim \Gamma(\alpha, \beta), \quad \alpha > 0, \beta > 0$
Posteriorna distribucija	hiperparametri $\lambda \{\mathbb{X} = \mathbb{x}\} \sim \Gamma(\alpha + n\bar{x}_n, \beta + n)$
Posteriorno očekivanje	$E[\lambda \mathbb{X} = \mathbb{x}] = \frac{\alpha + n\bar{x}_n}{\beta + n}$
Posteriorna varijanca	$Var[\lambda \mathbb{X} = \mathbb{x}] = \frac{\alpha + n\bar{x}_n}{(\beta + n)^2}$
Posteriorni medijan	nema zatvoreni oblik
Mod	$\frac{\alpha + n - 1}{\beta + n\bar{x}_n}$
Apriorna prediktivna distribucija	$X \sim Neg - Bin(\alpha, \beta)$
Posteriorna prediktivna distribucija	$X_{n+1} X_1 = x_1, \dots, X_n = x_n \sim Neg - Bin(\alpha + n\bar{x}_n, \beta + n)$

Posteriorna distribucija je kompromis između podataka i apriorne distribucije (početnog subjektivnog uvjerenja o parametru):

$$E[\lambda|\mathbb{X} = \mathbb{x}] = \frac{\overbrace{\beta}^{\text{težina}}}{\beta + n} \cdot \underbrace{\frac{\alpha}{\beta}}_{\text{apriorno čekivanje}} + \frac{\overbrace{n}^{\text{težina}}}{\beta + n} \cdot \underbrace{\bar{x}_n}_{\text{aritmetička sredina}}$$

Uočimo:

- ako je $n \gg \beta$:

$$E[\lambda|\mathbb{X} = \mathbb{x}] \approx \bar{x}_n, \quad Var[\lambda|\mathbb{X} = \mathbb{x}] \approx \frac{\bar{x}_n}{n}.$$

- ako je $\beta \gg n$:

$$E[\lambda|\mathbb{X} = \mathbb{x}] \approx \frac{\alpha}{\beta}.$$

- Interpretacija hiperparametara: β - broj apriornih “podataka”, α - suma β apriornih “podataka”.

38.1.2 Beta-Binomni model

Parametar od interesa	$\theta \in \langle 0, 1 \rangle$ - proporcija uspjeha
Vjerodostojnost	$f(x \theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x} I_{\{0,1,\dots,n\}}(x)$
Apriorna distribucija	$\theta \sim \text{Beta}(\alpha, \beta), \quad \alpha > 0, \beta > 0$
Posteriorna distribucija	hiperparametri $\theta \{\mathbb{X} = \mathbb{x}\} \sim \text{Beta}(\alpha + n\bar{x}_n, \beta + mn - n\bar{x}_n)$
Posteriorno očekivanje	$E[\theta \mathbb{X} = \mathbb{x}] = \frac{\alpha + n\bar{x}_n}{\alpha + \beta + mn}$
Posteriorna varijanca	$\text{Var}[\theta \mathbb{X} = \mathbb{x}] = \frac{(\alpha + n\bar{x}_n)(\beta + mn - n\bar{x}_n)}{(\alpha + \beta + mn)^2(\alpha + \beta + mn + 1)}$
Posteriorni medijan	nema zatvoreni oblik
Posteriorni Mod	$\frac{\alpha + n\bar{x}_n - 1}{\alpha + \beta + nm - 2}$
Apriorna prediktivna distribucija	$X \sim \text{Beta} - \text{Bin}(m, \alpha, \beta)$
Posteriorna prediktivna distribucija	$X_{n+1} X_1 = x_1, \dots, X_n = x_n \sim \text{Beta} - \text{Bin}(m, \alpha + n\bar{x}_n, \beta + nm - n\bar{x}_n)$

Posteriorna distribucija je kompromis između podataka i apriorne distribucije (početnog subjektivnog uvjerenja o parametru):

$$E[\theta|\mathbb{X} = \mathbb{x}] = \frac{\overbrace{\alpha + \beta}^{\text{težina}}}{\alpha + \beta + nm} \cdot \underbrace{\frac{\alpha}{\alpha + \beta}}_{\text{apriorno čekivanje}} + \frac{\overbrace{nm}^{\text{težina}}}{\alpha + \beta + nm} \cdot \underbrace{\frac{\bar{x}_n}{m}}_{\text{MLE procjenitelj proporcije}}$$

Uočimo:

- ako je $nm \gg \alpha + \beta$:

$$E[\theta|\mathbb{X} = \mathbb{x}] \approx \frac{\bar{x}_n}{m}, \quad \text{Var}[\theta|\mathbb{X} = \mathbb{x}] \approx \frac{1}{nm} \cdot \frac{\bar{x}_n}{m} \cdot \left(1 - \frac{\bar{x}_n}{m}\right).$$

- ako je $\alpha + \beta \gg nm$:

$$E[\theta|\mathbb{X} = \mathbb{x}] \approx \frac{\alpha}{\alpha + \beta}.$$

- Interpretacija hiperparametara: α - broj apriornih “uspjeha”, β - broj apriornih “neuspjeha”, $\alpha + \beta$ - apriorna dimenzija “uzorka”.

38.1.3 Gama-Eksponencijalni model

Parametar od interesa	$\lambda > 0$ - stopa pojavljivanja događaja
Vjerodostojnost	$f(x \lambda) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$
Apriorna distribucija	$\lambda \sim \Gamma(\alpha, \beta), \quad \alpha > 0, \beta > 0$
Posteriorna distribucija	hiperparametri $\lambda \{\mathcal{X} = \mathbf{x}\} \sim \Gamma(\alpha + n, \beta + n\bar{x}_n)$
Posteriorno očekivanje	$E[\lambda \mathcal{X} = \mathbf{x}] = \frac{\alpha + n}{\beta + n\bar{x}_n}$
Posteriorna varijanca	$Var[\lambda \mathcal{X} = \mathbf{x}] = \frac{\alpha + n}{(\beta + n\bar{x}_n)^2}$
Posteriorni medijan	nema zatvoreni oblik
Posteriorni Mod	$\frac{\alpha + n - 1}{\beta + n\bar{x}_n}$
Apriorna prediktivna distribucija	$X \sim Lomax(\alpha, \beta)$
Posteriorna prediktivna distribucija	$X_{n+1} \mathcal{X}_1 = \mathbf{x}_1, \dots, X_n = \mathbf{x}_n \sim Lomax(\alpha + n, \beta + n\bar{x}_n)$

Posteriorna distribucija je kompromis između podataka i apriorne distribucije (početnog subjektivnog uvjerenja o parametru):

$$E[\lambda|\mathcal{X} = \mathbf{x}] = \underbrace{\frac{\beta}{\beta + n\bar{x}_n}}_{\text{apriorno čekivanje}} \cdot \underbrace{\frac{\alpha}{\beta}}_{\text{apriorno očekivanje}} + \underbrace{\frac{n\bar{x}_n}{\beta + n\bar{x}_n}}_{\text{MLE procjenitelj stope}} \cdot \underbrace{\frac{1}{\bar{x}_n}}_{\text{MLE procjenitelj stope}}$$

Uočimo:

- ako je $n\bar{x}_n \gg \beta$:

$$E[\lambda|\mathcal{X} = \mathbf{x}] \approx \frac{1}{\bar{x}_n}, \quad Var[\lambda|\mathcal{X} = \mathbf{x}] \approx \frac{1}{n\bar{x}_n^2}.$$

- ako je $\beta \gg n\bar{x}_n$:

$$E[\lambda|\mathcal{X} = \mathbf{x}] \approx \frac{\alpha}{\beta}.$$

- Interpretacija hiperparametara: $\alpha - 1$ - broj događaja u eksponencijalnom modelu, β - ukupno vrijeme čekanja do realizacije tih događaja.

38.1.4 Normalni model

38.1.4.1 Nepoznato očekivanje (μ), poznata varijanca (σ^2)

Parametar od interesa	$\mu \in \mathbb{R}$ - očekivana vrijednost u modelu
Vjerodostojnost	$f(x \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Apriorna distribucija	$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2), \quad \mu_0 \in \mathbb{R}, \sigma_0^2 > 0$
Posteriorna distribucija	hiperparametri $\mu \{\mathcal{X} = \mathbf{x}\} \sim \mathcal{N}(\mu_n, \sigma_n^2)$, gdje je $\mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2} \bar{x}_n}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$
Posteriorno očekivanje	$E[\mu \mathcal{X} = \mathbf{x}] = \mu_n$
Posteriorna varijanca	$Var[\mu \mathcal{X} = \mathbf{x}] = \sigma_n^2$
Posteriorni medijan	μ_n
Posteriorni Mod	μ_n
Apriorna prediktivna distribucija	$X \sim \mathcal{N}(\mu_0, \sigma_0^2 + \sigma^2)$
Posteriorna prediktivna distribucija	$X_{n+1} \mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_n = \mathbf{x}_n \sim \mathcal{N}(\mu_n, \sigma_n^2 + \sigma^2)$

Obično se u Bayesovskom normalnom modelu govori o **preciznosti** (recipročna vrijednost varijance):

- $\tau^2 = 1/\sigma^2$: uzoračka preciznost (koliko su podaci “blizu” parametra μ)
- $\tau_0^2 = 1/\sigma_0^2$: apriorna preciznost
- $\tau_n^2 = 1/\sigma_n^2$: posteriorna preciznost.
- Uočimo da je preciznost aditivna: $\tau_n^2 = \tau_0^2 + n\tau^2$ Posteriorna distribucija je kompromis između podataka i apriorne distribucije (početnog subjektivnog uvjerenja o parametru):

$$E[\mu|\mathcal{X} = \mathbf{x}] = \mu_n = \frac{\overbrace{\tau_0^2}^{\text{težina}}}{\tau_0^2 + n\tau^2} \cdot \underbrace{\mu_0}_{\text{apriorno čekivanje}} + \frac{\overbrace{n\tau^2}^{\text{težina}}}{\tau_0^2 + n\tau^2} \cdot \underbrace{\bar{x}_n}_{\text{aritmetička sredina podataka}}$$

Uočimo:

- ako je $n \gg \tau_0^2/\tau^2$:

$$E[\mu|\mathbb{X} = \mathbb{x}] \approx \bar{x}_n, \quad Var[\mu|\mathbb{X} = \mathbb{x}] \approx \frac{\sigma^2}{n}.$$

- ako je $\tau_0^2/\tau^2 \gg n$:

$$E[\mu|\mathbb{X} = \mathbb{x}] \approx \mu_0.$$

- Interpretacija hiperparametara: μ_0 - očekivana apriorna vrijednost, $1/\sigma_0^2$ - apriorna preciznost.

38.1.4.2 Nepoznata varijanca (σ^2), poznato očekivanje (μ)

Parametar od interesa	$\sigma^2 > 0$ - varijanca u modelu
Vjerodostojnost	$f(x \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Apriorna distribucija	$\sigma^2 \sim InvGama(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2), \quad \nu_0 > 0, \sigma_0^2 > 0$
Posteriorna distribucija	hiperparametri $\sigma^2 \{\mathbb{X} = \mathbb{x}\} \sim InvGama\left(\frac{n+\nu_0}{2}, \frac{\nu_0\sigma_0^2 + \sum_{i=1}^n (x_i - \mu)^2}{2}\right)$

Interpretacija hiperparametara:

- ν_0 - veličina apriornog “uzorka”
- σ_0^2 - uzoračka varijanca

38.1.4.3 Nepoznato očekivanje (μ), nepoznata varijanca (σ^2)

Parametar od interesa	$(\mu, \sigma^2) \in \mathbb{R} \times \langle 0, +\infty \rangle$ - očekivanje i varijanca u modelu
Vjerodostojnost	$f(x \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Apriorna distribucija	$\sigma^2 \sim$ $InvGama(\frac{\nu_0}{2}, \frac{\nu_0}{2} \sigma_0^2), \quad \nu_0 >$ $0, \sigma_0^2 > 0$ hiperparametri $\mu \sigma^2 \sim$ $\mathcal{N}(\mu_0, \frac{\sigma^2}{k_0}), \quad \mu_0 \in$ $\mathbb{R}, k_0 \in \mathbb{N}$ hiperparametri
Posteriorna distribucija	$\sigma^2 \{\mathcal{X} = \mathcal{x}\} \sim$ $InvGama(\frac{\nu_n}{2}, \frac{\nu_n}{2} \sigma_n^2),$ gdje je $\nu_n =$ $\nu_0 + n, k_n = k_0 + n,$ $\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n -$ $1)s^2 + \frac{k_0 n}{k_n} (\bar{x}_n - \mu_0)^2$ $\mu \{\sigma^2, \mathcal{X} = \mathcal{x}\} \sim$ $\mathcal{N}(\mu_n, \frac{\sigma^2}{k_n}),$ gdje je $\mu_n = \frac{k_0 \mu_0 + n \bar{x}_n}{k_n}.$

Interpretacija hiperparametara:

- ν_0 - veličina apriornog “uzorka” za varijancu
- σ_0^2 - uzoračka varijanca
- k_0 - veličina apriornog “uzorka” za očekivanje
- μ_0 - apriorna “uzoračka srednja vrijednost k_0 ” podataka” uz varijancu σ^2 .

38.2 Zadaci

 Danger

39

Zadatak 39.1. Korištenjem veze između eksponencijalne familije i konjugiranih apriornih distribucija odredite konjugiranu apriornu distribuciju (i pripadnu posteriornu distribuciju) uz danu vjerodostojnost

$$f(x|k, \theta) = \frac{k}{\theta} x^{k-1} e^{-x^k/\theta} I_{(0, \infty)}(x), \quad k > 0, \theta > 0$$

i pretpostavku da je k zadana konstanta, a θ nepoznati parametar.

 Danger

40 Rješenje

$$\theta \sim \text{InvGama}(\alpha, \beta), \quad \theta | \{\mathbb{X} = \mathbb{x}\} \sim \text{InvGamma} \left(\alpha + n, \beta + \sum_{i=1}^n x_i^k \right)$$

 Danger

41

Zadatak 41.1. Odredite neinformativnu apriornu distribuciju koja osigurava konzistentnost s obzirom na reparametrizaciju uz danu vjerodostojnost

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \binom{m}{x_i} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{m-\sum_{i=1}^n x_i}, \quad m \in \mathbb{N}, \theta \in \langle 0, 1 \rangle, \mathbf{x} = (x_1, x_2, \dots, x_n).$$

 Danger

42 Rješenje

$f(\theta) \propto \theta^{-1/2} (1-\theta)^{-1/2} I_{\langle 0,1 \rangle}(\theta)$, tj. Jeffreyjeva apriorna distribucija je $\theta \sim \text{Beta}(1/2, 1/2)$.

 Danger

43

Zadatak 43.1. Odredite neinformativnu apriornu distribuciju koja osigurava konzistentnost s obzirom na reparametrizaciju uz danu vjerodostojnost

$$f(\mathbf{x}|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\lambda}, \quad \lambda > 0, \quad \mathbf{x} = (x_1, x_2, \dots, x_n).$$

 Danger

44 Rješenje

$f(\lambda) \propto \lambda^{-1/2} I_{(0,\infty)}(\lambda)$, što je Jeffreyjeva (neprava) apriorna distribucija.

45

Zadatak 45.1. Prikupljeni su podaci o međuvremenima (mjereni u satima) između potresa u jednom gradu i zabilježeno je 100 potresa u 63.09 sati. Seizmolog smatra da je (apriorna) očekivana stopa pojavljivanja potresa $4/3$ po satu, a pripadna varijabilnost (varijanca) $4/9$.

- Koji parametar nam je od interesa? Definirajte statistički model, odnosno pripadnu vjerodostojnost.
- Definirajte prirodnu konjugiranu apriornu distribuciju i pripadnu posteriornu distribuciju. Grafički prikažite.
- Koju (posteriornu) očekivanu stopu pojavljivanja potresa definira seizmolog nakon što analizira prikupljene podatke? Usporedite s apriornom stopom.
- Koliko iznosi (posteriorna) vjerojatnost da je stopa pojavljivanja potresa barem 2 po satu nakon što seizmolog analizira podatke, a koliko temeljem apriornog subjektivnog mišljenja? Komentirajte.

 Danger

46 Rješenje

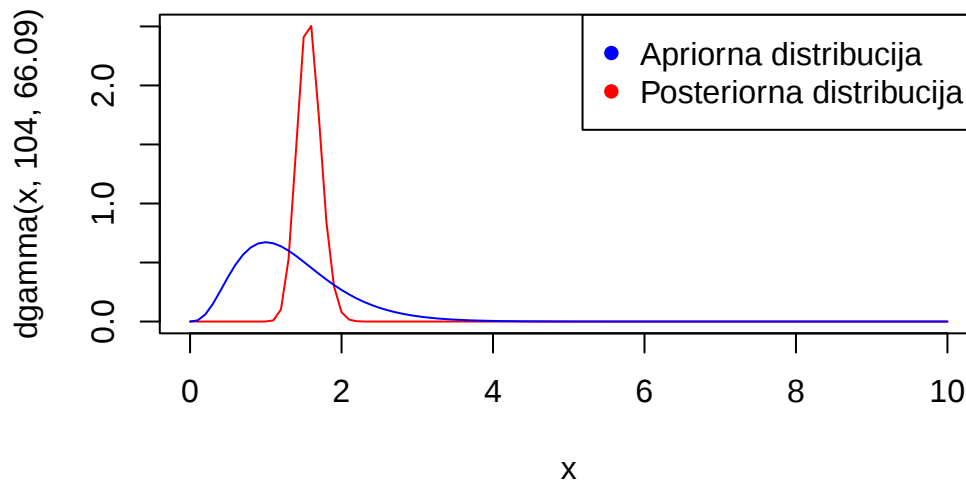
a) Zanima nas parametar $\lambda > 0$ uz vjerodostojnost

$$f(\mathbf{x}|\lambda) = \lambda^{100} e^{-\lambda \sum_{i=1}^{100} x_i}, \quad \lambda > 0, \quad \mathbf{x} = (x_1, \dots, x_{100}),$$

a koji predstavlja stopu pojavljivanja potresa po satu.

- b)
- *Apriorna distribucija*: $\lambda \sim \Gamma(4, 3)$
 - *Posteriorna distribucija*: $\lambda|\{\mathbb{X} = \mathbf{x}\} \sim \Gamma(4 + 100, 3 + 63.09) = \Gamma(104, 66.09)$
 - *Grafički prikaz*:

```
curve(dgamma(x, 104, 66.09), 0, 10, ylim = c(0, 2.5), col = 'red')  
curve(dgamma(x, 4, 3), add=TRUE, col = 'blue')  
legend("topright", legend = c("Apriorna distribucija", "Posteriorna distribucija"),
```



c) Seizmolog sada očekuje stopu pojavljivanja potresa po satu 1.57 uz standardnu devijaciju 0.154, dok je očekivana apriorna stopa bila 1.33 uz devijaciju 0.67.

d)

$$P(\lambda > 2 | \mathbb{X} = \mathbb{x}) \approx 0.00495458, \quad P(\lambda > 2) \approx 0.1512039$$

```
1-pgamma(2,104,66.09)
```

```
[1] 0.00495458
```

```
1-pgamma(2,4,3)
```

```
[1] 0.1512039
```

Uočavamo da se vjerojatnost pojavljivanja barem 2 potresa po satu znatno smanjila nakon što smo uzeli u obzir podatke.

47

Zadatak 47.1. Pretpostavimo da promatramo finale u streličarstvu u kojem sudjeluju dva ista natjecatelja u tri različite godine, pri čemu oba natjecatelja gađaju metu ukupno 60 puta, svaki puta nezavisno i s jednakom vjerojatnošću pogotka. Prikupili smo sljedeće podatke:

Natjecatelj	Prva godina	Druga godina	Treća godina
Natjecatelj 1	53	56	49
Natjecatelj 2	50	57	54

Na temelju dugogodišnjeg praćenja ovih natjecatelja očekujemo da prvi natjecatelj pogađa metu 80% puta, dok drugi 85% puta, pri čemu smatramo da je varijabilnost pogodaka (standardna devijacija) za oba natjecatelja 1%.

- Koji parametri su nam od interesa? Definirajte vjerodostojnost Bayesovog modela.
- Odredite konjugirane apriorne i pripadne posteriorne distribucije. Prikažite grafički dobivene distribucije i komentirajte.
- Koliko iznosi apriorna vjerojatnost da prvi (odn. drugi) natjecatelj ima vjerojatnost pogotka mete između 75% i 90%?
- Koliko iznosi posteriorna vjerojatnost da prvi (odn. drugi) natjecatelj ima vjerojatnost pogotka mete između 75% i 90%?
- Koliko iznosi posteriorna vjerojatnost da prvi natjecatelj ima veću vjerojatnost pogoditi metu nego drugi natjecatelj?
- Što je vjerojatnije, da natjecatelj 1 ili natjecatelj 2 na idućem zajedničkom natjecanju ostvari barem 55 pogodaka? Grafički prikažite.

 Danger

48 Rješenje

a) Zanimaju nas parametri

- $\theta_1 \in \langle 0, 1 \rangle$ vjerojatnost da natjecatelj 1 pogodi metu
- $\theta_2 \in \langle 0, 1 \rangle$ vjerojatnost da natjecatelj 2 pogodi metu Vjerodostojnosti:

$$f(\mathbf{x}|\theta_1) = \prod_{i=1}^3 \binom{60}{x_i} \theta_1^{x_i} (1 - \theta_1)^{60-x_i}, \quad f(\mathbf{y}|\theta_2) = \prod_{i=1}^3 \binom{60}{y_i} \theta_2^{y_i} (1 - \theta_2)^{60-y_i},$$

gdje je

- $\mathbf{x} = (x_1, x_2, x_3) = (53, 56, 49)$, $\mathbf{y} = (y_1, y_2, y_3) = (50, 57, 54)$.

b) *Apriorne distribucije:*

- $\theta_1 \sim \text{Beta}(12, 3)$
- $\theta_2 \sim \text{Beta}(9.9875, 1.7625)$

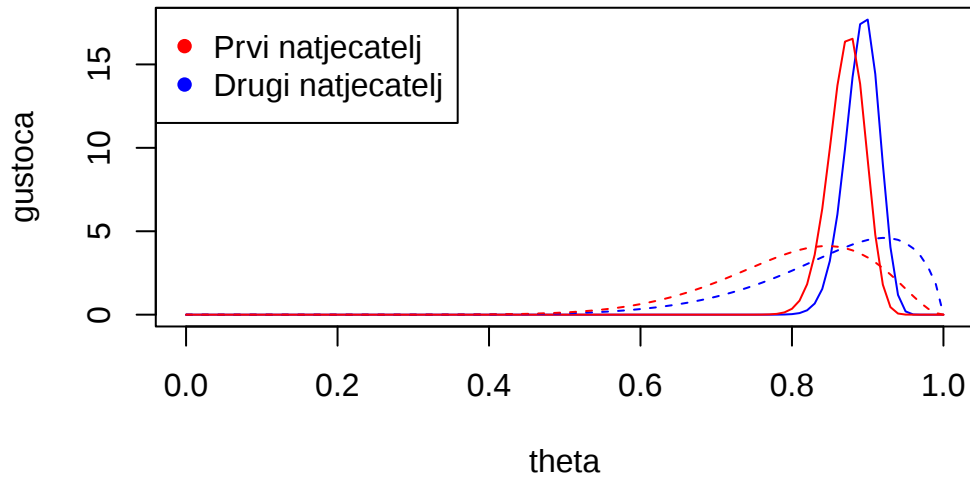
Posteriorne distribucije:

- $\theta_1|\mathbf{x} \sim \text{Beta}\left(\sum_{i=1}^n x_i + 12, 60 \cdot 3 - \sum_{i=1}^n x_i + 3\right) = \text{Beta}(170, 25)$
- $\theta_2|\mathbf{y} \sim \text{Beta}\left(\sum_{i=1}^n y_i + 9.9875, 60 \cdot 3 - \sum_{i=1}^n y_i + 1.7625\right) = \text{Beta}(170.9875, 20.7625)$

Grafički prikaz:

```
theta<-seq(from=0, to=1, by=0.01)
plot(theta, dbeta(theta,170.9875, 20.7625), type="l", col="blue", ylab="gustoća", m
lines(theta, dbeta(theta,170,25), type="l", col="red")
lines(theta, dbeta(theta,9.9875, 1.7625), type="l", lty="dashed", col="blue")
lines(theta, dbeta(theta,12,3), type="l", lty="dashed", col="red")
legend("topleft", legend = c("Prvi natjecatelj", "Drugi natjecatelj"), col = c("red
```

Apriorna gustoca (iscrtkano) i Posteriorna gustoca (puna lii)



c)

$$P(\theta_1 \in [0.75, 0.9]) \approx 0.5605124, \quad P(\theta_2 \in [0.75, 0.9]) \approx 0.4709752$$

```
pbeta(0.90,12,3)-pbeta(0.75,12,3)
```

```
[1] 0.5605124
```

```
pbeta(0.90,9.9875, 1.7625)-pbeta(0.75,9.9875, 1.7625)
```

```
[1] 0.4709752
```

d)

$$P(\theta_1 | \{\mathbb{X} = \mathbb{x}\} \in [0.75, 0.9]) \approx 0.886503, \quad P(\theta_2 | \mathbb{Y} = y \in [0.75, 0.9]) \approx 0.6239969$$

```
pbeta(0.90,170,25)-pbeta(0.75,170,25)
```

```
[1] 0.886503
```

```
pbeta(0.90,170.9875, 20.7625)-pbeta(0.75,170.9875, 20.7625)
```

```
[1] 0.6239969
```

e) Izračunajmo

$$P(\theta_1|\{\mathbb{X} = x\} > \theta_2|\{\mathbb{Y} = y\})$$

numerički, koristeći tzv. Monte Carlo metodu. Simulirajmo 1000 realizacija slučajnih varijabli $\theta_1|\{\mathbb{X} = x\}$ i $\theta_2|\{\mathbb{Y} = y\}$ te izračunajmo relativnu frekvenciju događaja $\theta_1|\{\mathbb{X} = x\} > \theta_2|\{\mathbb{Y} = y\}$

```
theta1 <- rbeta(1000, 170,25)
theta2 <- rbeta(1000, 170.9875, 20.7625)
mean(theta1>theta2)
```

```
[1] 0.253
```

f) Znamo da je u Beta-binomnom modelu prediktivna posteriorna distribucija Beta-binomna distribucija, tj.

$$X_4|\{\mathbb{X} = x\} \sim BetaBin(60, 170, 25) \Rightarrow P(X_4 > 55|\{\mathbb{X} = x\}) \approx 0.136$$

$$Y_4|\{\mathbb{Y} = y\} \sim BetaBin(60, 170.9875, 20.7625) \Rightarrow P(Y_4 > 55|\{\mathbb{Y} = y\}) \approx 0.246$$

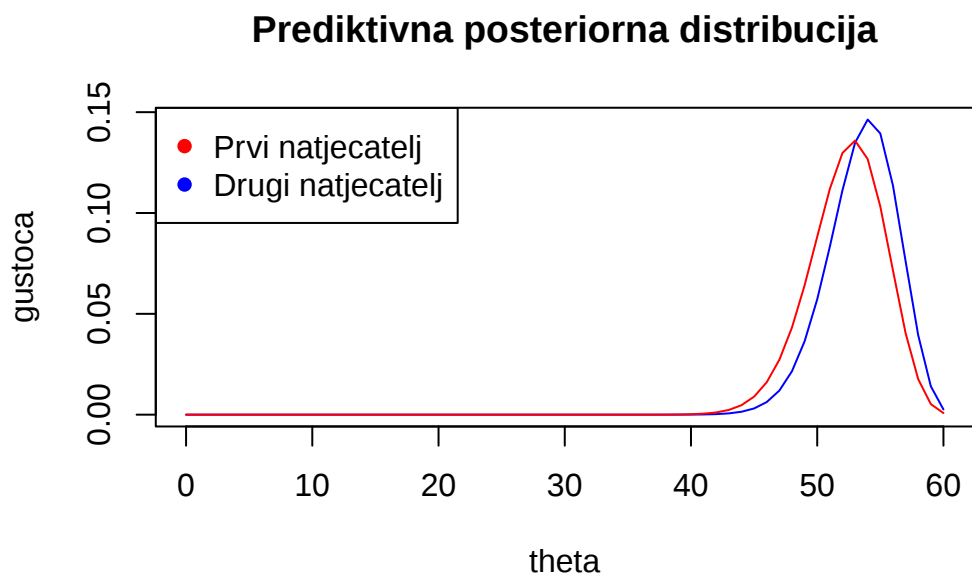
```
1-pbbinom(55,60,170,25)
```

```
[1] 0.1356478
```

```
1-pbbinom(55,60,170.9875,20.7625)
```

```
[1] 0.2458845
```

```
theta<-seq(from=0, to=60, by=1)
plot(theta, dbbinom(theta,60, 170.9875, 20.7625), type="l", col="blue", ylab="gustoća",
lines(theta, dbbinom(theta,60, 170,25), type="l", col="red")
legend("topleft", legend = c("Prvi natjecatelj", "Drugi natjecatelj"), col = c("red", "b
```



Vjerojatnije je da će na idućem natjecanju barem 55 pogodaka mete imati natjecatelj 2.

49

Zadatak 49.1. 1990-tih prikupljeni su podaci o broju djece i stupnju obrazovanja 155 žena u SAD-u. U tijeku provedbe ankete žene su imale 40 godina. Poznato je da je 1970-tih bilo razdoblje niske stope nataliteta u SAD-u (anketirane žene su tada bile u svojim dvadesetima). U sljedećoj tablici prikazani su sumarni podaci o broju djece i stupnju obrazovanja žena:

Obrazovanje žene	Broj anketiranih žena	Ukupan broj djece svih žena
Srednja stručna sprema ili niže	111	217
Prvostupnica ili više	44	66

- Koji parametri su nam od interesa? Definirajte vjerodostojnost Bayesovog modela.
- Odredite konjugirane apriorne i pripadne posteriorne distribucije. Pretpostavite da apriorno očekujete dvoje djece uz varijabilnost (varijancu) 2 za obje skupine. Prikažite grafički dobivene distribucije i komentirajte.
- Usporedite apriorno/posteriorno očekivanje i varijabilnost broja djece žena.
- Koliko iznosi posteriorna vjerojatnost da žene s najviše srednjom stručnom spremom očekivano imaju veći broj djece od žena koje su barem prvostupnice?
- Koliko iznosi posteriorna vjerojatnost da slučajno odabrana žena s najviše srednjom stručnom spremom/barem prvostupnica ima troje ili više djece? Grafički usporedite predikcije za broj djece slučajno odabranih žena koje imaju najviše srednju stručnu spremu i onih koje su barem prvostupnice. Komentirajte.

 Danger

50 Rješenje

a) Zanimaju nas parametri

- $\lambda_1 > 0$ - očekivan broj djece žena s najviše srednjom stručnom spremom
- $\lambda_2 > 0$ - očekivan broj djece žena koje su barem prvostupnice

Vjerodostojnosti:

$$f(\mathbf{x}|\lambda_1) = \frac{\lambda_1^{\sum_{i=1}^n x_i}}{111!} e^{-\lambda_1}, \quad f(\mathbf{y}|\lambda_2) = \frac{\lambda_2^{\sum_{i=1}^n y_i}}{111!} e^{-\lambda_2},$$

gdje je $\mathbf{x} = (x_1, x_2, x_3, \dots, x_{111})$, $\mathbf{y} = (y_1, y_2, \dots, y_{44})$.

b) *Apriorne distribucije:*

- $\lambda_1 \sim \Gamma(2, 1)$
- $\lambda_2 \sim \Gamma(2, 1)$

Posteriorne distribucije:

- $\lambda_1|\mathbf{x} \sim \Gamma\left(2 + \sum_{i=1}^n x_i, 1 + n_1\right) = \Gamma(219, 112)$
- $\lambda_2|\mathbf{y} \sim \Gamma\left(2 + \sum_{i=1}^n y_i, 1 + n_2\right) = \Gamma(68, 45)$

c)

```
alpha <- 2
beta <- 1
n1 <- 111
n2 <- 44
x111<-217
y44<-66

#apriorno:
#obje skupine
alpha/beta #ocekivanje
```

```
[1] 2
```

```
alpha/beta^2 #varijanca
```

```
[1] 2
```

```
#posteriorno:  
#zene sa najviše srednjom strucnom spremom  
(alpha+x111)/(beta+n1) #ocekivanje
```

```
[1] 1.955357
```

```
(alpha+x111)/(beta+n1)^2 #varijanca
```

```
[1] 0.01745855
```

```
#zene koje su barem prvostupnice:  
(alpha+y44)/(beta+n2) #ocekivanje
```

```
[1] 1.511111
```

```
(alpha+x111)/(beta+n1)^2 #varijanca
```

```
[1] 0.01745855
```

d) Izračunajmo

$$P(\lambda_1|\{\mathbb{X} = \mathbb{x}\} > \lambda_2|\{\mathbb{Y} = \mathbb{y}\})$$

numerički, koristeći tzv. Monte Carlo metodu. Simulirajmo 1000 realizacija slučajnih varijabli $\lambda_1|\{\mathbb{X} = \mathbb{x}\}$ i $\lambda_2|\{\mathbb{Y} = \mathbb{y}\}$ te izračunajmo relativnu frekvenciju događaja $\lambda_1|\{\mathbb{X} = \mathbb{x}\} > \lambda_2|\{\mathbb{Y} = \mathbb{y}\}$

```
lambda1 <- rgamma(1000, 219,112)  
lambda2 <- rgamma(1000, 68,45)  
mean(lambda1>lambda2)
```

```
[1] 0.969
```

Prema tome, vjerojatnost da žene s najviše srednjom stručnom spremom očekivano imaju veći broj djece u odnosu na žene koje su barem prvostupnice iznosi približno 97%.

e) Znamo da je u Gama-Poissonovom modelu prediktivna posteriorna distribucija negativna-binomna distribucija, tj.

$$X_{112}|\{\mathbb{X} = x\} \sim \text{NegBin}(219, 112) \Rightarrow P(X_4 \geq 3|\{\mathbb{X} = x\}) \approx 0.311$$

$$Y_{45}|\{\mathbb{Y} = y\} \sim \text{NegBin}(68, 45) \Rightarrow P(Y_4 \geq 3|\{\mathbb{Y} = y\}) \approx 0.195$$

```
1-pnbinom(2,size=219,mu=219/112)
```

```
[1] 0.3112938
```

```
1-pnbinom(2,size=68,mu=68/45)
```

```
[1] 0.1952806
```

```
vjer<-c()
for(i in c(100,1000,10000,100000))
{
  x112 <- rnbinom(i, size=219,mu=219/112)
  y45 <- rnbinom(i, size=68,mu=68/45)
  vjer<-c(vjer,mean(x112>y45))
}
vjer
```

```
[1] 0.52000 0.47200 0.48330 0.48189
```

51

Zadatak 51.1. U bazi podataka **midge** (iz paketa **Flury**) nalaze se podaci o duljini krila dviju vrsta mušica (*Amerohelea fasciata* (Af) i *Pseudofasciata* (Apf)). Pretpostavka je da mjerenja dolaze iz normalne distribucije s poznatim varijancama ($\sigma_1^2 = 0.017$, $\sigma_2^2 = 0.008$). Prethodna istraživanja sugeriraju da je očekivana duljina krila za mušice 1.9mm . Napomena: Baza **Flury** je nedostupna u najnovijoj verziji R-a, te ju možemo dohvatiti putem arhiva:

```
install.packages("remotes")
library(remotes)
install_version("Flury", "0.1-3")
library(Flury)
```

- Odredite konjugiranu apriornu distribuciju i pripadnu posteriornu distribuciju za očekivanu duljinu krila mušica pojedine vrste uz korištenje činjenice da očekivana vrijednost umanjena za dvije standardne devijacije treba biti pozitivna.
- Odredite Bayesov procjenitelj očekivane duljine krila mušica pojedine vrste uz apsolutnu udaljenost kao funkciju rizika.
- Nakon što uzmemo podatke u obzir, koliko iznosi vjerojatnost da je očekivana duljina krila pojedine vrste mušice veća od 2 mm? Usporedite s istom vjerojatnošću prije nego uzmemo podatke u obzir.
- Nakon što uzmemo u obzir podatke, koliko je vjerojatno da slučajnim odabirom jedne mušice svake vrste ustvrdimo da je mušica vrste *Amerohelea fasciata* veća od mušice vrste *Pseudofasciata*?

 Danger

52 Rješenje

```
data(midge)
attach(midge)
```

- a) S obzirom da je vjerodostojnost iz normalne distribucije, zaključujemo da je konjugirana apriorna distribucija normalna, tj. imamo:

$$\mu_1 \sim \mathcal{N}(1.9, 0.95^2), \quad f(x|\mu_1) = \frac{1}{\sqrt{2\pi \cdot 0.017}} e^{-\frac{(x-\mu_1)^2}{2 \cdot 0.017}},$$

pri čemu je varijanca apriorne distribucije dobivena temeljem zahtjeva da očekivana vrijednost umanjena za dvije standardne devijacije treba biti pozitivna:

$$\mu_0 - 2\sigma_0 > 0 \Rightarrow \sigma_0 < 0.95.$$

Pripadna posteriorna distribucija je

$$\mu_1 | \{\mathbb{X} = \mathbb{x}\} \sim \mathcal{N}(1.805, 0.002)$$

što lako dobijemo iz sljedećeg koda:

```
mu0.Af <- 1.9
sigma0.Af <- 0.95
sigma.Af <- sqrt(0.017)
hat.Af <- mean(Wing.Length[Species=="Af"])
n.Af <- length(Wing.Length[Species=="Af"])

mun.Af <- (mu0.Af/sigma0.Af^2+n.Af/sigma.Af^2*hat.Af)/(1/sigma0.Af^2+n.Af/sigma.Af^2)
mun.Af
```

```
[1] 1.804644
```

```
sigman.Af <- 1/(1/sigma0.Af^2+n.Af/sigma.Af^2)
sigman.Af
```

```
[1] 0.001884944
```

Slično, za očekivanu duljinu krila mušice vrste **Pseudofasciata** imamo:

$$\mu_2 \sim \mathcal{N}(1.9, 0.95^2), \quad f(y|\mu_2) = \frac{1}{\sqrt{2\pi \cdot 0.008}} e^{-\frac{(y-\mu_2)^2}{2 \cdot 0.008}},$$

uz pripadnu posteriornu distribuciju

$$\mu_2|\{\mathbb{Y} = y\} \sim \mathcal{N}(1.927, 0.001)$$

što se lako dobije iz koda:

```
mu0.Apfb <- 1.9
sigma0.Apfb <- 0.95
sigma.Apfb <- sqrt(0.008)
hat.Apfb <- mean(Wing.Length[Species=="Apfb"])
n.Apfb <- length(Wing.Length[Species=="Apfb"])

mun.Apfb <- (mu0.Apfb/sigma0.Apfb^2+n.Apfb/sigma.Apfb^2*hat.Apfb)/(1/sigma0.Apfb^2+n.Apfb/sigma.
mun.Apfb
```

```
[1] 1.926627
```

```
sigman.Apfb <-1/(1/sigma0.Apfb^2+n.Apfb/sigma.Apfb^2)
sigman.Apfb
```

```
[1] 0.001331366
```

- b) Znamo da je Bayesov procjenitelj uz apsolutni rizik medijan posteriorne distribucije, što su u ovom slučaju isto što i pripadna očekivanja, tj.:

$$\text{medijan}(\mu_1|\{\mathbb{X} = x\}) = 1.805, \quad \text{medijan}(\mu_2|\{\mathbb{Y} = y\}) = 1.927.$$

c) Imamo:

```
1-pnorm(2, 1.805, sqrt(0.002)) #posteriorna vjerojatnost Af
```

```
[1] 6.493246e-06
```

```
1-pnorm(2, 1.927,sqrt(0.001)) #posteriorna vjerojatnost Apfb
```

```
[1] 0.01048671
```

```
1-pnorm(2, 1.9, 0.95) #apriorna vjerojatnost Af
```

```
[1] 0.4580835
```

```
1-pnorm(2, 1.9,0.95) #apriorna vjerojatnost Apf
```

```
[1] 0.4580835
```

to jest

$$P(\mu_1 > 2) \approx 0.46, \quad P(\mu_1 > 2|\{\mathbb{X} = \mathbb{x}\}) < 0.000006$$

i

$$P(\mu_2 > 2) \approx 0.46, \quad P(\mu_2 > 2|\{\mathbb{Y} = \mathbb{y}\}) \approx 0.01$$

d) Traženu vjerojatnost odredimo temeljem posteriornih prediktivnih distribucija, tj. distribucija

$$X_{10}|\{\mathbb{X} = (x_1, \dots, x_9)\} \sim \mathcal{N}(1.805, 0.037), \quad Y_7|\{\mathbb{Y} = (y_1, \dots, y_6)\} \sim \mathcal{N}(1.927, 0.009).$$

Koristeći Monte Carlo aproksimaciju

```
vjer<-c()
for(i in c(100,1000,10000,100000,1000000))
{
  x10 <- rnorm(i, 1.805,0.037)
  y7 <- rnorm(i, 1.927,0.009)
  vjer<-c(vjer,mean(x10>y7))
}
vjer
```

```
[1] 0.000000 0.000000 0.000300 0.000740 0.000701
```

za traženu vjerojatnost dobivamo

$$P(X_{10} > Y_7|\{\mathbb{X} = (x_1, \dots, x_9), \mathbb{Y} = (y_1, \dots, y_6)\}) \approx 0.0007.$$

53 Vjerodostojna područja (intervali)

i Note

54 α - vjerodostojno područje

Za skup $C_{\mathbf{x}}$ kažemo da je **α -vjerodostojno područje** (uz dane podatke $\mathbf{x}$) za parametar θ ako je

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

Specijalno, ako je $C_{\mathbf{x}} = [l(\mathbf{x}), d(\mathbf{x})] \subset \mathbb{R}$ kažemo da se radi o **α -vjerodostojnom intervalu**.

i Note

55 područje najveće posteriorne gustoće (HPD područje)

Za α -vjerodostojno područje kažemo da je **područje najveće posteriorne gustoće** $s(\mathbf{x})$ i pišemo **HPD područje** (od eng. highest posterior density) ako vrijedi:

- $P(\theta \in C_{\mathbf{x}} | \mathbb{X} = \mathbf{x}) = 1 - \alpha.$
- $\forall \theta_a \in s(\mathbf{x}), \theta_b \notin s(\mathbf{x})$ je $f(\theta_a | \mathbf{x}) > f(\theta_b | \mathbf{x}).$

Primijetimo da sve točke HPD područja imaju veću posteriornu gustoću nego točke izvan područja uz zadani α .



Tip

56 Primjer

Pretpostavimo da 10 puta nezavisno bacamo novčić i da padnu dvije glave. Odredimo vjerodostojni i pouzdani interval uz $\alpha = 0.05$ za nepoznatu proporciju glava θ : - Vjerodostojnost: $f(x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}$ - Podaci: $x = 2$ - Iskoristimo konjugiranu apriornu distribuciju, odnosno pripadni Beta-Binomni model: $f(\theta) = I_{(0,1)}(\theta) \Rightarrow \theta \sim U(0,1)$, odnosno $\theta \sim \text{Beta}(1,1)$ - posteriorna distribucija: $\theta|X=2 \sim \text{Beta}(1+2, 1+10-2) = \text{Beta}(3,9)$

```
f<-2 #podaci (frekvencija glava)
n<-10 #broj bacanja novcica
xhat<-f/n #relativna frekvencija glava
c(xhat-qnrm(0.975)*sqrt(xhat*(1-xhat)/n),xhat+qnrm(0.975)*sqrt(xhat*(1-xhat)/n)) #frek
```

```
[1] -0.04791801  0.44791801
```

Dakle, 95%-tni pouzdani interval za proporciju glava (frekvencionistički pristup) je

[−0.04791801, 0.44791801]

```
qbeta(c(0.025,0.975),3,9) #Bayesov vjerodostojni interval
```

```
[1] 0.06021773 0.51775585
```

S druge strane, 0.05-vjerodostojni interval (Bayesov pristup) je

[0.06021773, 0.51775585]

```
hdi(qbeta,shape1=3,shape2=9,credMass=0.95)
```

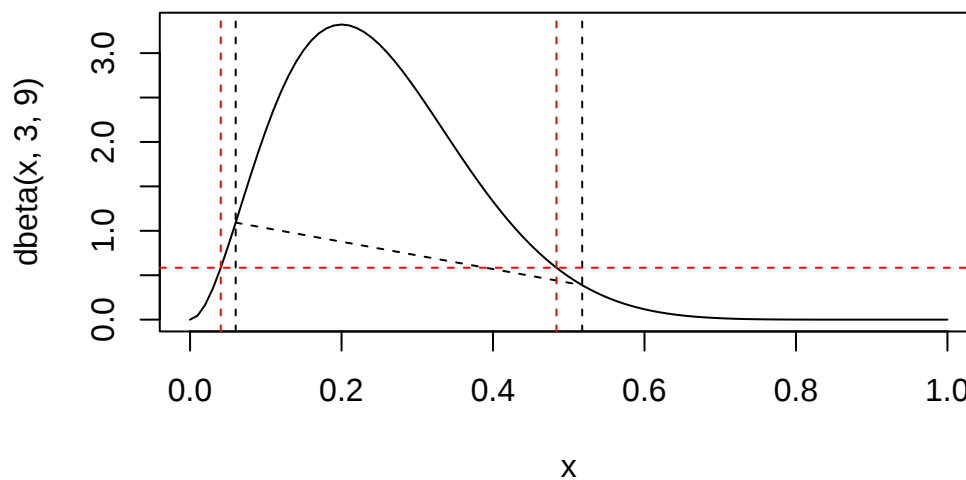
```
      lower      upper
0.04055517 0.48372366
attr(,"credMass")
[1] 0.95
```

Pripadni HPD (highest posterior density) interval je

[0.04055517, 0.48372366]

Sljedeći kod grafički ilustrira “standardni” vjerodostojni interval i HPD interval:

```
curve(dbeta(x,3,9))
abline(v=qbeta(0.025,3,9),lty="dashed")
abline(v=qbeta(0.975,3,9),lty="dashed")
lines(c(qbeta(0.025,3,9),qbeta(0.975,3,9)),c(dbeta(qbeta(0.025,3,9),3,9),dbeta(qbeta(0.975,3,9),3,9)),col="red",lty="dashed")
abline(v=0.04055517 , col="red",lty="dashed")
abline(v=0.48372366 , col="red",lty="dashed")
abline(h=dbeta(0.04055517,3,9),col="red", lty="dashed")
```



Možemo uočiti da je HPD interval uži u odnosu na vjerodostojni interval dobiven kvantilima, pri čemu oba imaju 95% vjerojatnosti da sadrže θ .

57 Zadaci

 Danger

58

Zadatak 58.1. Za nepoznate parametre u zadacima 3.4. - 3.7. odredite 0.05-vjerodostojne intervale standardno i korištenjem metode najveće posteriorne gustoće.

 Danger

59 Rješenje

- Zadatak 3.4. Posteriorna distribucija: $\lambda|\{\mathbb{X} = \mathbb{x}\} \sim \Gamma(4 + 100, 3 + 63.09) = \Gamma(104, 66.09)$

a) “standardni” 0.05-vjerodostojni interval: [1.285755, 1.890111]

```
qgamma(c(0.025,0.975),104,66.09)
```

```
[1] 1.285755 1.890111
```

b) HPD 0.05-vjerodostojni interval: [1.276348, 1.879410]

```
hdi(qgamma,shape=104,rate=66.09,credMass=0.95)
```

```
      lower      upper  
1.276348 1.879410  
attr(,"credMass")  
[1] 0.95
```

- Zadatak 3.5. Posteriorne distribucije:

$$- \theta_1 | \mathbb{x} \sim \text{Beta} \left(\sum_{i=1}^n x_i + 12, 60 \cdot 3 - \sum_{i=1}^n x_i + 3 \right) = \text{Beta}(170, 25)$$

$$- \theta_2 | \mathbb{y} \sim \text{Beta} \left(\sum_{i=1}^n y_i + 9.9875, 60 \cdot 3 - \sum_{i=1}^n y_i + 1.7625 \right) = \text{Beta}(170.9875, 20.7625)$$

a) "standardni" 0.05-vjerodostojni intervali:

- za θ_1 : [0.8215389, 0.9148389]
- za θ_2 : [0.8441667, 0.9315524]

```
qbeta(c(0.025,0.975),170,25)
```

```
[1] 0.8215389 0.9148389
```

```
qbeta(c(0.025,0.975),170.9875, 20.7625)
```

```
[1] 0.8441667 0.9315524
```

b) HPD 0.05-vjerodostojni intervali:

- za θ_1 : [0.8243324, 0.9170963]
- za θ_2 : [0.8471935, 0.9339220]

```
hdi(qbeta,shape1=170,shape2=25,credMass=0.95)
```

```
      lower      upper  
0.8243324 0.9170963  
attr(,"credMass")  
[1] 0.95
```

```
hdi(qbeta,shape1=170.9875,shape2=20.7625,credMass=0.95)
```

```
      lower      upper  
0.8471935 0.9339220  
attr(,"credMass")  
[1] 0.95
```

- Zadatak 3.6. Posteriorne distribucije:

$$\begin{aligned} - \lambda_1 | x &\sim \Gamma \left(2 + \sum_{i=1}^n x_i, 1 + n_1 \right) = \Gamma(219, 112) \\ - \lambda_2 | y &\sim \Gamma \left(2 + \sum_{i=1}^n y_i, 1 + n_2 \right) = \Gamma(68, 45) \end{aligned}$$

a) "standardni" 0.05-vjerodostojni intervali:

- za λ_1 : [1.704943, 2.222679]
- za λ_2 : [1.173437, 1.890836]

```
qgamma(c(0.025,0.975),219,112)
```

```
[1] 1.704943 2.222679
```

```
qgamma(c(0.025,0.975),68, 45)
```

```
[1] 1.173437 1.890836
```

b) HPD 0.05-vjerodostojni intervali:

- za λ_1 : [1.699263, 2.216473]
- za λ_2 : [1.159872, 1.874921]

```
hdi(qgamma,shape=219,rate=112,credMass=0.95)
```

```
      lower      upper
1.699263 2.216473
attr(,"credMass")
[1] 0.95
```

```
hdi(qgamma,shape=68,rate=45,credMass=0.95)
```

```
      lower      upper
1.159872 1.874921
attr(,"credMass")
[1] 0.95
```

- Zadatak 3.7.
 - $\mu_1 | \{\mathbb{X} = \mathbb{x}\} \sim \mathcal{N}(1.805, 0.002)$
 - $\mu_2 | \{\mathbb{Y} = \mathbb{y}\} \sim \mathcal{N}(1.927, 0.001)$
- a) "standardni" 0.05-vjerodostojni intervali:
 - za μ_1 : [1.717348, 1.892652]
 - za μ_2 : [1.86502, 1.98898]

```
qnorm(c(0.025,0.975),1.805,sqrt(0.002))
```

```
[1] 1.717348 1.892652
```

```
qnorm(c(0.025,0.975),1.927, sqrt(0.001))
```

```
[1] 1.86502 1.98898
```

- b) HPD 0.05-vjerodostojni intervali:

- za μ_1 : [1.717348, 1.892652]
- za μ_2 : [1.86502, 1.98898]

```
hdi(qnorm,mean=1.805,sd=sqrt(0.002),credMass=0.95)
```

```
      lower    upper  
1.717348 1.892652  
attr(,"credMass")  
[1] 0.95
```

```
hdi(qnorm,mean=1.927,sd=sqrt(0.001),credMass=0.95)
```

```
      lower    upper  
1.86502 1.98898  
attr(,"credMass")  
[1] 0.95
```

60 Monte Carlo metode

Monte Carlo metode zasnivaju se na simuliranju/generiranju velikog broja slučajnih brojeva u svrhu dobivanja aproksimacija/numeričkih rezultata nekog problema. U kontekstu Bayesovske statistike koristimo ih u svrhu aproksimacije posteriorne distribucije ili njezinih pojedinačnih karakteristika.

Neka je θ parametar od interesa uz zadanu vjerodostojnost $f(x_1, \dots, x_n | \theta)$. Ideja je simulirati/generirati S nezavisnih realizacija parametra θ iz posteriorne distribucije, tj.

$$\theta^{(1)}, \dots, \theta^{(S)} \sim \text{n.j.d. iz } f(\theta | x_1, \dots, x_n).$$

Posteriornu distribuciju $f(\theta | x_1, \dots, x_n)$ možemo aproksimirati empirijskom distribucijom uzorka $\{\theta^{(1)}, \dots, \theta^{(S)}\}$.

Note

61 Monte Carlo aproksimacija

Neka je $\theta^{(1)}, \dots, \theta^{(S)} \sim \text{n.j.d. iz } p(\theta | x_1, \dots, x_n)$. Empirijska distribucija uzorka $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ naziva se **Monte Carlo aproksimacija** posteriorne distribucije $f(\theta | x_1, \dots, x_n)$.

Koristeći Monte Carlo metodu možemo aproksimirati razne numeričke karakteristike nepoznatog parametra θ (kao sl. varijable). Naime, prema jakom zakonu velikih brojeva za (gotovo svaku) funkciju g vrijedi:

$$\frac{1}{S} \sum_{s=1}^S g(\theta^{(s)}) \rightarrow E[g(\theta) | X_1 = x_1, \dots, X_n = x_n], \quad S \rightarrow \infty, \quad \text{g.s.}$$

Prema tome, kada $S \rightarrow \infty$ imamo:

- $\bar{\theta} \rightarrow E[\theta | X_1 = x_1, \dots, X_n = x_n]$

- $\hat{\sigma}^2 = \frac{1}{S-1} \sum_{s=1}^S (\theta^{(s)} - \bar{\theta})^2 \rightarrow \text{Var}[\theta|X_1 = x_1, \dots, X_n = x_n]$
- $\frac{1}{S} \sum_{s=1}^S I_{\{\theta^{(s)} \leq c\}} \rightarrow P(\theta \leq c|X_1 = x_1, \dots, X_n = x_n)$.
- empirijska distribucija od $\{\theta^{(1)}, \dots, \theta^{(S)}\} \rightarrow f(\theta|x_1, \dots, x_n)$.
- $\text{medijan}(\theta^{(1)}, \dots, \theta^{(S)}) \rightarrow \text{medijan}(\theta)$
- p-kvantil od $\{\theta^{(1)}, \dots, \theta^{(S)}\} \rightarrow \theta_p$.

Kod Monte Carlo aproksimacija najveći je problem znati je li došlo do konvergencije, odnosno jesmo li generirali dovoljno slučajnih brojeva (dovoljno velik S). U tu svrhu često se koriste grafički prikazi kojim možemo dobiti uvid u konvergenciju.



Tip

62 Primjer

Promotrimo problem konvergencije Monte-Carlo aproksimacija u kontekstu aproksimacije posteriornog očekivanja. Prema centralnom graničnom teoremu imamo:

$$\frac{\bar{\theta} - E[\theta|X_1 = x_1, \dots, X_n = X_n]}{\sqrt{\text{Var}[\theta|X_1 = x_1, \dots, X_n = X_n]/S}} \Rightarrow \mathcal{N}(0, 1), \quad S \rightarrow \infty.$$

Uzevši u obzir navedeno, aproksimativni 95%-tni Monte Carlo pouzdani interval za posteriorno očekivanje je

$$[\bar{\theta} - 2\sqrt{\hat{\sigma}^2/S}, \bar{\theta} + 2\sqrt{\hat{\sigma}^2/S}].$$

Uobičajena je praksa izabrati dovoljno veliki S tako da standardna greška ($\sqrt{\hat{\sigma}^2/S}$) bude manja od preciznosti do na koju želimo izvjestiti posteriorno očekivanje $E[\theta|X_1 = x_1, \dots, X_n = X_n]$.

Pretpostavimo da smo generirali uzorak dimenzije $S = 100$ za kojeg je aproksimacija za $\text{Var}[\theta|X_1 = x_1, \dots, X_n = X_n]/S \approx 0.025$, a želimo preciznost 0.01. Tada imamo zahtjev oblika

$$2\sqrt{\hat{\sigma}^2/S} < 0.01 \Rightarrow S > 1000.$$

Prema tome trebali bismo napraviti Monte Carlo aproksimaciju na temelju barem $S = 1000$ simuliranih vrijednosti kako bismo imali pouzdanost 0.01 s velikom vjerojatnošću.

62.1 Posteriorno zaključivanje za proizvoljne funkcije parametra

Pretpostavimo da nas zanima statističko zaključivanje o $g(\theta)$ umjesto originalnog parametra θ . Monte Carlo metode nam to jednostavno pružaju kroz sljedeći postupak:

- generiraj $\theta^{(1)} \sim f(\theta|x_1, \dots, x_n)$, izračunaj $\gamma^{(1)} = g(\theta^{(1)})$
- generiraj $\theta^{(2)} \sim f(\theta|x_1, \dots, x_n)$, izračunaj $\gamma^{(2)} = g(\theta^{(2)})$
- $\vdots$
- generiraj $\theta^{(S)} \sim f(\theta|x_1, \dots, x_n)$, izračunaj $\gamma^{(S)} = g(\theta^{(S)})$

Tako generirani uzorak $\{\gamma^{(1)}, \dots, \gamma^{(S)}\}$ sastoji se od S nezavisnih realizacije iz $f(\gamma|x_1, \dots, x_n)$ (iz posteriorne distribucije u terminu $\gamma = g(\theta)$).

Sada slično kao u prethodnom poglavlju, za $S \rightarrow \infty$ imamo:

- $\bar{\gamma} = \frac{1}{S} \sum_{s=1}^S \gamma^{(s)} \rightarrow E[\gamma|X = x_1, \dots, X = x_n]$
- $\hat{\sigma}^2 = \frac{1}{S-1} \sum_{s=1}^S (\gamma^{(s)} - \bar{\gamma})^2 \rightarrow Var[\gamma|X_1 = x_1, \dots, X_n = x_n]$
- empirijska distribucija od $\{\gamma^{(1)}, \dots, \gamma^{(S)}\} \rightarrow f(\gamma|x_1, \dots, x_n)$

...

 Tip

63 Primjer

U Binomnom modelu $B(n, \theta)$ mogu nas zanimati logaritmirane šanse, tj. $\ln \frac{\theta}{1-\theta}$, umjesto originalnog parametra vjerojatnosti θ . U tom kontekstu želimo zaključivati o $\gamma = g(\theta) = \ln \frac{\theta}{1-\theta}$. Uočimo da je $\theta \in \langle 0, 1 \rangle$, a $\gamma \in \mathbb{R}$.

63.1 Monte Carlo Markovljevi lanci

63.1.1 Metropolis-Hastings algoritam

U kontekstu Bayesovske statistike ovaj algoritam se koristi u svrhu aproksimacije posteriorne distribucije, a onda Monte Carlo metodama aproksimiramo i sve druge numeričke karakteristike posteriorne distribucije.

Algoritam se bazira na iterativnom generiranju vrijednosti iz posteriorne distribucije korištenjem Markovljevih lanaca, a potom se koristi empirijska distribucija generiranih vrijednosti kao aproksimacija posteriorne distribucije. Iz gore navedenih razloga Metropolis-Hastings algoritam spada u MCMC metode (Markov Chain Monte Carlo).

U svakom Bayesovskom modelu zadan je parametar θ , pripadna apriorna distribucija $f(\theta)$ i vjerodostojnosti $f(\mathbf{z}|\theta)$. To je dovoljno za aproksimaciju posteriorne distribucije $f(\theta|\mathbf{z})$ MCMC metodom:

i Note

64 Metropolis Hastings algoritam (Bayesovska verzija)

1. Generiraj $\theta^* \sim T(\theta|\theta^*)$
2. Izračunaj Hastingsov omjer (“acceptance rate”): $r = a(\theta^{(s)}, \theta^*) = \frac{f(\theta^*|\mathbf{z})T(\theta^{(s)}|\theta^*)}{f(\theta^{(s)}|\mathbf{z})T(\theta^*|\theta^{(s)})}$
3. Generiraj $U \sim \mathcal{U}(0, 1)$
4. $\theta^{(s+1)} = \begin{cases} \theta^* & , \quad U \leq r \\ \theta^{(s)} & , \quad U > r \end{cases}$

65 Konvergencija Metropolis Hastings algoritma

Teorem 65.1. *Niz $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(s)}, \dots)$ konstruiran Metropolis Hastings algoritmom je (reverzibilan) Markovljev lanac sa stacionarnom distribucijom $f(\theta|\mathbf{x})$.*

Više informacija o teoretskom aspektu algoritma (kao i dokaz prethodnog teorema) moguće je pronaći u knjigama Hoff (2009) i Dobrow (2016).

Primijetimo da u koraku 2. za izračun Hastingsovog omjera nije potrebno znati normalizacijsku konstantu posteriorne distribucije jer se ona u omjeru pokradi.

Usporedba U i r u koraku 4. se često radi na logaritamskoj skali radi numeričke stabilnosti algoritma.

U slučaju da koristimo simetričnu distribuciju kandidat, tj. ukoliko vrijedi

$$T(\theta^{(a)}|\theta^{(b)}) = T(\theta^{(b)}|\theta^{(a)})$$

algoritam se specijalno zove Metropolis algoritam, a izračun Hastingsovog omjera se svodi na

$$r = a(\theta^{(s)}, a^*) = \frac{f(x|\theta^*)f(\theta^*)}{f(x|\theta^{(s)})f(\theta^{(s)})}.$$

Za kandidat (simetričnu) distribuciju često koristimo:

- $T(\theta^*|\theta^{(s)}) \sim \mathcal{U}(\theta^{(s)} - \delta, \theta^{(s)} + \delta)$
- $T(\theta^*|\theta^{(s)}) \sim \mathcal{N}(\theta^{(s)}, \delta^2)$ gdje je $\delta > 0$ parametar kojeg treba birati tako da algoritam bude efikasan (r između 20% i 50%).

Efikasnost algoritma ponajviše ovisi o izboru početne vrijednosti i distribucije kandidata.



Tip

66 Pravilo sagorijevanja (“burn in period”)

S obzirom da Metropolis-Hastings algoritam generira niz vrijednosti $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(s)}, \dots\}$ koje su zavisne koristimo pravilo sagorijevanja, tj. “burn in period”:

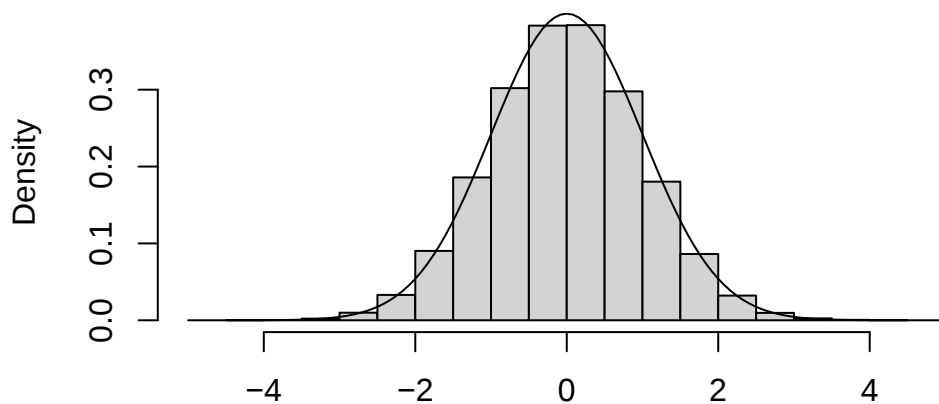
1. Izvršavaj algoritam do iteracije B za koju je Markovljev lanac postigao stacionarnost
2. Izvršavaj algoritam dodatnih S puta
3. Odbaci $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)}\}$ i koristi empirijsku distribuciju od $\{\theta^{(B+1)}, \theta^{(2)}, \dots, \theta^{(B+S)}\}$ kao aproksimaciju za $f(\theta|\mathbb{x})$.



Tip

67 Primjer - uzorkovanje standardne normalne distribucije

```
trials <- 1000000
simlist <- numeric(trials)
state <- 0
for (i in 2:trials) {
  prop <- runif(1, state-1, state+1)
  acc <- exp(-(prop^2-state^2)/2)
  if (runif(1) < acc) state <- prop
  simlist[i] <- state
}
hist(simlist, xlab="", main="", prob=T)
curve(dnorm(x), -4, 4, add=T)
```





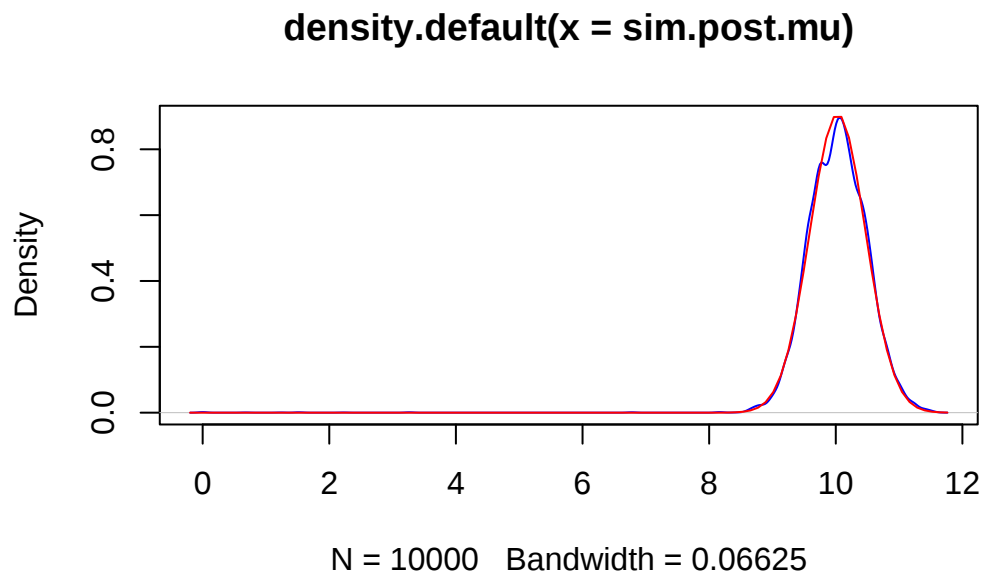
Tip



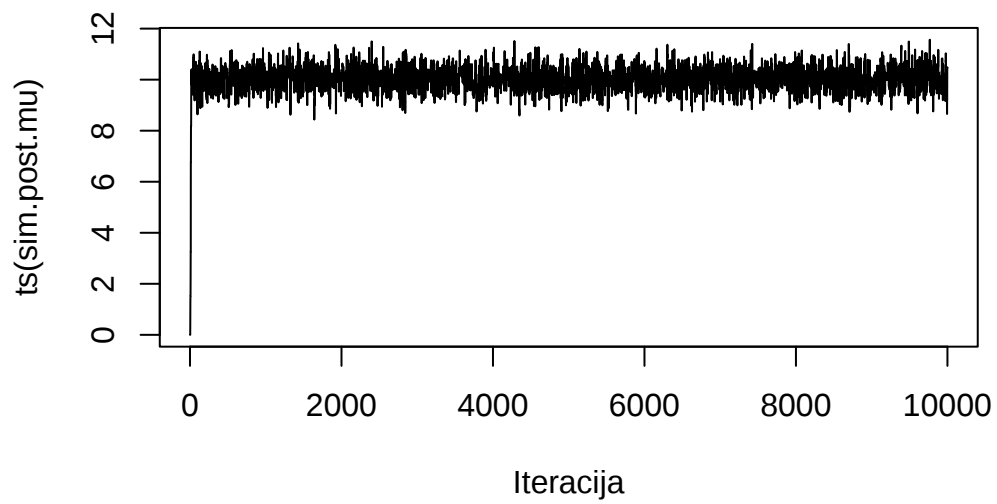
68 Primjer - normalni konjugirani model (poznata varijanca)

```
S <- 10000
sigma2<-1
mu0 <-5
sigma.02<-10
mu <- 0
delta2 <- 2
x <- c(9.37, 10.18, 9.16, 11.6, 10.33)
sim.post.mu <- NULL
for(s in 1:S)
{mu.star<-rnorm(1,mu, sqrt(delta2))
log.r <- (sum(dnorm(x,mu.star, sqrt(sigma2), log=TRUE))+dnorm(mu.star,mu0, sqrt(sigma.02),
  (sum(dnorm(x,mu, sqrt(sigma2), log=TRUE))+dnorm(mu,mu0, sqrt(sigma.02), log=TRUE))
if (log(runif(1)) < log.r) {mu <-mu.star}
sim.post.mu <-c(sim.post.mu, mu)
}

plot(density(sim.post.mu), col="blue")
curve(dnorm(x,10.03, 0.44), add=T, col="red")
```



```
plot(ts(sim.post.mu), xlab="Iteracija")
```



68.0.1 Gibbsovo uzorkovanje

Gibbsovo uzorkovanje je specijalizirana varijanta Metropolis Hastings algoritma koja se koristi u slučaju višedimenzionalnog parametra.

Pretpostavimo da je parametar od interesa p dimenzionalni vektor $\phi = (\phi_1, \dots, \phi_p)$ uz vjerodostojnost $f(\mathbf{x}|\phi)$ i apriornu distribuciju $f(\phi)$.

Algoritam se svodi na iterativno generiranje jednodimenzionalnih komponenti parametra ϕ_k iz uvjetne distribucije $f(\phi_k|\phi_1, \dots, \phi_p, \mathbf{x})$ uz poznate ostale vrijednosti komponenti parametra i podatke. Pri tome se uvijek koriste zadnje generirane vrijednosti komponenti parametra.

 Note

69 Gibbsovo uzorkovanje (Bayesovska verzija) - algoritam

1. Izaberi početnu vrijednost parametra $\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)})$.
2. Generiraj $\phi^{(s)}$ iz $\phi^{(s-1)}$ na sljedeći način:
 - 2.1. Generiraj $\phi_1^{(s)} \sim f(\phi_1|\phi_2^{(s-1)}, \phi_3^{(s-1)}, \dots, \phi_p^{(s-1)}, \mathbf{x})$
 - 2.2. Generiraj $\phi_2^{(s)} \sim f(\phi_2|\phi_1^{(s)}, \phi_3^{(s-1)}, \dots, \phi_p^{(s-1)}, \mathbf{x})$
 - ...
 - 2.p. Generiraj $\phi_p^{(s)} \sim f(\phi_p|\phi_1^{(s)}, \phi_2^{(s)}, \dots, \phi_{p-1}^{(s)}, \mathbf{x})$

Više informacija o teoretskom aspektu algoritma (kao i dokaz prethodnog teorema) moguće je pronaći u knjigama Hoff (2009) i Dobrow (2016).

Efikasnost Gibbsovog uzorkovanja ima naslijeđene probleme kao i ostale MCMC metode kao posljedica činjenice da se radi o specifičnoj verziji Metropolis Hastings algoritmu.

Gibbsovo uzorkovanje podrazumijeva da uvjetne distribucije jednodimenzionalnih komponenti parametra uz ostale poznate vrijednosti znamo (vidi korak 2. algoritma). Ukoliko pak iste ne znamo, mogli bismo za svaku od njih provesti Metropolis Hastings algoritam ("Metropolis with Gibbs"), međutim to može smanjiti efikasnost i brzinu konvergencije algoritma te povećati osjetljivost na početnu vrijednost.



Tip

70 Ergodski teorem

- MC metode oslanjaju se na jaki zakon velikih brojeva kako bi opravdali korištenje empirijske distribucije generiranih vrijednosti u svrhu aproksimacije posteriorne distribucije
- MCMC metode oslanjaju se na ergodski teorem za Markovljeve lance (svojevrсна generalizacija JZVB u slučaju Markovljevih lanaca) kako bi donijeli zaključke:

$$- P(\phi^{(s)} \in A) \rightarrow \int_A f(\phi|\mathbf{z})d\phi, \quad S \rightarrow \infty$$

$$- \frac{1}{S} \sum_{s=1}^S g(\phi^{(s)}) \rightarrow E[g(\phi)|\mathbf{Z} = \mathbf{z}] = \int g(\phi)f(\phi|\mathbf{z})d\phi, \quad S \rightarrow \infty$$

Više o Ergodskom teoremu za Markovljeve lance moguće je pronaći u knjizi Dobrow (2016).

 Tip

71 Primjer - normalni konjugirani model (nepoznato očekivanje, nepoznata varijanca)

Imamo $(X_1, \dots, X_n)$ j.s.u. iz $\mathcal{N}(\mu, \sigma^2)$ distribucije, gdje je $\phi = (\mu, \sigma^2)$ nepoznati parametar. Želimo simulirati vrijednosti iz posteriorne distribucije

$$f(\mu, \sigma^2 | \mathbf{x}).$$

Sljedeći kod provodi Gibbsovo uzorkovanje koje možemo sažeti na idući način:

- početna vrijednost: $\phi^{(0)} = (\mu^{(0)}, \sigma^{2(0)}) = (\bar{x}_n, s_n^2)$ (početna vrijednost može biti proizvoljna, ali radi ubrzavanja algoritma dobro je birati vrijednosti koje su koncentrirane u dijelovima velike gustoće posteriorne distribucije, npr. mod)
- generiraj ϕ^s iz $\phi^{(s-1)}$ korištenjem uvjetnih distribucija:
 - generiraj $\mu^{(s)} \sim f(\mu | \sigma^{2(s-1)}, \mathbf{x})$ (normalni model uz poznatu varijancu)
 - generiraj $\sigma^{2(s)} \sim f(\sigma^2 | \mu^{(s)}, \mathbf{x})$ (normalni model uz poznato očekivanje)

U konačnici dobivamo niz S parova

$$\{(\mu^{(0)}, \sigma^{2(0)}), \dots, (\mu^{(S)}, \sigma^{2(S)})\}$$

iz posteriorne distribucije.

```

# Hiperparametri apriorne distribucije
mu0<-1.9 ; t20<-0.95^2
s20<-0.01 ; nu0<-1

# Podaci
y<-c(1.64,1.70,1.72,1.74,1.82,1.82,1.82,1.90,2.08)
n<-length(y) ; mean.y<-mean(y) ; var.y<-var(y)

## Početne vrijednosti za Gibbsovo uzorkovanje
S<-1000
PHI<-matrix(nrow=S,ncol=2)
PHI[1,]<-phi<-c( mean.y, 1/var.y)

## Gibbsovo uzorkovanje
for(s in 2:S) {

# generiramo novi *mu* iz njegove uvjetne distribucije (normalni model s poznatom varijancijom)
mun<- ( mu0/t20 + n*mean.y*phi[2] ) / ( 1/t20 + n*phi[2] )
t2n<- 1/( 1/t20 + n*phi[2] )
phi[1]<-rnorm(1, mun, sqrt(t2n) )

#generiraj novi *sigma^2* iz njegove uvjetne distribucije (normalni model s poznatim očekivanim srednjim)
nun<- nu0+n
s2n<- (nu0*s20 + (n-1)*var.y + n*(mean.y-phi[1])^2 ) /nun
phi[2]<- rgamma(1, nun/2, nun*s2n/2)

PHI[s,]<-phi
}

```

Radi veće efikasnosti algoritma prilikom ažuriranja drugog parametra kod generiranja novog σ^2 korišten je sljedeći identitet:

$$\sum_{i=1}^n (x_i - \mu)^2 = (n-1)s_n^2 + n(\bar{x}_n - \mu)^2.$$

Sljedeći dio koda grafički ilustrira iteracije u Gibbsovom uzorkovanju:

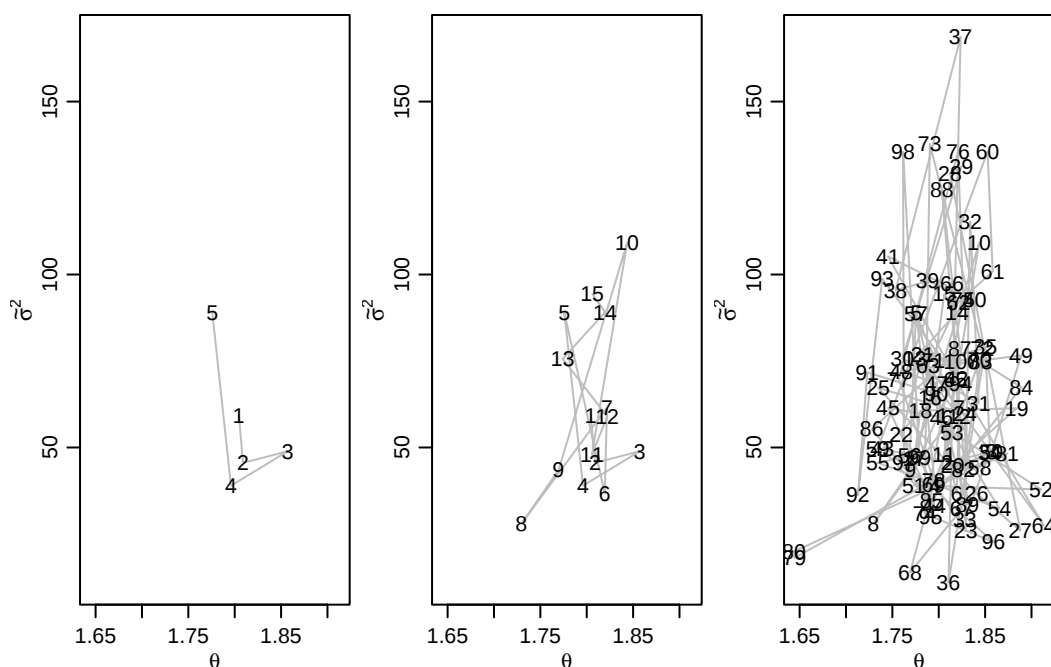
```

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0))
m1<-5
plot( PHI[1:m1,],type="l",xlim=range(PHI[1:100,1]), ylim=range(PHI[1:100,2]),
      lty=1,col="gray",xlab=expression(theta),ylab=expression(tilde(sigma)^2))
text( PHI[1:m1,1], PHI[1:m1,2], c(1:m1) )

m1<-15
plot( PHI[1:m1,],type="l",xlim=range(PHI[1:100,1]), ylim=range(PHI[1:100,2]),
      lty=1,col="gray",xlab=expression(theta),ylab=expression(tilde(sigma)^2))
text( PHI[1:m1,1], PHI[1:m1,2], c(1:m1) )

m1<-100
plot( PHI[1:m1,],type="l",xlim=range(PHI[1:100,1]), ylim=range(PHI[1:100,2]),
      lty=1,col="gray",xlab=expression(theta),ylab=expression(tilde(sigma)^2))
text( PHI[1:m1,1], PHI[1:m1,2], c(1:m1) )

```



Na sljedećim prikazima možemo vidjeti uzorak generiran Gibbsovim uzorkovanjem, empirijske marginalne distribucije s vertikalnim linijama koje reprezentiraju pouzdani interval tako dobivene empirijske distribucije i rezultata koji bi dao frekvencionistički pouzdani interval za očekivanje.

```

# grid
G<-100 ; H<-100

mean.grid<-seq(1.505,2.00,length=G)
prec.grid<-seq(1.75,175,length=H)

post.grid<-matrix(nrow=G,ncol=H)

for(g in 1:G) {
  for(h in 1:H) {

    post.grid[g,h]<- dnorm(mean.grid[g], mu0, sqrt(t20)) *
                     dgamma(prec.grid[h], nu0/2, s20*nu0/2 ) *
                     prod( dnorm(y,mean.grid[g],1/sqrt(prec.grid[h])) )
  }
}

post.grid<-post.grid/sum(post.grid)
par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0))
sseq<-1:1000

image( mean.grid,prec.grid,post.grid,col=gray( (10:0)/10 ),
       xlab=expression(theta), ylab=expression(tilde(sigma)^2) ,
       xlim=range(PHI[,1]),ylim=range(PHI[,2]) )
points(PHI[sseq,1],PHI[sseq,2],pch=".",cex=1.25 )

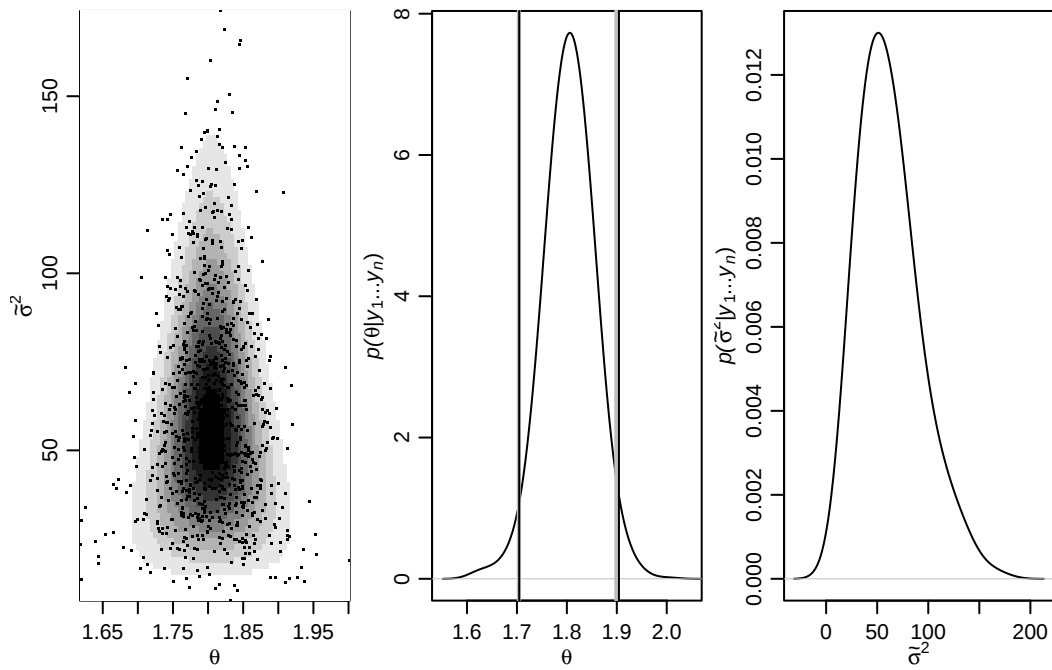
plot(density(PHI[,1],adj=2), xlab=expression(theta),main="",
      xlim=c(1.55,2.05),
      ylab=expression( paste(italic("p("),
                             theta,"|",italic(y[1]),"...",italic(y[n]),")",sep="")))
abline(v=quantile(PHI[,1],prob=c(.025,.975)),lwd=2,col="gray")

n<-length(y) ; ybar<-mean(y) ; s2<-var(y);

abline( v= ybar+qt( c(.025,.975), n-1) *sqrt(s2/n), col="black",lwd=1)

plot(density(PHI[,2],adj=2), xlab=expression(tilde(sigma)^2),main="",
      ylab=expression( paste(italic("p("),
                             tilde(sigma)^2,"|",italic(y[1]),"...",italic(y[n]),")",sep="")))

```



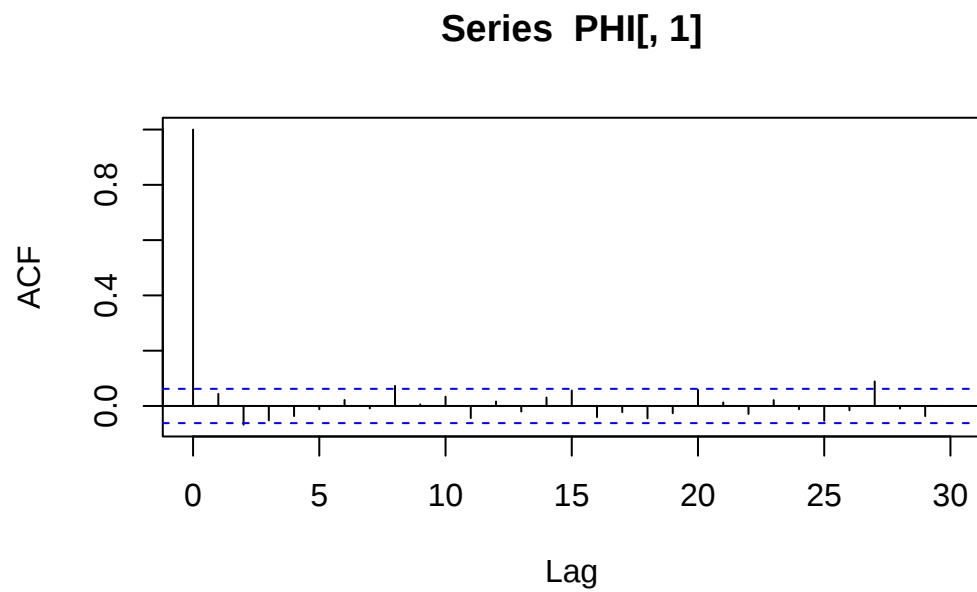
Efektivne veličine uzoraka i pripadne autokorelacije na koraku 1:

$$S_{eff}(\theta) = 1152, \quad S_{eff}(\sigma^2) = 1166, \quad acf_{\theta}(1) = 0.044, \quad acf_{\sigma^2}(1) = 0.096$$

```
effectiveSize(PHI[,1])
```

```
var1  
1151.948
```

```
acf(PHI[,1])[1]
```



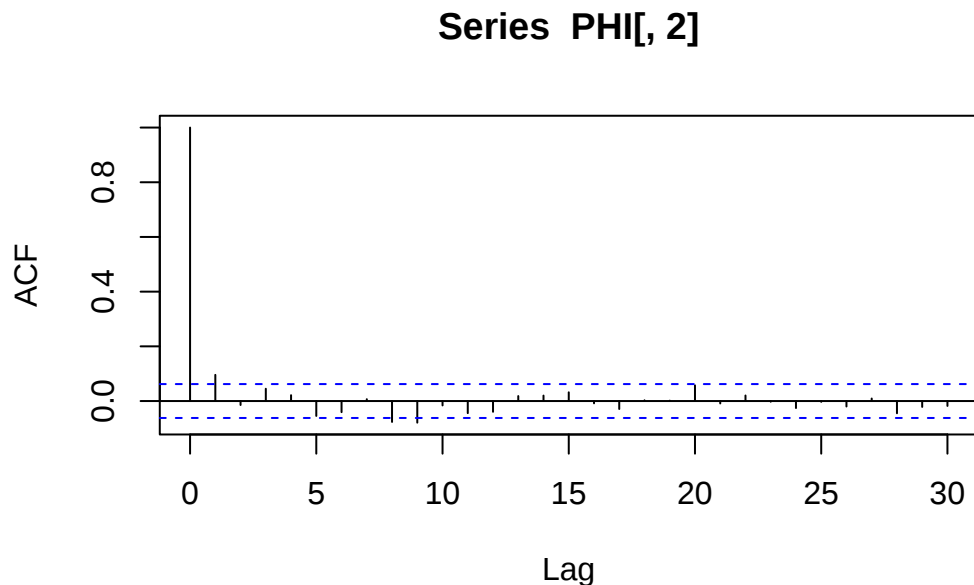
Autocorrelations of series 'PHI[, 1]', by lag

```
1  
0.044
```

```
effectiveSize(PHI[,2])
```

```
var1  
1166.018
```

```
acf(PHI[,2])[1]
```



Autocorrelations of series 'PHI[, 2]', by lag

```

1
0.096

```

Dobivene mjere sugeriraju da je MCMC lanac konvergirao i da je aproksimacija približno kvalitetna kao MC aproksimacija dimnezije \$ 1200\$. Autokorelacija na lagu 1 je vrlo mala.

71.0.1 O brzini konvergencije MCMC metoda

Za MCMC aproksimaciju integrala $\int_A f(\phi) d\phi$, gdje je f neka gustoća, od presudne važnosti je da Markovljev lanac provede vrijeme u skupu A proporcionalno sa stvarnom vjerojatnošću $P(A) = \int_A f(\phi) d\phi$.

U svrhu ilustracije, neka su A_1 , A_2 i A_3 skupovi takvi da je $P(A_2) < P(A_1) \approx P(A_3)$. Da bi MCMC aproksimacija bila dobra ML bi trebao najmanje vremena provesti u skupu A_2 , a podjednako (više) vremena u A_1 i A_3 . Općenito, ukoliko ML kreće iz stanja koje se nalazi u području male vjerojatnosti, trebat će nam velik broj iteracija MCMC algoritma (S) dok ML

- 1) ne izađe iz A_2 i prijeđe u područje veće vjerojatnosti i
- 2) ne prošeće između A_1 i A_3 kao i ostalih skupova velike vjerojatnosti.

Pri tome, 1) dovodimo u vezu sa stacionarnošću, odnosno konvergencijom lanca (ukoliko ML kreće iz područja s velikom vjerojatnošću stacionarnost nije problem).

S druge strane 2) dovodimo u vezu s brzinom kretanja ML kroz parametarski prostor koje se formalno zove “brzina miješanja” (eng. speed of mixing). MC metode imaju savršenu brzinu miješanja (mixing) jer nezavisno uzorkujemo (nema korelacije). S druge strane, MCMC metode imaju lošu brzinu miješanja zbog koreliranosti u uzorku.

Postavlja se pitanje koliko koreliranost utječe na brzinu miješanja/kvalitetu aproksimacije? Ilustrirajmo to na primjeru aproksimacije očekivanja $E[\phi] = \phi_0 = \int \phi f(\phi) d\phi$. Pretpostavimo da raspoložemo slučajnim uzorkom $(\phi^{(1)}, \dots, \phi^{(S)})$. Nepristran (frekvencionistički) procjenitelj za ϕ_0 je $\bar{\phi} = \frac{1}{S} \sum_{s=1}^S \phi^{(s)}$.

- MC metoda (nezavisno uzorkovanje)

$$Var_{MC} [\bar{\phi}] = E [(\bar{\phi} - \phi_0)^2] = \frac{Var\phi}{S}, \quad Var\phi = \int (\phi - \phi_0)^2 f(\phi) d\phi.$$

Pri tome je $\sqrt{Var_{MC} [\bar{\phi}]}$ pripadna Monte Carlo standardna greška (govori nam koliko dobro očekujemo da $\bar{\phi}$ aproksimira ϕ_0). Drugim riječima (prema CGT), interval

$$\left[\bar{\phi} - 2\sqrt{Var_{MC}(\bar{\phi})}, \bar{\phi} + 2\sqrt{Var_{MC}(\bar{\phi})} \right]$$

bi u pravilu u približno 95% slučajeva sadržavao stvarnu vrijednost ϕ_0 .

- MCMC metoda (npr. Metropolis-Hastings algoritam, Gibbsovo uzorkovanje)

$$\begin{aligned} Var_{MC} [\bar{\phi}] &= E [(\bar{\phi} - \phi_0)^2] = E \left[\left(\frac{1}{S} \sum_{s=1}^S (\phi^{(s)} - \phi_0) \right)^2 \right] \\ &= \frac{1}{S^2} \sum_{s=1}^S E [(\phi^{(s)} - \phi_0)^2] + \frac{1}{S^2} \sum_{s \neq t} E [(\phi^{(s)} - \phi_0) (\phi^{(t)} - \phi_0)] \\ &= Var_{MC} [\bar{\phi}] + \frac{1}{S^2} \sum_{s \neq t} E [(\phi^{(s)} - \phi_0) (\phi^{(t)} - \phi_0)] \end{aligned}$$

, pri čemu zadnji član ovisi o korelaciji u ML i uvijek je pozitivan.

U tom kontekstu uočavamo

$$Var_{MCMC} [\bar{\phi}] > Var_{MC} [\bar{\phi}].$$

Riječima, MCMC aproksimacija je uvijek lošija od MC aproksimacije, a što je autokorelacija u lancu veća to je aproksimacija lošija.

Korisna usporedba MC i MCMC metoda radimo s tzv. efektivnom veličinom uzorka (eng. effective sample size):

i Note

72 efektivna veličina uzorka

Broj S_{eff} t.d. je $Var_{MCMC}[\bar{\phi}] = Var[\phi]/S_{eff}$ nazivamo efektivna veličina uzorka.

Interpretiramo ga kao broj nezavisnih generiranja MC metode potreban da dobijemo istu preciznost kao MCMC metoda koju koristimo.

Generalno, MC metoda oslanjanju se na jaki zakon velikih brojeva kako bi opravdali korištenje empirijske distribucije simuliranih vrijednosti u svrhu aproksimacije posteriorne distribucije.

S druge strane, MCMC metode oslanjanju se na ergodski teorem za Markovljeve lance (svojevrsna generalizacija JZVB u kontekstu ML) koji osigurava:

-
-

$$P(\phi^{(s)} \in A) \rightarrow \int_A f(\phi|\mathbf{x})d\phi, \quad S \rightarrow \infty$$

$$\frac{1}{S} \sum_{s=1}^S g(\phi^{(s)}) \rightarrow E[g(\phi)|\mathbf{X} = \mathbf{x}] = \int g(\phi)f(\phi|\mathbf{x})d\phi, \quad S \rightarrow \infty.$$

Zbog strukture zavisnosti niza generiranih vrijednosti $(\phi^{(1)}, \dots, \phi^{(S)})$ kao Markovljevog lanca, postoji snažna koreliranost među vrijednostima niza, pa (nerijetko) trebamo simulirati daleko više vrijednosti nego u slučajnu MC metoda kako bi osigurali zadovoljavajuće aproksimacije.

Zbog toga je uvijek dobro provjeriti kakva je koreliranost lanca korištene MCMC metode pomoću

`acf()`

na različitim koracima (“lagovima”), npr. 1, 10, 50.

72.0.1 (Ostale) MCMC metode i njihova implementacija

U prethodnim poglavljima detaljnije smo objasnili Metropolis-Hastings algoritam i Gibbsovo uzorkovanje. Intuitivno je jasno da u jednodimenzionalnim parametarskim modelima koristimo Metropolis-Hastings algoritam, dok u multivarijatnom slučaju koristimo Gibbsovo uzorkovanje. Samostalna (“ručna”) implementacija ovih algoritama postaje redundantna s povećanjem kompleksnosti Bayesovskog modela. Iz tog razloga nastale su mnoge implementacije ovih (i drugih) MCMC algoritama u svrhu aproksimacije posteriorne distribucije kao što su:

- JAGS (“Just another Gibbs samples”): kao što i sam naziv kaže, radi se o implementaciji Gibbsovog uzorkovanja koja za nas obavlja tehnički posao, a mi trebamo samo unijeti komponente Bayesovskog modela (<https://mcmc-jags.sourceforge.io/>)
- BUGS (“Bayesian inference Using Gibbs Sampling”): iako ne u potpunosti ista implementacija, radi slično što i JAGS (<http://www.bayesianscientific.org/resource/bugs-openbugs-winbugs/>)
- STAN: radi se o implementaciji više vrsta MCMC metoda, a koje nisu Gibbsovo uzorkovanje, primjerice trenutno vrlo popularnoj Hamiltonian Monte Carlo (HMC) metodi (<https://mc-stan.org/>)

```
install.packages("rstan", repos = c("https://mc-stan.org/r-packages/", getOption("repos"))
library("rstan")
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
```

STAN struktura u R-u tipično izgleda ovako:

```
MojeImeModela <- "      #Ime modela
data {
  real <lower = _, upper = _> x[broj_podataka];          // podaci
}
parameters {
  real parametar1;          //realni parametar
  real<lower = 0> parametar2;
}
model {
  x ~ ime_distribucije(parametri_distribucije); // vjerodostojnost
  parametar1 ~ ime_distribucije(hiperparametri) //apriorne distribucije (ukoliko nisu defi
  parametar2 ~ ime_distribucije(hiperparametri)
}

generated quantities {                                     //funkcije parametra koje želimo generirati (
```

```

    real x_tilda = imedistribucije_rng(parametri);
}"
simulacija<-stan(
  model_code = _____,      #IME modela
  data = list(x=_____),      # Podaci
  chains = _____,          # broj korištenih Markovljevih lanaca
  warmup = _____,          # burn in period za svaki lanac
  iter = _____,            # ukupan broj iteracija svakog lanca
  cores = _____,           # broj korištenih jezgri
  refresh = _____,         # no progress show
  seed=_____
)

```

Nakon generiranja simuliranih vrijednosti posteriorne distribucije, analizu brzine konvergencije i brzine miješanja algoritma, kao i vizualni/sumarni pregled posteriorne distribucije dobivamo na sljedeći način:

```

# grafički prikaz gustoće/histograma za pojedinihi parametar koristeći sve lanace algoritma
# pars je ime parametra za kojeg crtamo usporedbu
mcmc_dens(my_sim, pars = "___") +
  yaxis_text(TRUE) +
  ylab("density")

mcmc_hist(my_sim, pars = "___") +
  yaxis_text(TRUE) +
  ylab("count")

# Vizualna dijagnostika: trajektorija simuliranih vrijednosti parametara
mcmc_trace(my_sim, pars = "___", size = 0.5)
mcmc_trace(my_sim, pars = "___", size = 0.5, window = c(0, ___)) # Zumirani prikaz

# Vizualna dijagnostika: prikaz gustoće za pojedini parametar za svaki lanac posebno - slu
mcmc_dens_overlay(my_sim, pars = "___")

# Vizualna dijagnostika: autokorelacijska funkcija za pojedini parametar - govori o brzini
mcmc_acf(my_sim, pars = "___")

```

```
# Numerička dijagnostika: Rhat, efektivna veličina uzorka, sumarni pregled numeričkih karak
print(MojeImeModela)

#dohvacanje simuliranih vrijednosti
#parametra
rstan::extract(sim.MCMC.stan, permuted = TRUE)
```

Koju od implementacija izabrati za korištenje nije jednostavno odgovoriti i ovisi o više faktora, posebice o vrsti problema koji proučavamo. Primjerice, u slučaju (semi)-konjugatnih Bayesovskih modela obično JAGS-ova implementacija Gibbsovog uzorkovanja funkcionira jako dobro, dok u općenitijem i kompleksnijem slučaju HMC metode implementirane u platformi STAN daju bolje rezultate.

72.1 Zadaci

 Danger

73

Zadatak 73.1. Za zadatak 3.4 na temelju zadane vjerodostojnosti i apriorne distribucije odredite:

- Monte Carlo aproksimaciju pripadnih posteriornih distribucija uz grafičku vizualizaciju konvergencije aproksimacije.
- Monte Carlo aproksimaciju posteriornog očekivanja uz grafičku vizualizaciju konvergencije aproksimacije.
- Odredite Monte Carlo aproksimaciju posteriorne vjerojatnosti da je stopa pojavljivanja potresa barem 2 po satu uz usporedbu sa stvarnom vjerojatnosti.

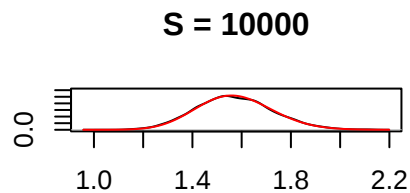
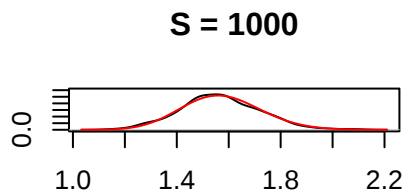
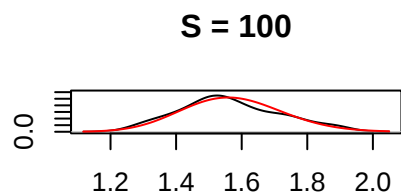
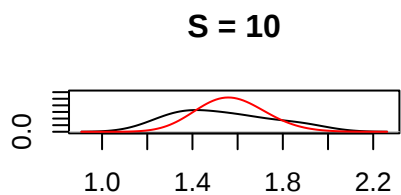
 Danger

74 Rješenje

- a) Vizualizacija konvergencije Monte Carlo aproksimacije posteriorne distribucije $f(\lambda|x_1, \dots, x_n)$ prema stvarnoj distribuciji (crvena boja).

```
alpha <- 4
beta <- 3
sum.x <- 63.09
n <- 100

par(mfrow=c(2,2))
for(i in c(10,100,1000,10000))
{
  lambda.mc <- rgamma(i,alpha+n,beta+sum.x)
  plot(density(rgamma(i,104,66.09)),ylim = c(0, 3), main = paste("S =",i), xlab="", ylab="")
  curve(dgamma(x, alpha+n,beta+sum.x), col = 'red', add=T)
}
```



b) Monte Carlo aproksimacija posteriornog očekivanja:

```
aprox.po<-c()
for(i in c(10,100,1000,10000))
{
  lambda.mc <- rgamma(i,alpha+n,beta+sum.x)
  aprox.po<-c(aprox.po,mean(lambda.mc))
}
aprox.po
```

```
[1] 1.558213 1.583106 1.565242 1.573105
```

Prema tome Monte Carlo aproksimacija posteriornog očekivanja dimenzije $S = 10000$ je

$$\frac{1}{10000} \sum_{s=1}^{10000} \lambda^{(10000)} = 1.573105$$

dok je stvarna vrijednost posteriornog očekivanja

$$E[\lambda|X_1 = x_1, \dots, X_n = x_n] = 1.573612.$$

```
(alpha+n)/(beta+sum.x)
```

```
[1] 1.573612
```

c) Monte Carlo aproksimacija posteriorne vjerojatnosti $P(\lambda > 2|X_1 = x_1, \dots, X_n = x_n)$:

```
aprox.pv2<-c()
for(i in c(10,100,1000,10000,100000,1000000))
{
  lambda.mc <- rgamma(i,alpha+n,beta+sum.x)
  aprox.pv2<-c(aprox.pv2,mean(lambda.mc>2))
}
aprox.pv2
```

```
[1] 0.000000 0.020000 0.007000 0.005700 0.005110 0.005003
```

Stoga je tražena aproksimacija

$$\frac{1}{10000} \sum_{s=1}^{10000} I_{\{\lambda^{(s)} > 2\}} = 0.005003$$

dok je stvarna vrijednost

$$P(\lambda > 2|X_1 = x_1, \dots, X_n = x_n) = 0.00495458.$$

 Danger

75

Zadatak 75.1. Riješite zadatke 3.4 - 3.7 koristeći Monte Carlo aproksimacije uz prikladne usporedbe sa stvarnim vrijednostima.

76

Zadatak 76.1. U bazi podataka **midge** (iz paketa **Flury**) nalaze se podaci o duljini krila dviju vrsta mušica (*Amerohelea fasciata* (Af) i *Pseudofasciata* (Apf)). Pretpostavka je da mjerenja dolaze iz normalne distribucije. Prethodna istraživanja sugeriraju da je očekivana duljina krila za mušice 1.9mm uz standardnu devijaciju 0.1.

- Koji su parametri od interesa? Zapišite vjerodostojnost Bayesovog modela za pojedinu vrstu
- Odredite konjugiranu apriornu distribuciju i pripadnu posteriornu distribuciju za pojedine vrste.
- Aproksimirajte posteriornu očekivanu vrijednost duljine krila mušice za pojedinu vrstu korištenjem Monte Carlo aproksimacija.
- Nakon što uzmemo podatke u obzir, koliko iznosi aproksimacija vjerojatnosti da je očekivana duljina krila pojedine vrste mušice veća od 2 mm?
- Odredite posteriorne 0.05-vjerodostojne intervale za očekivanja pojedinih vrsta

```
install.packages("remotes")
library(remotes)
install_version("Flury", "0.1-3")
library(Flury)
```

 Danger

77 Rješenje

- a)
- $(\mu_1, \sigma_1^2) \in \mathbb{R} \times \langle 0, +\infty \rangle$, gdje su μ_1 i σ_1^2 redom očekivana duljina i varijanca krila mušice vrste Af
 - $(\mu_2, \sigma_2^2) \in \mathbb{R} \times \langle 0, +\infty \rangle$, gdje su μ_2 i σ_2^2 redom očekivana duljina i varijanca krila mušice vrste Af

Pripadne vjerodostojnosti:

$$f(x|\mu_1, \sigma_1^2) = \frac{1}{\sqrt{2\pi \cdot \sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2 \cdot \sigma_1^2}}, \quad f(y|\mu_2, \sigma_2^2) = \frac{1}{\sqrt{2\pi \cdot \sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2 \cdot \sigma_2^2}}.$$

b) Radi se o normalnom modelu uz oba parametra nepoznata.

Apriorne distribucije: S obzirom da nemamo nikakve dodatne informacije o (novim) vrstama mušica, za apriorne parametre koristimo zajedničke vrijednosti

$$\nu_0 = 1, k_0 = 1, \mu_0 = 1.9, \sigma_0^2 = 0.01$$

Pri tome treba napomenuti da μ_0 i σ_0^2 reflektiraju informacije o prijašnjim istraživanjima za mušice općenito, a kako se radi o novim mušicama, postavljamo apriorne dimenzije “uzoraka” $\nu_0 = k_0 = 1$.

Stoga je:

- $\sigma_1^2 \sim \text{InvGama}(\frac{1}{2}, \frac{1}{2} \cdot 0.01)$, $\mu_1 | \sigma_1^2 \sim \mathcal{N}(1.9, \frac{\sigma_1^2}{1})$
- $\sigma_2^2 \sim \text{InvGama}(\frac{1}{2}, \frac{1}{2} \cdot 0.01)$, $\mu_2 | \sigma_2^2 \sim \mathcal{N}(1.9, \frac{\sigma_2^2}{1})$.

Posteriorne distribucije: Znamo da je uz zadane konjugirane apriorne distribucije, posteriorna distribucija oblika:

- $\sigma_1^2 \sim \text{InvGama}(\frac{\nu_9}{2}, \frac{\nu_9}{2} \cdot \sigma_9^2)$, $\mu_1 | \sigma_1^2 \sim \mathcal{N}(\mu_9, \frac{\sigma_9^2}{\nu_9})$
- $\sigma_2^2 \sim \text{InvGama}(\frac{\nu_6}{2}, \frac{\nu_6}{2} \cdot \sigma_6^2)$, $\mu_2 | \sigma_2^2 \sim \mathcal{N}(\mu_6, \frac{\sigma_6^2}{\nu_6})$,

gdje su

- $\nu_9 = 10, k_9 = 10, \mu_9 = 1.814, \sigma_9^2 = \frac{1}{10} \left(1 \cdot 0.01 + (9-1)s_9^2 + \frac{1 \cdot 9}{10} (\bar{x}_9 - 1.9)^2 \right) = 0.006$
- $\nu_6 = 7, k_6 = 7, \mu_6 = 1.923, \sigma_6^2 = \frac{1}{7} \left(1 \cdot 0.01 + (6-1)s_6^2 + \frac{1 \cdot 6}{7} (\bar{y}_6 - 1.9)^2 \right) = 0.01$

```

data(midge)
attach(midge)
# vrsta Af
nu0.Af <- 1
k0.Af <- 1
mu0.Af <- 1.9
sigma0.Af <- 0.1
bar.Af <- mean(Wing.Length[Species=="Af"])
n.Af <- length(Wing.Length[Species=="Af"])
s2.Af <- var(Wing.Length[Species=="Af"])
kn.Af <- k0.Af + n.Af
nun.Af <- nu0.Af + n.Af
mun.Af <- (k0.Af * mu0.Af + n.Af * bar.Af) / kn.Af
sigman.Af <- sqrt((nu0.Af * sigma0.Af^2 + (n.Af - 1) * s2.Af + k0.Af * n.Af * (bar.Af - mu0.Af) / kn.Af) / n.Af)

mun.Af

```

[1] 1.814

```
sigman.Af^2
```

[1] 0.005902222

```

# vrsta Apf
nu0.Apf <- 1
k0.Apf <- 1
mu0.Apf <- 1.9
sigma0.Apf <- 0.1
bar.Apf <- mean(Wing.Length[Species=="Apf"])
n.Apf <- length(Wing.Length[Species=="Apf"])
s2.Apf <- var(Wing.Length[Species=="Apf"])
kn.Apf <- k0.Apf + n.Apf
nun.Apf <- nu0.Apf + n.Apf
mun.Apf <- (k0.Apf * mu0.Apf + n.Apf * bar.Apf) / kn.Apf
sigman.Apf <- sqrt((nu0.Apf * sigma0.Apf^2 + (n.Apf - 1) * s2.Apf + k0.Apf * n.Apf * (bar.Apf - mu0.Apf) / kn.Apf) / n.Apf)

mun.Apf

```

[1] 1.922857

```
sigman.Apf^2
```

```
[1] 0.01025578
```

c) Primijetimo da nas zanima Monte Carlo aproksimacija $E[\mu_1|X_1 = x_1, \dots, X_9 = x_9]$ i $E[\mu_2|Y_1 = y_1, \dots, Y_6 = x_6]$.

```
#vrsta Af
sigma2.mc.Af <- 1/rgamma(10000, nun.Af, nun.Af*sigman.Af^2/2)
mu.mc.Af <- rnorm(10000, mun.Af, sqrt(sigma2.mc.Af/kn.Af))
mean(mu.mc.Af)
```

```
[1] 1.81379
```

```
#vrsta Apf
sigma2.mc.Apf <- 1/rgamma(10000, nun.Apf, nun.Apf*sigman.Apf^2/2)
mu.mc.Apf <- rnorm(10000, mun.Apf, sqrt(sigma2.mc.Apf/kn.Apf))
mean(mu.mc.Apf)
```

```
[1] 1.922698
```

Dakle, Monte Carlo aproksimacije očekivanih duljina krila su:

$$E[\mu_1|X_1 = x_1, \dots, X_9 = x_9] \approx 1.814, \quad E[\mu_2|Y_1 = y_1, \dots, Y_6 = x_6] \approx 1.923$$

S druge strane, Monte Carlo aproksimacije vjerojatnosti $P(\mu_1 > 2|\{X = x\})$ i $P(\mu_2 > 2|\{Y = y\})$ lako izračunamo sljedećim kodom:

```
mean(mu.mc.Af >2)
```

```
[1] 0
```

```
mean(mu.mc.Apf >2)
```

```
[1] 0.008
```

e)

```
quantile(mu.mc.Af, c(0.025, 0.975))
```

```
      2.5%    97.5%
1.778559 1.849136
```

```
quantile(mu.mc.Apf, c(0.025, 0.975))
```

```
      2.5%    97.5%
1.865124 1.981767
```

Zaključujemo:

$$P(\mu_1 \in \langle 1.78, 1.85 \rangle | X_1 = x_1, \dots, X_9 = x_9) \approx 0.95$$

$$P(\mu_2 \in \langle 1.863, 1.981 \rangle | Y_1 = x_1, \dots, Y_6 = y_6) \approx 0.95$$

 Danger

78

Zadatak 78.1. Zadan je gama-eksponencijalni model uz hiperparametre $\alpha = 1, \beta = 2$ i podatke $\mathbf{x} = (0.7, 1.2, 1.05, 0.9, 0.8, 1.1)$.

- Napravite MCMC aproksimaciju posteriorne distribucije parametra modela i grafički usporedite sa stvarnom distribucijom.
- Napravite MCMC aproksimaciju posteriorne prediktivne distribucije modela i grafički usporedite sa stvarnom distribucijom.

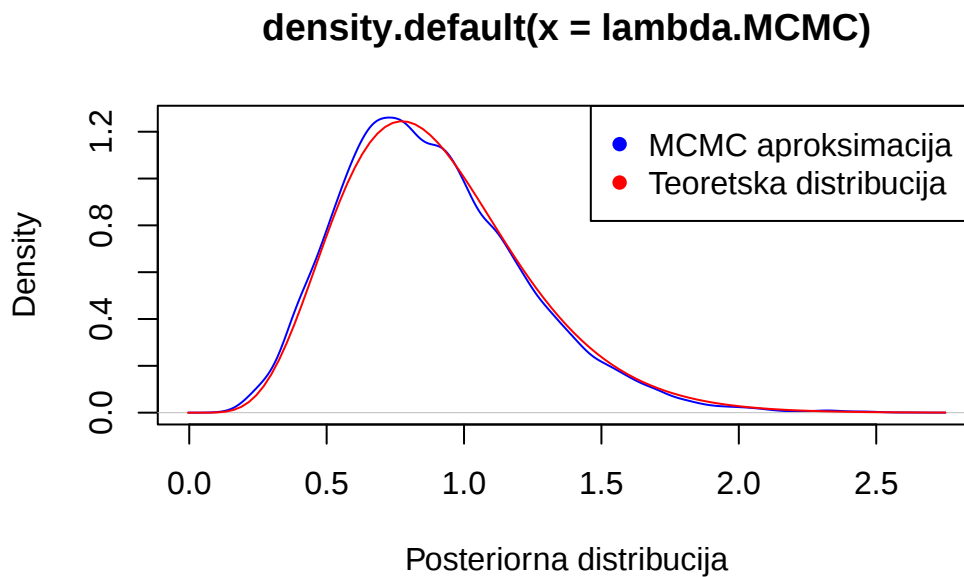
 Danger

79 Rješenje

- a) Kako se radi o jednodimenzionalnom parametru provest ćemo Metropolis-Hastings algoritam za generiranje vrijednosti iz posteriorne distribucije uz simetričnu distribuciju kandidata (specijalno, radi se o Metropolis algoritmu).

```
S <- 20000
x <- c(0.7, 1.2, 1.05, 0.9, 0.8, 1.1)
alpha<-1
beta<-2
alphan <- alpha+length(x)
betan <-beta+sum(x)
delta <-0.2
lambda <- 0.9
lambda.MCMC <- NULL
for(s in 1:S)
{lambda.star<-runif(1,lambda-delta, lambda+delta)
log.r <- (sum(dexp(x,lambda.star, log=TRUE))+dgamma(lambda.star,alpha, beta, log=TRUE))-
  (sum(dexp(x,lambda, log=TRUE))+dgamma(lambda,alpha,beta, log=TRUE))
if (log(runif(1)) < log.r) {lambda <-lambda.star}
lambda.MCMC <-c(lambda.MCMC, lambda)
}

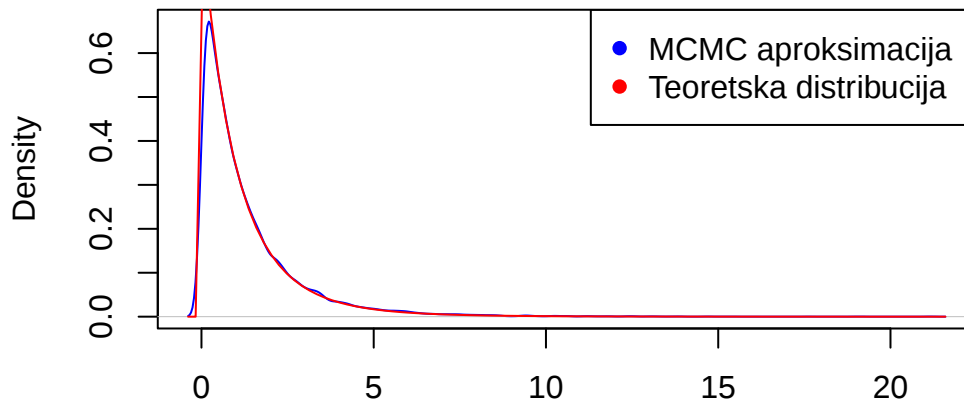
plot(density(lambda.MCMC), col="blue", xlab="Posteriorna distribucija")
curve(dgamma(x,alphan, betan), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c("blue", "red"))
```



b) Koristeći MCMC generirane vrijednosti za parametar λ možemo generirati nove “podatke”, odnosno predikcije iz vjerodostojnosti:

```
x.pred.MCMC <- rexp(S,lambda.MCMC)
plot(density(x.pred.MCMC), col="blue", xlab="Prediktivna posteriorna distribucija")
curve(dlomax(x,1/(betan),alphan), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```

density.default(x = x.pred.MCMC)



Prediktivna posteriorna distribucija

 Danger

80

Zadatak 80.1. Zadan je beta-binomni model (broj izvođenja pokusa $m = 10$) uz hiperparametre $\alpha = 2, \beta = 2$ i podatke $\mathbf{x} = (3, 7, 5, 4, 7, 9, 4, 5, 2)$.

- Napravite MCMC aproksimaciju posteriorne distribucije parametra modela i grafički usporedite sa stvarnom distribucijom.
- Napravite MCMC aproksimaciju posteriorne prediktivne distribucije modela i grafički usporedite sa stvarnom distribucijom.
- Napravite MCMC aproksimaciju posteriorne distribucije logaritma šansi parametra.

 Danger



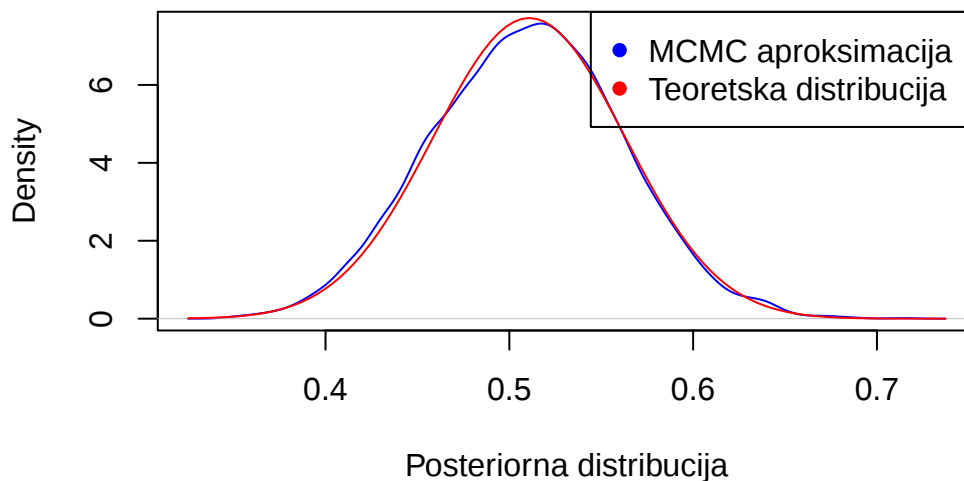
81 Rješenje

- a) Kako se radi o jednodimenzionalnom parametru provest ćemo Metropolis-Hastings algoritam za generiranje vrijednosti iz posteriorne distribucije uz simetričnu distribuciju kandidata (specijalno, radi se o Metropolis algoritmu).

```
S <- 20000
x <- c(3,7,5,4,7,9,4,5,2)
alpha<-2
beta<-2
m<-10
alphan <- alpha+sum(x)
betan <-beta+m*length(x)-sum(x)
delta <-0.05
theta <- 0.6
theta.MCMC <- NULL
for(s in 1:S)
{theta.star<-runif(1,theta-delta, theta+delta)
log.r <- (sum(dbinom(x,m,theta.star, log=TRUE))+dbeta(theta.star,alpha, beta, log=TRUE))
  (sum(dbinom(x,m,theta, log=TRUE))+dbeta(theta,alpha,beta, log=TRUE))
if (log(runif(1)) < log.r) {theta <-theta.star}
theta.MCMC <-c(theta.MCMC, theta)
}

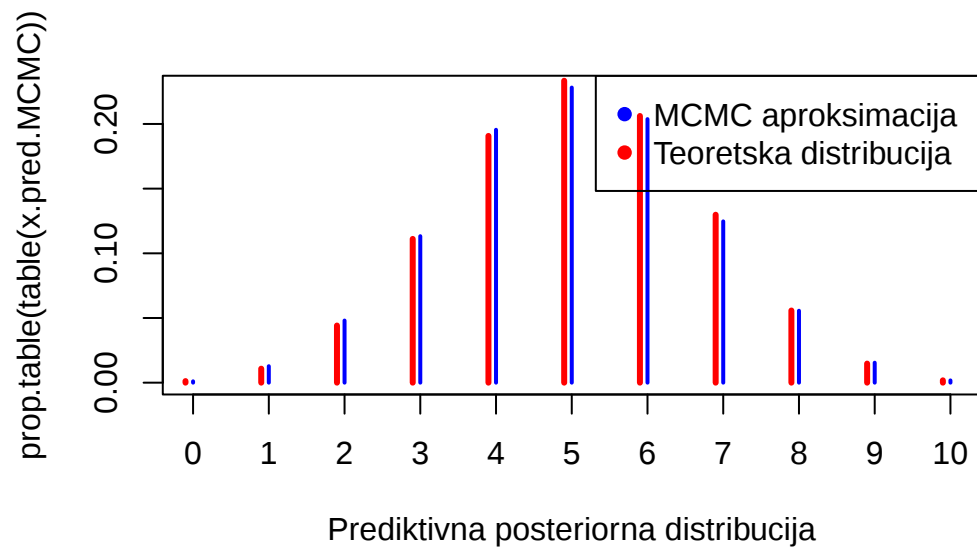
plot(density(theta.MCMC), col="blue", xlab="Posteriorna distribucija")
curve(dbeta(x,alphan, betan), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```

density.default(x = theta.MCMC)



b) Koristeći MCMC generirane vrijednosti za parametar λ možemo generirati nove “podatke”, odnosno predikcije iz vjerodostojnosti:

```
x.pred.MCMC <- rbinom(S,m,theta.MCMC)
plot(prop.table(table(x.pred.MCMC)), col="blue", xlab="Prediktivna posteriorna distribucija",
points(c(0:10)-0.1, dbinom(c(0:10), m,alphan,betan), col="red",type="h",
      ylab=expression(paste(italic("p("),y[n+1], "|",y[1], "...",y[n],")", sep="")),
      xlab=expression(italic(y[n+1])),ylim=c(0,.5),lwd=3)
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c("blue", "red"))
```



82

Zadatak 82.1. Neka je zadan miješani model tri normalne distribucije

$$\begin{aligned}\theta|\delta &\sim \mathcal{N}(\mu_\delta, \sigma_\delta^2) \\ \delta &\sim \begin{pmatrix} 1 & 2 & 3 \\ 0.45 & 0.1 & 0.45 \end{pmatrix} \\ (\mu_1, \mu_2, \mu_3) &= (-3, 0, 3), \quad (\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1/3, 1/3, 1/3).\end{aligned}$$

- Grafički prikažite gustoću modela.
- Generirajte uzorak dimenzije $S = 2000$ za (θ, δ) korištenjem Monte Carlo (MC) metode. Na temelju MC aproksimacije grafički usporedite marginalnu teorijsku distribuciju za θ sa pripadnom aproksimacijom.
- Generirajte uzorak dimenzije $S = 10000$ za (θ, δ) korištenjem metode Monte Carlo Markovljevih lanaca (MCMC metode) uz početnu vrijednost $\theta = 0$. Na temelju MCMC aproksimacije grafički usporedite marginalnu teorijsku distribuciju za θ sa pripadnom aproksimacijom.
- Koja aproksimacija daje bolje rezultate. Zašto?
- Generirajte novu MCMC aproksimaciju s početnom vrijednosti $\theta = -6$.
- Odredite efektivnu veličinu uzorka za MCMC metodu te autokorelaciju dobivenog uzorka za θ na lagu 10 i 50. Komentirajte.

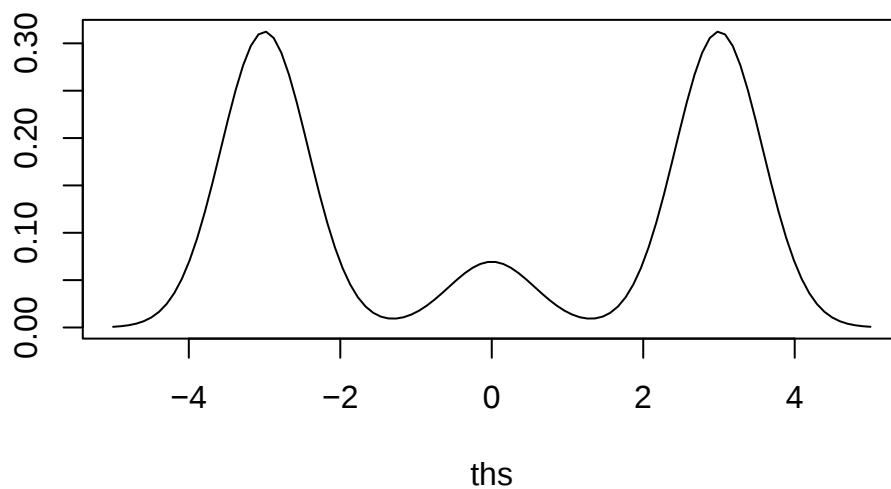
 Danger

83 Rješenje

a)

```
mu<-c(-3,0,3)
s2<-c(.33,.33,.33)
w<-c(.45,.1,.45)

ths<-seq(-5,5,length=100)
plot(ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])) ,type="l", ylab="" )
```



b) Uzorkovanje MC metodom podrazumijeva da znamo generirati vrijednosti izravno iz tražene distribucije $f(\theta, \delta)$. Ovdje to možemo (neposredno) napraviti jer je

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

```

S<-2000
d<-sample(1:3,S, prob=w,replace=TRUE)
th<-rnorm(S,mu[d],sqrt(s2[d]))
THD.MC<-cbind(th,d)

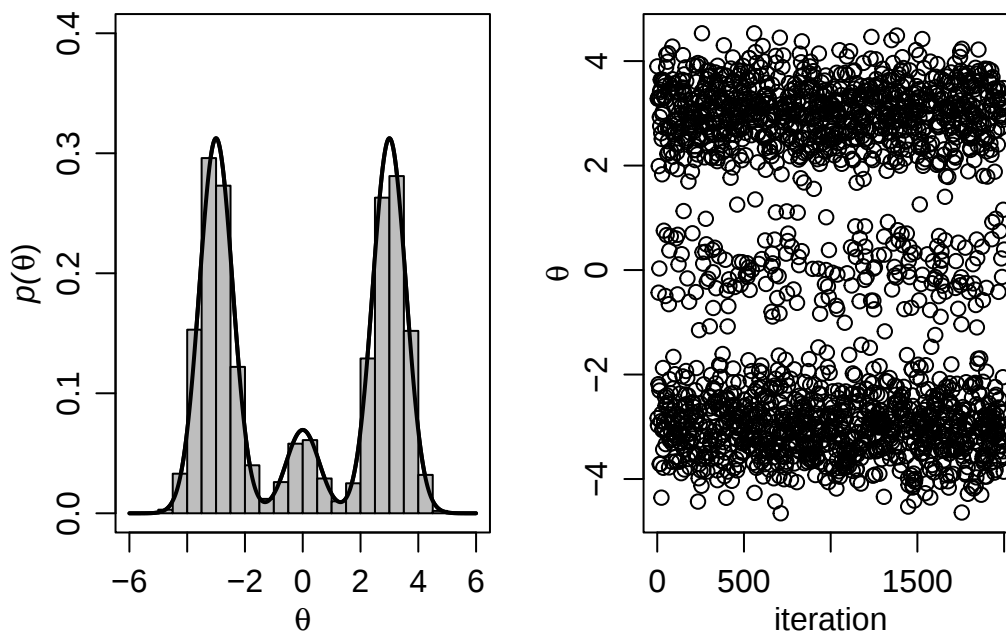
```

Grafički prikaz:

```

par(mfrow=c(1,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0))
ths<-seq(-6,6,length=1000)
plot(ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])) ,type="l" ,
      xlab=expression(theta),ylab=
        expression( paste( italic("p("),theta,")",sep="" ) ),lwd=2 ,ylim=c(0,.40))
hist(THD.MC[,1],add=TRUE,prob=TRUE,nclass=20,col="gray")
lines( ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
        w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
        w[3]*dnorm(ths,mu[3],sqrt(s2[3])),lwd=2 )
plot(THD.MC[1:2000,1],xlab="iteration",ylab=expression(theta))

```



c) MCMC metoda (Gibbsovo uzorkovanje) podrazumijeva generiranje vrijednosti iz uvjetnih distribucija parametra, tj. najprije generiramo iz

$$f(\delta|\theta) = \frac{f(\theta|\delta) \cdot f(\delta)}{\sum_{i=1}^3 f(\theta|\delta_i) \cdot f(\delta_i)}$$

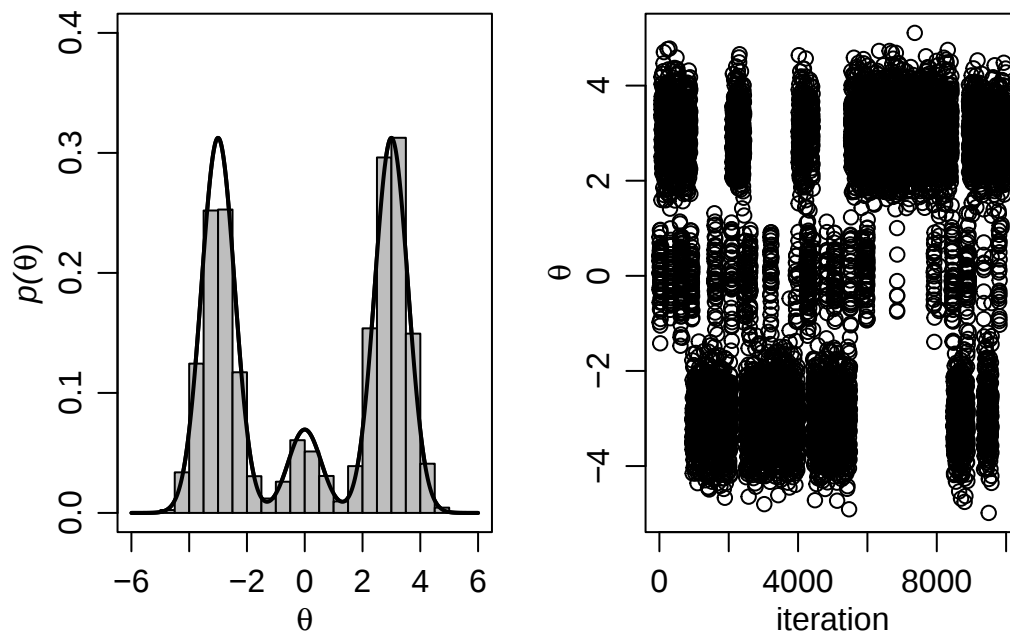
a zatim korištenjem simuliranog δ generiramo θ iz uvjetne distribucije

$$f(\theta|\delta).$$

```
th<-0
THD.MCMC<-NULL
S<-10000
for(s in 1:S) {
  d<-sample(1:3,1,prob= w*dnorm(th,mu,sqrt(s2)))
  th<-rnorm(1,mu[d],sqrt(s2[d]))
  THD.MCMC<-rbind(THD.MCMC,c(th,d))
}
```

Grafički prikaz:

```
par(mfrow=c(1,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0))
Smax<-1000
ths<-seq(-6,6,length=1000)
plot(ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])),type="l", xlab=expression(theta),
      ylab=expression( paste( italic("p("),theta,")",sep="" ) ,lwd=2 ,ylim=c(0,.40))
hist(THD.MCMC[1:10000,1],add=TRUE,prob=TRUE,nclass=20,col="gray")
lines( ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])),lwd=2 )
plot(THD.MCMC[1:10000,1],xlab="iteration",ylab=expression(theta))
```



- d) MC metoda daje bolje rezultate u odnosu na MCMC metodu (Gibbsovo uzorkovanje) unatoč tome što je MC metoda bazirana na $S = 2000$, a MCMC metoda na $S = 10000$ vrijednosti. Naime, uzorak dobiven MCMC metodom je snažno koreliran niz vrijednosti (bazira se na Markovljevom lancu), dok uzorak generiran standardnom MC metodom nema strukturu zavisnosti, tj. sve vrijednosti

$$(\theta^{(1)}, \delta^{(1)}), \dots, (\theta^{(S)}, \delta^{(S)})$$

su nezavisno generirane .

e)

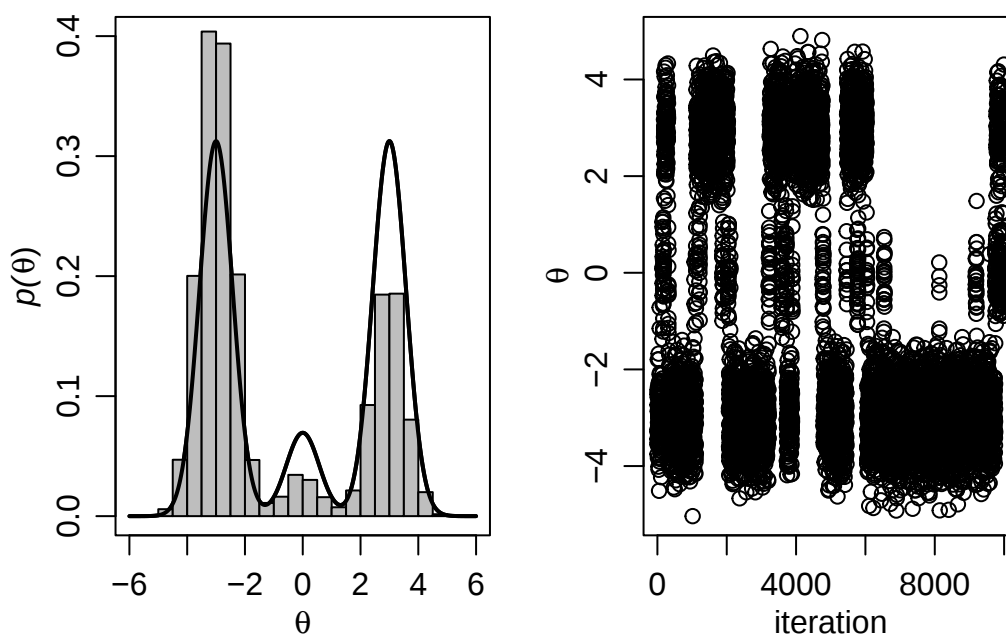
```
th<--6
THD.MCMC<-NULL
S<-10000
for(s in 1:S) {
  d<-sample(1:3,1,prob=w*dnorm(th,mu,sqrt(s2)) )
  th<-rnorm(1,mu[d],sqrt(s2[d]) )
  THD.MCMC<-rbind(THD.MCMC,c(th,d) )
}
```

Grafički prikaz:

```

par(mfrow=c(1,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0))
Smax<-1000
ths<-seq(-6,6,length=1000)
plot(ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])) ,type="l" , xlab=expression(theta),
      ylab=expression( paste( italic("p("),theta,")",sep="" ) ,lwd=2 ,ylim=c(0,.40))
hist(THD.MCMC[1:10000,1],add=TRUE,prob=TRUE,nclass=20,col="gray")
lines( ths, w[1]*dnorm(ths,mu[1],sqrt(s2[1])) +
      w[2]*dnorm(ths,mu[2],sqrt(s2[2])) +
      w[3]*dnorm(ths,mu[3],sqrt(s2[3])),lwd=2 )
plot(THD.MCMC[1:10000,1],xlab="iteration",ylab=expression(theta))

```



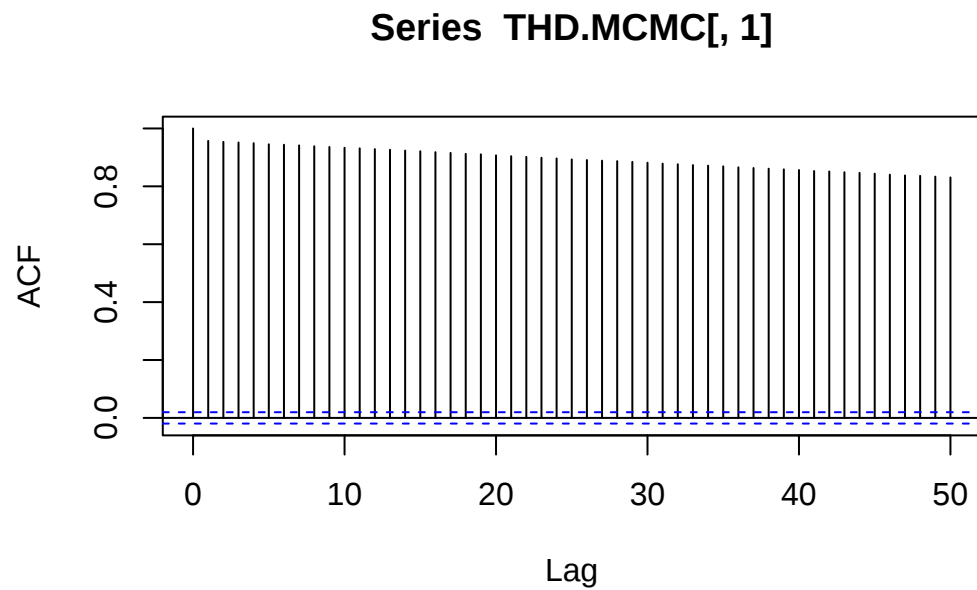
Možemo primijetiti da je odabir početne vrijednosti $\theta = -6$ posljedično značio veću zastupljenost “lijeve” regije te je usporilo konvergenciju Markovljevog lanca. Idealno bi bilo započinjati Gibbsovo uzorkovanje u području s velikom vjerojatnošću (iako to u praksi ne znamo, možemo samo naslutiti).

f) Efektivna veličina uzorka iznosi

```
effectiveSize(THD.MCMC[,1])
```

```
var1  
14.92217
```

```
acf(THD.MCMC[,1], lag=50)[c(10,50)]
```



Autocorrelations of series 'THD.MCMC[, 1]', by lag

	10	50
	0.933	0.831

dok je autokorelacija na lagu 10 0.933, a na lagu 50 0.831. Dakle, kvaliteta od MCMC aproksimacije dimenzije $S = 10000$ ekvivalentna je aproksimaciji dobivenoj MC metodom dimenzije $S = 15$. Uočavamo da se radi o značajnoj korelaciji u lancu, što sugerira sporu konvergenciju lanca prema traženoj distribuciji.

 Danger

84

Zadatak 84.1. Riješite zadatke 5.4 i 5.5 MCMC metodom koristeći STAN platformu.

 Danger

85 Rješenje

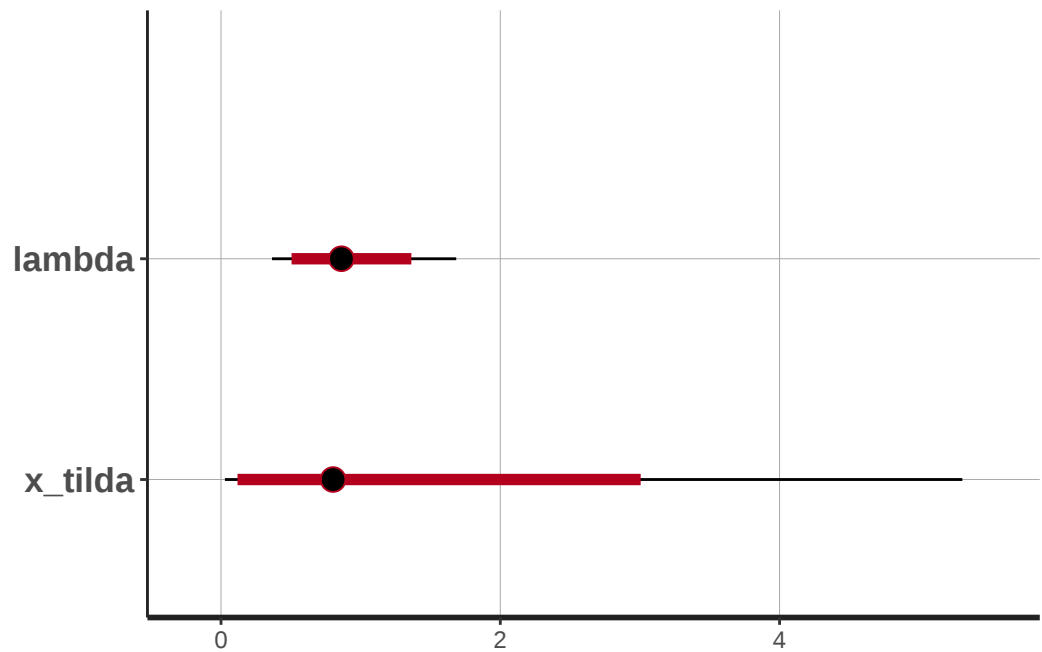
- Zadatak 5.4

-a)

```
modelGamaExp <- "  
data {  
  real <lower = 0> x[6];          // podaci  
}  
parameters {  
  real<lower = 0> lambda;        // lambda  
}  
model {  
  x ~ exponential(lambda); // vjerodostojnost  
  lambda ~ gamma(1, 2); // apriorna distribucija  
}  
  
generated quantities {  
  real<lower = 0> x_tilda = exponential_rng(lambda);  
}  
"  
podaci <- list(x=c(0.7, 1.2, 1.05, 0.9, 0.8, 1.1))  
sim.MCMC.stan <-stan(  
  model_code = modelGamaExp,  
  data = podaci,      # named list of data  
  chains = 4,          # broj korištenih Markovljevih lanaca  
  warmup = 10000,      # burn-in period  
  iter = 20000,        # ukupan broj iteracija po Markovljevom lancu  
  cores = 1,           # broj jezgri  
  refresh = 0,         # bez prikazivanja napretka  
  seed=934  
)  
  
plot(sim.MCMC.stan) #crta vjerodostojne intervale
```

ci_level: 0.8 (80% intervals)

outer_level: 0.95 (95% intervals)



```
print(sim.MCMC.stan) #informacije o MCMC aproksimaciji
```

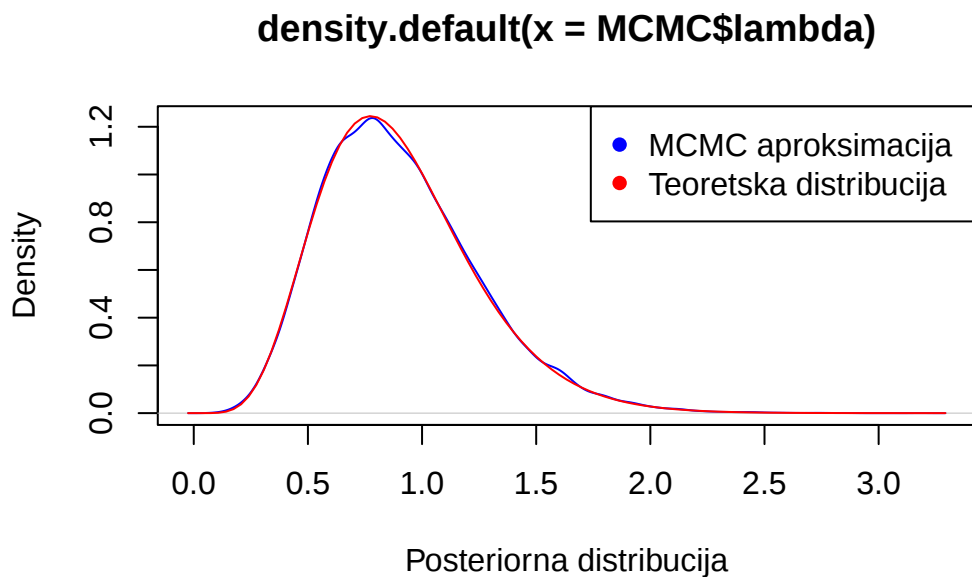
Inference for Stan model: anon_model.
4 chains, each with iter=20000; warmup=10000; thin=1;
post-warmup draws per chain=10000, total post-warmup draws=40000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda	0.91	0.00	0.34	0.37	0.65	0.86	1.11	1.68	13683	1
x_tilda	1.29	0.01	1.54	0.03	0.32	0.80	1.70	5.31	29310	1
lp__	-8.23	0.01	0.72	-10.28	-8.39	-7.95	-7.77	-7.71	14057	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:39:06 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```
MCMC <- rstan::extract(sim.MCMC.stan, permuted = TRUE) #dohvacanje simuliranih vrijednos
#parametra
```

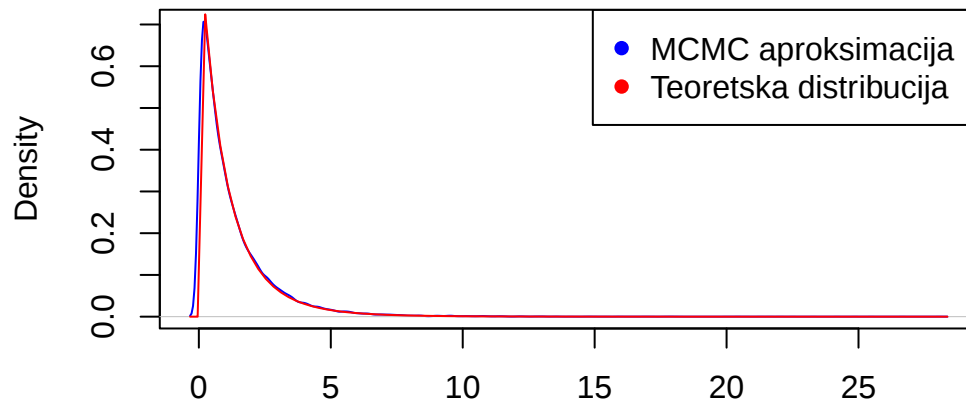
```
plot(density(MCMC$lambda), col="blue", xlab="Posteriorna distribucija")
curve(dgamma(x,7, 7.75), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```



-b)

```
plot(density(MCMC$x_tilda), col="blue", xlab="Prediktivna posteriorna distribucija")
curve(dlomax(x,1/(7.5),7), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```

density.default(x = MCMC\$x_tilda)



Prediktivna posteriorna distribucija

- Zadatak 5.5
-a)

```

modelBetaBin <- "
data {
  int <lower = 0, upper=10> x[9];          // podaci
}
parameters {
  real<lower = 0, upper=1> theta;          // parametar theta
}
model {
  x ~ binomial(10,theta); // vjerodostojnost
  theta ~ beta(2, 2); // apriorna distribucija
}

generated quantities {
  int<lower = 0, upper=10> x_tilda = binomial_rng(10, theta);
}
"

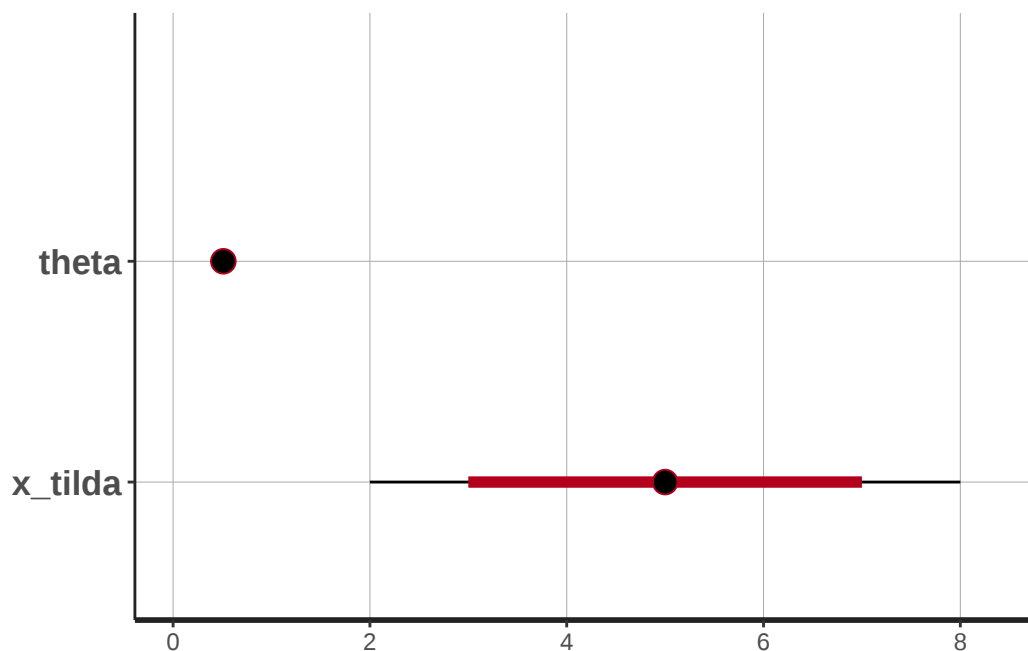
podaci <- list(x=c(3,7,5,4,7,9,4,5,2))
sim.MCMC.stan <-stan(
  model_code = modelBetaBin,
  data = podaci,      # named list of data
  chains = 4,         # broj korištenih Markovljevih lanaca
  warmup = 10000,     # burn-in period
  iter = 20000,       # ukupan broj iteracija po Markovljevom lancu
  cores = 1,          # broj jezgri
  refresh = 0,        # bez prikazivanja napretka
  seed=934
)

plot(sim.MCMC.stan) #crta vjerodostojne intervale

```

ci_level: 0.8 (80% intervals)

outer_level: 0.95 (95% intervals)



```
print(sim.MCMC.stan) #informacije o MCMC aproksimaciji
```

Inference for Stan model: anon_model.

4 chains, each with iter=20000; warmup=10000; thin=1;

post-warmup draws per chain=10000, total post-warmup draws=40000.

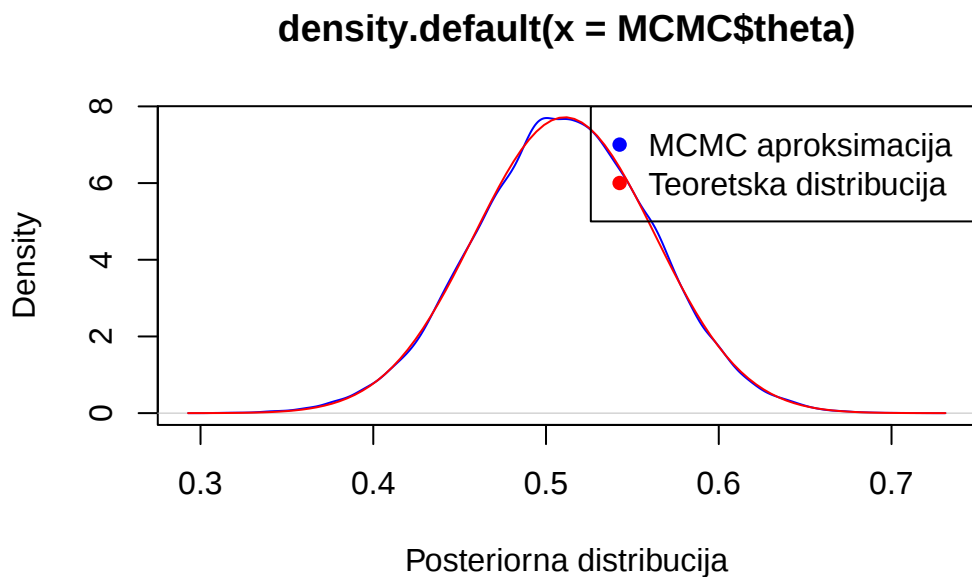
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta	0.51	0.00	0.05	0.41	0.48	0.51	0.55	0.61	14968	1
x_tilda	5.11	0.01	1.65	2.00	4.00	5.00	6.00	8.00	35039	1
lp__	-65.64	0.01	0.72	-67.70	-65.79	-65.36	-65.18	-65.13	16408	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:39:50 2024.

For each parameter, `n_eff` is a crude measure of effective sample size, and `Rhat` is the potential scale reduction factor on split chains (at convergence, `Rhat`=1).

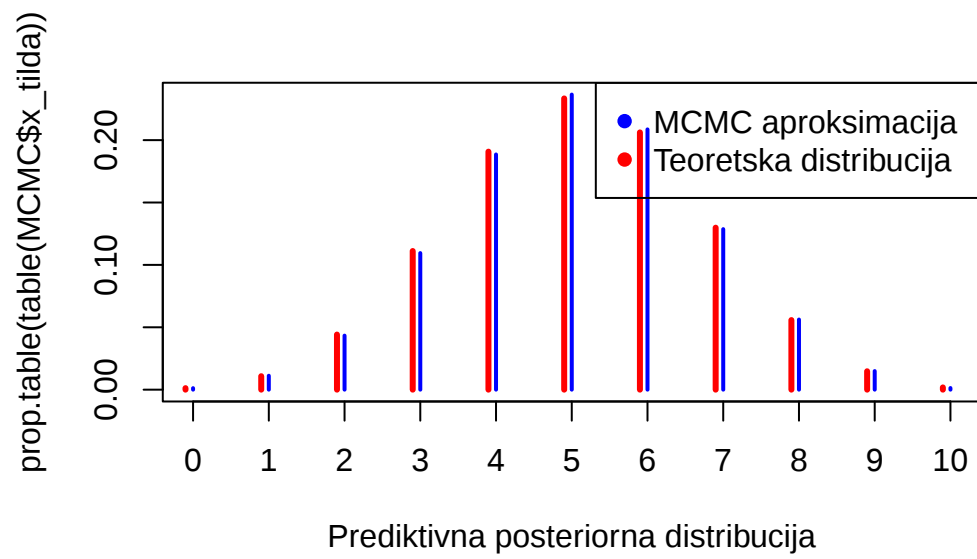
```
MCMC <- rstan::extract(sim.MCMC.stan, permuted = TRUE) #dohvacanje simuliranih vrijednos
#parametra
```

```
plot(density(MCMC$theta), col="blue", xlab="Posteriorna distribucija")
curve(dbeta(x,48, 46), add=T, col="red")
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```



-b)

```
plot(prop.table(table(MCMC$x_tilda)), col="blue", xlab="Prediktivna posteriorna distribu
points(c(0:10)-0.1, dbinom(c(0:10), 10, 48, 46), col="red" ,type="h",
      ylab=expression(paste(italic("p("),y[n+1],"|",y[1],"...",y[n],")"),sep="")),
      xlab=expression(italic(y[n+1])),ylim=c(0,.5),lwd=3)
legend("topright", legend = c("MCMC aproksimacija ", "Teoretska distribucija"), col = c(
```



86

Zadatak 86.1. U bazi podataka **midge** (iz paketa **Flury**) nalaze se podaci o duljini krila dviju vrsta mušica (*Amerohalea fasciata* (Af) i *Pseudofasciata* (Apf)). Pretpostavka je da mjerenja dolaze iz normalne distribucije. Prethodna istraživanja sugeriraju da je očekivana duljina krila za mušice 1.9mm uz standardnu devijaciju 0.1

- a) Napravite aproksimaciju posteriorne distribucije parametara za mušicu vrste Apf MCMC metodom koristeći STAN platformu. Pri tome koristite 4 Markovljeva lanca dimenzije 20000, pri čemu polovicu simulacija koristite kao “burn-in” period.
- b) Analizirajte kvalitetu dobivene aproksimacije, odnosno je li korištena metoda konvergirala.
- c) Što možete reći o brzini miješanja dobivene aproksimacije?
- d) Koliko iznosi efektivna uzoračka veličina? Interpretirajte.
- e) Odredite posteriorne 0.05-vjerodostojne intervale za očekivanja i varijance.
- f) Koliko iznosi vjerojatnost da slučajno odabrana mušica ove vrste ima duljinu krila između 1.8 mm i 2 mm ?
- g) Ukoliko zanemarimo prethodna istraživanja kakve rezultate i procjene dobivamo?

 Danger

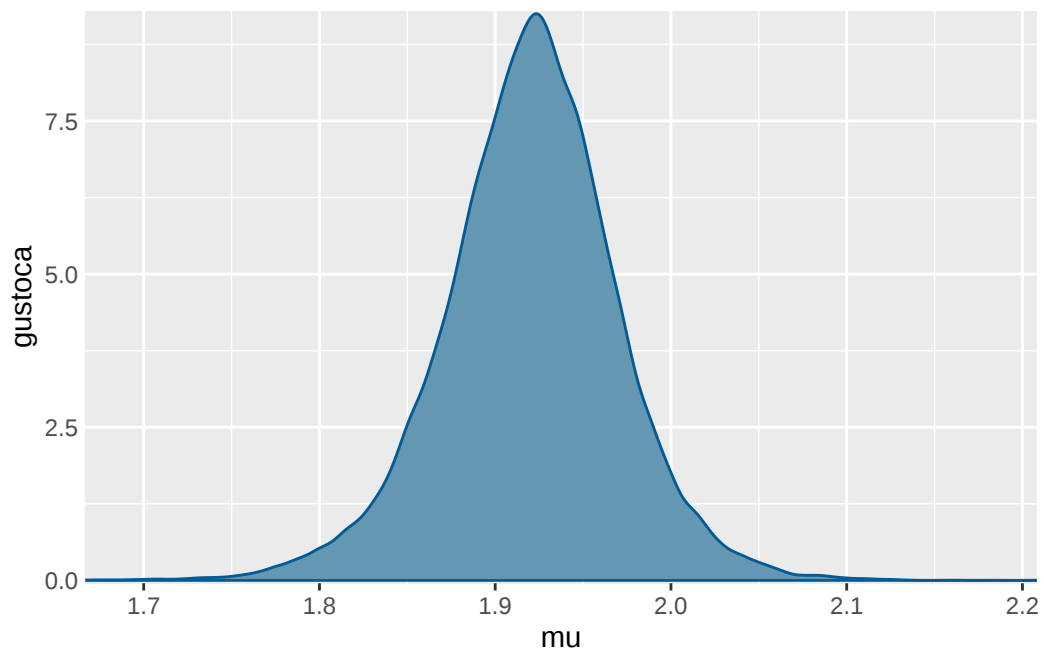
87 Rješenje

a) Model i implementacija MCMC metode korištenjem STAN platforme:

```
modelMusiceApf <- "  
data {  
  real <lower = 0> x[6];          // podaci  
}  
parameters {  
  real mu;          // parametar mu  
  real<lower = 0> sigma2; // parametar sigma2  
}  
model {  
  x ~ normal(mu,sqrt(sigma2)); // vjerodostojnost  
  mu ~ normal(1.9,0.1); //apriorna distribucija  
}  
  
generated quantities {  
  real x_tilda = normal_rng(mu,sqrt(sigma2));  
}"  
sim.MCMC.stan.Apf <-stan(  
  model_code = modelMusiceApf,  
  data = list(x=midge$Wing.Length[Species=="Apf"]), # named list of data  
  chains = 4,          # broj korištenih Markovljevih lanaca  
  warmup = 10000,      # burn in period za svaki lanac  
  iter = 20000,        # ukupan broj iteracija svakog lanca  
  cores = 1,          # broj korištenih jezgri  
  refresh = 0,        # no progress show  
  seed=934  
)
```

Aproksimacija posteriorne distribucije za μ :

```
mcmc_dens(sim.MCMC.stan.Apf, pars = "mu") +  
  yaxis_text(TRUE) +  
  ylab("gustoća")
```

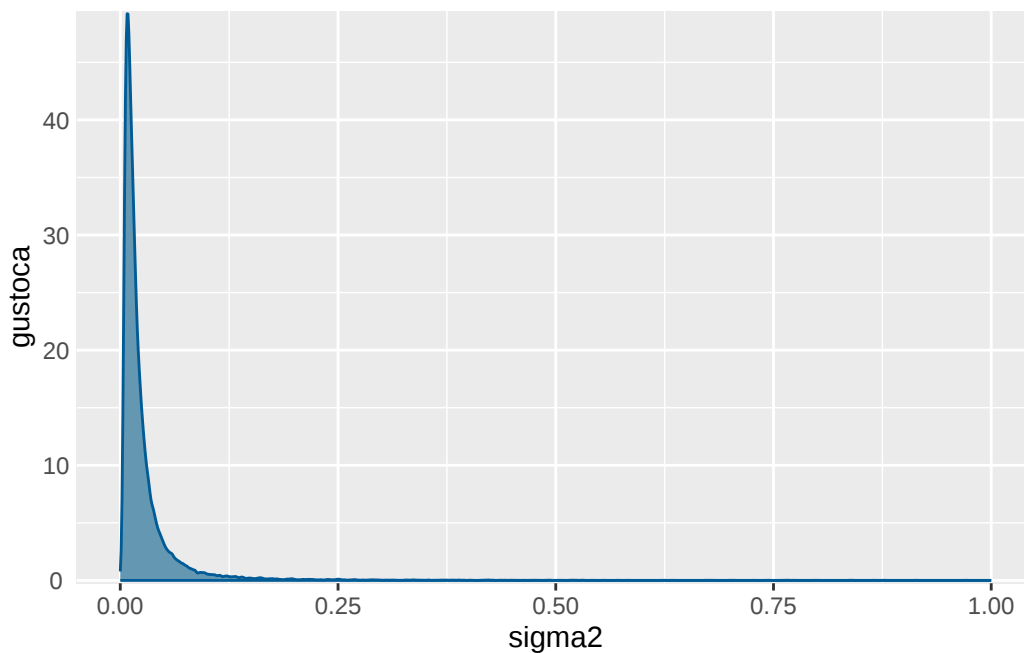


Aproksimacija posteriorne distribucije za σ^2 :

```
mcmc_dens(sim.MCMC.stan.Apriori, pars = "sigma2") +  
  yaxis_text(TRUE) + xlim(0,1)+  
  ylab("gustoća")
```

Scale for x is already present.

Adding another scale for x, which will replace the existing scale.



b)

Sljedećom naredbom dobivamo pregled dobivene posteriorne aproksimacije (prosjeak, standardna greška, standardna devijacija, kvantili, efektivna veličina uzorka, $\hat{R}$ - u slučaju konvergenције treba biti 1).

```
print(sim.MCMC.stan.Apf)
```

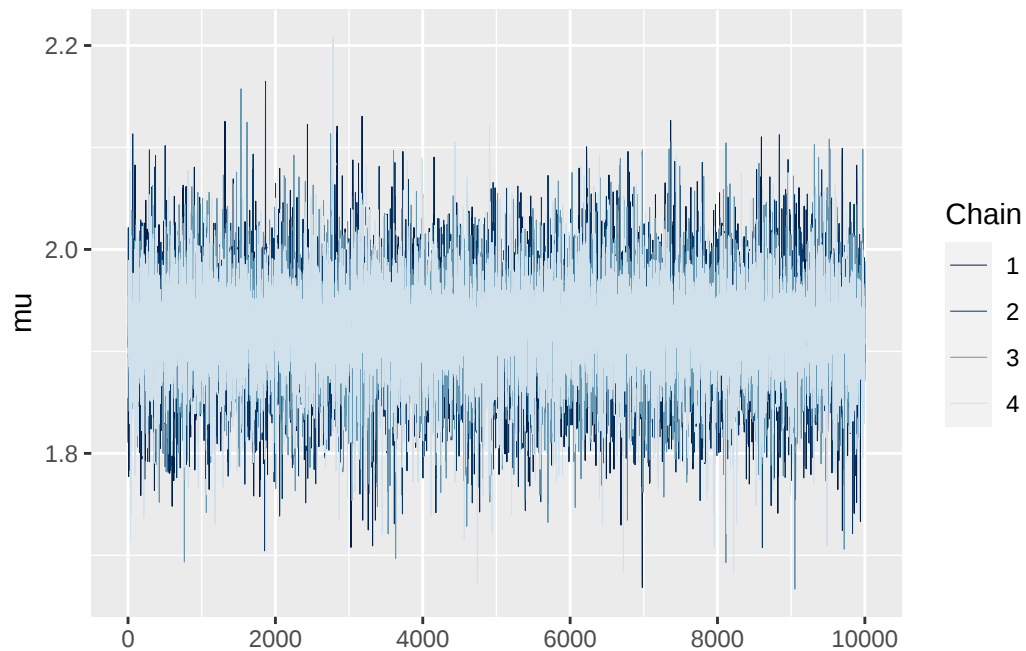
```
Inference for Stan model: anon_model.
4 chains, each with iter=20000; warmup=10000; thin=1;
post-warmup draws per chain=10000, total post-warmup draws=40000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu	1.92	0.00	0.05	1.82	1.89	1.92	1.95	2.02	16453	1
sigma2	0.03	0.00	0.06	0.00	0.01	0.01	0.03	0.12	15596	1
x_tilda	1.92	0.00	0.17	1.58	1.83	1.92	2.01	2.26	36346	1
lp__	6.07	0.01	1.12	3.07	5.61	6.41	6.89	7.20	11175	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:40:35 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

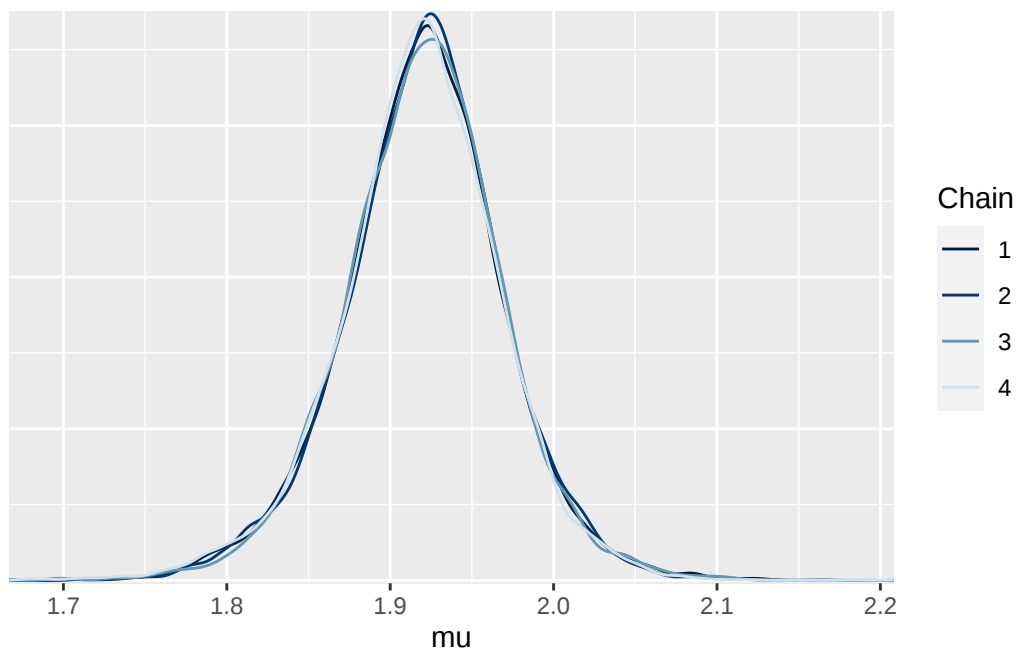
Usporedba trajektorija generiranih vrijednosti korištenjem četiriju lanaca:

```
mcmc_trace(sim.MCMC.stan.Apf, pars = "mu", size = 0.5)
```

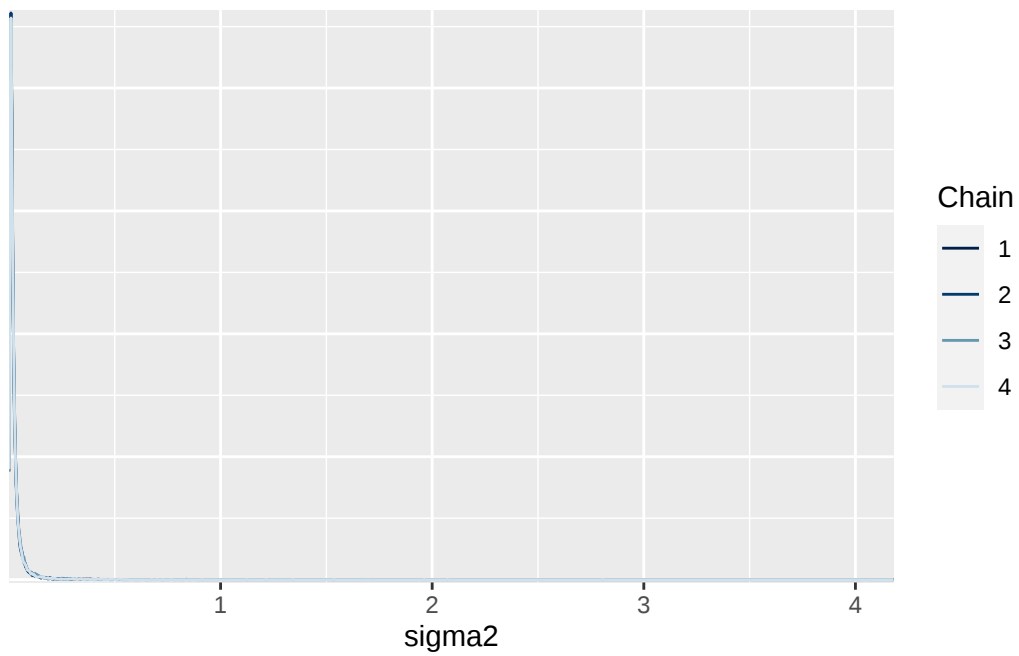


Ne uočavamo nikakve trendove niti strukturu u trajektorijama, sugerira konvergenciju.
Usporedba posteriornih distribucija četiriju lanaca:

```
mcmc_dens_overlay(sim.MCMC.stan.Apf , pars = "mu")
```



```
mcmc_dens_overlay(sim.MCMC.stan.Apf , pars = "sigma2")
```

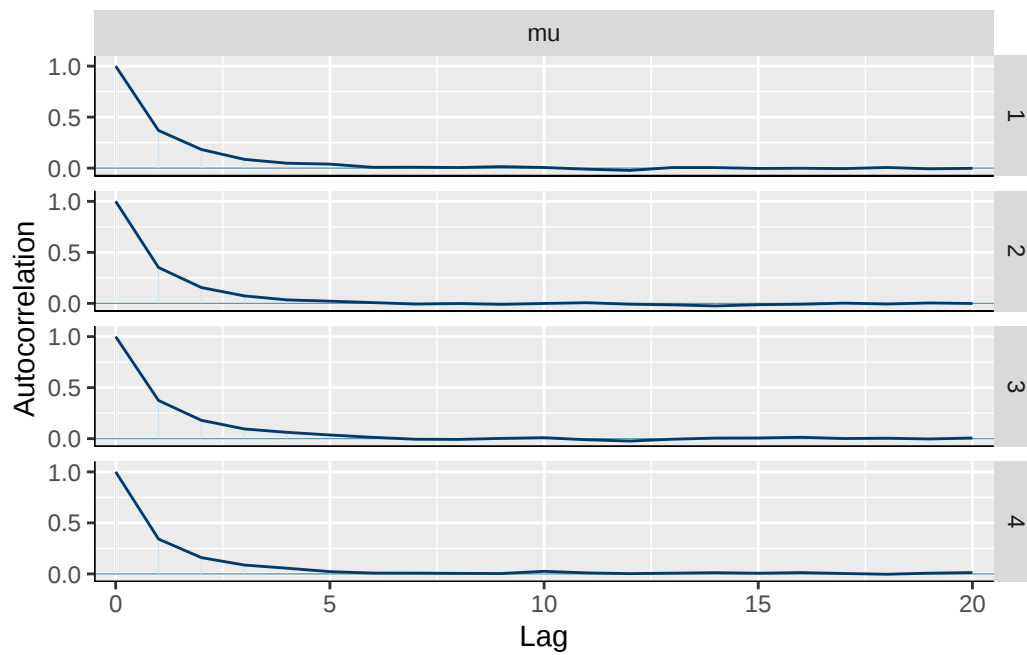


Vidimo da su ujednačene što sugerira da je došlo do konvergencije.

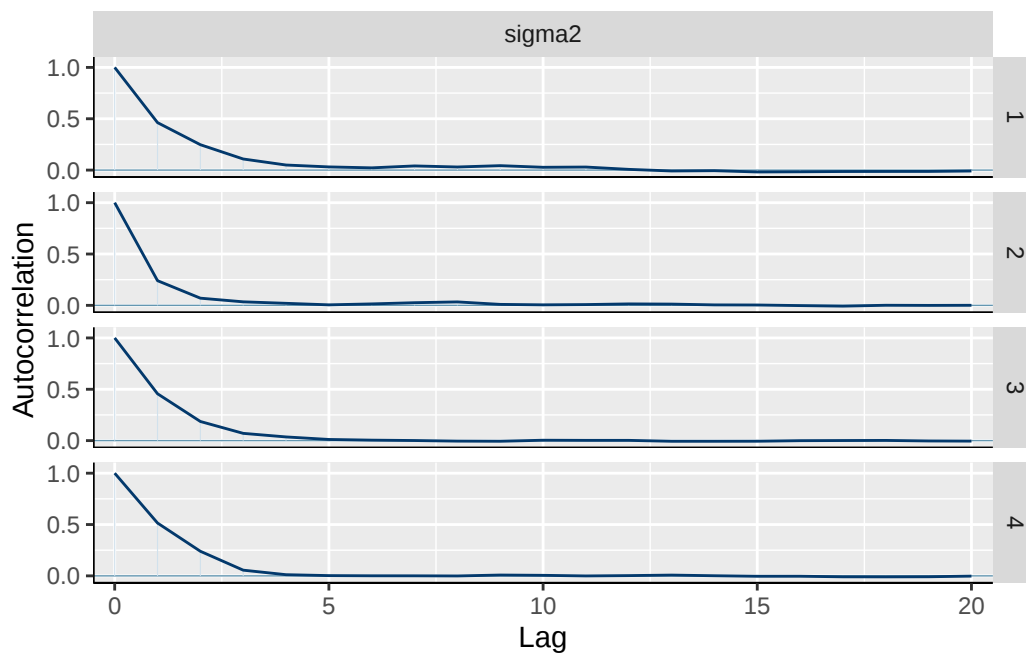
c)

Autokorelacijske funkcije četiriju lanaca:

```
mcmc_acf(sim.MCMC.stan.Apf, pars = "mu")
```



```
mcmc_acf(sim.MCMC.stan.Apf, pars = "sigma2")
```



Vidimo da autokorelacijska funkcija jednako brzo opada za oba parametra kod sva 4 lanca, što sugerira dobru brzinu miješanja.

d)

Efektivne uzoračke veličine:

- Za μ : $S_{\text{eff}}=8829$
- Za σ^2 : 10208

Dobivene vrijednosti sugeriraju da su aproksimacije jednako kvalitetne kao da smo koristili Monte Carlo metodu sa 8829 (10208) simuliranih vrijednosti za μ (σ^2).

- e)
- Za μ : $[1.78, 2.07]$
 - Za σ^2 : $[0.11, 0.17]$

f)

$$P(X_7 \in [1.8, 2] | X_1 = x_1, \dots, X_6 = x_6) \approx 0.51515.$$

```
MCMC <- rstan::extract(sim.MCMC.stan.Apfn)
ecdf(MCMC$x_tilda)(2)-ecdf(MCMC$x_tilda)(1.8)
```

[1] 0.5451

g) Ovdje podrazumijevamo da nemamo apriornih informacija, odnosno da je prikladna apriorna distribucija “uniformna” (u ovom slučaju nepravda za oba parametra). To jednostavno postizemo tako da ne postavimo nikakvu apriornu distribuciju u kodu (default je slabo informativna apriorna distribucija koja doprinosi efikasnom izvršenju algoritma - potiče se koristiti svoje apriorne distribucije).

```
modelMusiceApf2 <- "  
data {  
  real <lower = 0> x[6];          // podaci  
}  
parameters {  
  real mu;          // parametar mu  
  real<lower = 0> sigma2; // parametar sigma2  
}  
model {  
  x ~ normal(mu,sqrt(sigma2)); // vjerodostojnost  
}  
generated quantities {  
  real x_tilda = normal_rng(mu,sqrt(sigma2));  
}  
"  
sim.MCMC.stan.Ap2 <-stan(  
  model_code = modelMusiceApf2,  
  data = list(x=midge$Wing.Length[Species=="Apf"]), # named list of data  
  chains = 4,          # broj korištenih Markovljevih lanaca  
  warmup = 10000,      # burn in period za svaki lanac  
  iter = 20000,        # ukupan broj iteracija svakog lanca  
  cores = 1,           # broj korištenih jezgri  
  refresh = 0,         # no progress show  
  seed=934  
)
```

Usporedimo dva modela:

```
print(sim.MCMC.stan.Ap2)
```

Inference for Stan model: anon_model.
4 chains, each with iter=20000; warmup=10000; thin=1;
post-warmup draws per chain=10000, total post-warmup draws=40000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu	1.92	0.00	0.05	1.82	1.89	1.92	1.95	2.02	16453	1
sigma2	0.03	0.00	0.06	0.00	0.01	0.01	0.03	0.12	15596	1
x_tilda	1.92	0.00	0.17	1.58	1.83	1.92	2.01	2.26	36346	1
lp__	6.07	0.01	1.12	3.07	5.61	6.41	6.89	7.20	11175	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:40:35 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```
print(sim.MCMC.stan.Ap2)
```

Inference for Stan model: anon_model.
4 chains, each with iter=20000; warmup=10000; thin=1;
post-warmup draws per chain=10000, total post-warmup draws=40000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu	1.93	0.00	0.07	1.78	1.89	1.93	1.96	2.07	8829	1
sigma2	0.03	0.00	0.11	0.00	0.01	0.02	0.03	0.17	10208	1
x_tilda	1.93	0.00	0.20	1.54	1.83	1.93	2.02	2.31	27634	1
lp__	5.97	0.01	1.27	2.53	5.45	6.36	6.89	7.23	7436	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:41:23 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

Ne uočavamo posebne razlike.

 Danger

88

Zadatak 88.1. Riješite zadatak 3.5 korištenjem MCMC metode na platformi STAN.

 Danger

89 Rješenje

Aproksimacija posteriornih distribucija na platformi STAN:

Paralelno ćemo napraviti 2 rješenja:

- `modelUsporedbaProporcija1` : (bez postavljanja apriornih distribucija, prepuštamo STAN-u odabir)
- `modelUsporedbaProporcija2` : (koristimo konjugirane apriorne distribucije)

```

modelUsporedbaProporcija1 <- "
data {
  int <lower = 0, upper=60> x[3];          // podaci1
  int <lower = 0, upper=60> y[3];          // podaci2
}
parameters {
  real<lower = 0, upper=1> theta1;          // parametar theta1
  real<lower = 0, upper=1> theta2; // parametar theta2
}
model {
  x ~ binomial(60,theta1); // vjerodostojnost1
  y ~ binomial(60,theta2); // vjerodostojnost2
}
generated quantities {
  int <lower = 0, upper=60> x_tilda = binomial_rng(60,theta1);
  int <lower = 0, upper=60> y_tilda = binomial_rng(60,theta2);
}
"
podaci <- list(x=c(53,56,49), y=c(50,57,54))
sim.MCMC.stan1 <-stan(
  model_code = modelUsporedbaProporcija1,
  data = podaci,      # named list of data
  chains = 4,          # broj korištenih Markovljevih lanaca
  warmup = 10000,      # burn in period za svaki lanac
  iter = 20000,        # ukupan broj iteracija svakog lanca
  cores = 1,           # broj korištenih jezgri
  refresh = 0,         # no progress show
  seed=934
)

print(sim.MCMC.stan1)

```

Inference for Stan model: anon_model.

4 chains, each with iter=20000; warmup=10000; thin=1;

post-warmup draws per chain=10000, total post-warmup draws=40000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta1	0.87	0.00	0.02	0.82	0.86	0.87	0.89	0.92	36345	1
theta2	0.89	0.00	0.02	0.84	0.87	0.89	0.91	0.93	37397	1
x_tilda	52.42	0.02	2.97	46.00	51.00	53.00	55.00	58.00	37629	1
y_tilda	53.41	0.01	2.79	47.00	52.00	54.00	55.00	58.00	37672	1

lp__	-133.09	0.01	1.01	-135.80	-133.47	-132.79	-132.38	-132.11	18378
------	---------	------	------	---------	---------	---------	---------	---------	-------

1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:42:06 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```

MCMC1 <- rstan::extract(sim.MCMC.stan1, permuted=TRUE)

#####

modelUsporedbaProporcija2 <- "
data {
  int <lower = 0, upper=60> x[3];          // podaci1
  int <lower = 0, upper=60> y[3];          // podaci2
}
parameters {
  real<lower = 0, upper=1> theta1;          // parametar theta1
  real<lower = 0, upper=1> theta2; // parametar theta2
}
model {
  x ~ binomial(60,theta1); // vjerodostojnost1
  y ~ binomial(60,theta2); // vjerodostojnost2
  theta1 ~ beta(12,3); // konjugirana apriorna distribucija
  theta2 ~ beta(9.9875, 1.7625); // konjugirana apriorna distribucija
}
generated quantities {
  int <lower = 0, upper=60> x_tilda = binomial_rng(60,theta1);
  int <lower = 0, upper=60> y_tilda = binomial_rng(60,theta2);
}
"
podaci <- list(x=c(53,56,49), y=c(50,57,54))
sim.MCMC.stan1 <-stan(
  model_code = modelUsporedbaProporcija2,
  data = podaci,      # named list of data
  chains = 4,          # broj korištenih Markovljevih lanaca
  warmup = 10000,      # burn in period za svaki lanac
  iter = 20000,        # ukupan broj iteracija svakog lanca
  cores = 1,           # broj korištenih jezgri
  refresh = 0,         # no progress show
  seed=934
)
print(sim.MCMC.stan1)

```

Inference for Stan model: anon_model.
 4 chains, each with iter=20000; warmup=10000; thin=1;
 post-warmup draws per chain=10000, total post-warmup draws=40000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta1	0.87	0.00	0.02	0.82	0.86	0.87	0.89	0.91	33802	1
theta2	0.89	0.00	0.02	0.84	0.88	0.89	0.91	0.93	34013	1
x_tilda	52.31	0.02	2.96	46.00	50.00	53.00	54.00	57.00	38486	1
y_tilda	53.50	0.01	2.75	48.00	52.00	54.00	55.00	58.00	38428	1
lp__	-141.44	0.01	1.01	-144.17	-141.83	-141.14	-140.73	-140.45	17213	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:42:50 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```
MCMC2 <- rstan::extract(sim.MCMC.stan1, permuted=TRUE)
```

d)

```
#Prvi model:
ecdf(MCMC1$theta1)(0.9)-ecdf(MCMC1$theta1)(0.75)
```

```
[1] 0.8626
```

```
ecdf(MCMC1$theta2)(0.9)-ecdf(MCMC1$theta2)(0.75)
```

```
[1] 0.646075
```

```
#Drugi model:
ecdf(MCMC2$theta1)(0.9)-ecdf(MCMC2$theta1)(0.75)
```

```
[1] 0.88615
```

```
ecdf(MCMC2$theta2)(0.9)-ecdf(MCMC2$theta2)(0.75)
```

```
[1] 0.62255
```

e)

```
#Prvi model:
mean(MCMC1$theta1>MCMC1$theta2)
```

```
[1] 0.31285
```

```
#Drugi model:  
mean(MCMC2$theta1>MCMC2$theta2)
```

```
[1] 0.26895
```

f)

```
#Prvi model:  
mean(MCMC1$x_tilda>55)
```

```
[1] 0.146775
```

```
mean(MCMC1$y_tilda>55)
```

```
[1] 0.23965
```

```
#Drugi model:  
mean(MCMC2$x_tilda>55)
```

```
[1] 0.13555
```

```
mean(MCMC2$y_tilda>55)
```

```
[1] 0.246325
```

90 Bayesovski hijerarhijski modeli

 Note

91 Izmjenljivost (eng. Exchangeability)

Neka je $f(x_1, \dots, x_n)$ zajednička gustoća slučajnog vektora $(X_1, \dots, X_n)$. Ako je $f(x_1, \dots, x_n) = f(x_{\pi_1}, \dots, x_{\pi_n})$ za sve permutacije π skupa $\{1, \dots, n\}$, tada kažemo da su slučajne varijable $X_1, \dots, X_n$ izmjenljive.

Sljedeća dva teorema povezuju Bayesovski koncept i koncept izmjenljivosti.

 Important

92 j.s.u. i izmjenljivost

Teorem 92.1. *Neka je $\theta \sim f(\theta)$ i $X_1, \dots, X_n$ uvjetno nezavisne i jednako distribuirane uz dani θ . Tada su $X_1, \dots, X_n$ izmjenljive.*

Drugim riječima:

$$X_1, \dots, X_n | \theta \text{ n.j.d.} \Rightarrow X_1, \dots, X_n \text{ su izmjenljive} .$$
$$\theta \sim f(\theta)$$

O suprotnom smjeru govori sljedeći teorem.

! Important

93 de Finetti

Teorem 93.1. *Neka $(X_n, n \in \mathbb{N})$ niz slučajnih varijabli takvih da je $\mathcal{R}(X_i) = \mathcal{R}(X_j), \forall i, j \in \mathbb{N}$. Neka su $X_1, \dots, X_n$ izmjenljive za svaki $n \in \mathbb{N}$. Tada je*

$$f(x_1, \dots, x_n) = \int \prod_{i=1}^n f(x_i|\theta) f(\theta) d\theta$$

za neki parametar θ uz apriornu distribuciju $f(\theta)$ i neku vjerodostojnost $f(x|\theta)$. Pri tome apriorna distribucija i vjerodostojnost ovise o $f(x_1, \dots, x_n)$.

Zaključujemo:

$$X_1, \dots, X_n | \theta \text{ n.j.d. za svaki } n \in \mathbb{N} \Leftrightarrow \begin{matrix} \theta \sim f(\theta) \\ X_1, \dots, X_n \text{ su izmjenljive za svaki } n \in \mathbb{N}. \end{matrix}$$

Realni scenariji za “ $X_1, \dots, X_n$ izmjenljive za svaki $n \in \mathbb{N}$ ”:

- $X_1, \dots, X_n$ su dobiveni uzorkovanjem iz konačne populacije s vraćanjem
- $X_1, \dots, X_n$ su dobiveni uzorkovanjem iz beskonačne populacije s vraćanjem (u praksi, ako je populacije dimenzije N bit će dovoljno $N \gg n$)
- $X_1, \dots, X_n$ su ishodi ponovljivog pokusa.

Terminologija za hijerarhijske modele nije konzistentna te se isti mogu pronaći pod nazivom “fixed effects” ili “random effects” modeli.

Pretpostavimo da imamo hijerarhijske podatke $(\mathbb{X}_1, \dots, \mathbb{X}_m)$, gdje je $\mathbb{X}_j = (X_{1,j}, \dots, X_{n_j,j})$ (usporedba m grupa svaka dimenzije uzorka $n_j, j \in \{1, \dots, m\}$).

Pri tome:

- Ako je jedino što znamo da je $(X_{1,j}, \dots, X_{n_j,j})$ slučajan uzorak iz grupe j , tretiramo ih kao izmjenljive.

- Ako je $N_j \gg n_j$: $\{X_{1,j}, \dots, X_{n_j,j} | \phi_j\}$ j.s.u. iz $f(x|\phi_j)$, $j = 1, \dots, m$ (prema de Finettijevom teoremu).
- Ako su grupe uzorkovane iz veće populacije grupa i za njih možemo primijeniti izmjenljiv model: $\{\phi_1, \dots, \phi_m | \psi\}$ j.s.u. iz $f(\phi|\psi)$ (ponovo prema de Finettijevom teoremu).

Ovom uzastopnom primjenom de Finettijevog teorema dobivamo:

i Note

94 Bayesovski hijerarhijski model

$X_{1,j}, \dots, X_{n_j,j} | \phi_j$ j.s.u. iz $f(x|\phi_j)$ (varijabilnost uzorkovanja unutar grupe)

$\phi_1, \dots, \phi_m | \psi$ j.s.u. iz $f(\phi|\psi)$ (varijabilnost uzorkovanja među grupama)

$\psi \sim f(\psi)$ (apriorna distribucija)

Primijetimo da iz podataka procjenjujemo uzoračke distribucije $f(x|\phi)$ i $f(\phi|\psi)$, dk se apriorna distribucija ne procjenjuje iz podataka.

94.1 Hijerarhijski normalni model (homogene varijance)

Radi se o popularnom modelu za opisivanje heterogenosti očekivanja više populacija.

95 Bayesovski normalni hijerarhijski model

$X_{1,j}, \dots, X_{n_j,j} | \theta_j, \sigma^2$ j.s.u. iz $\mathcal{N}(\theta_j, \sigma^2), j \in \{1, \dots, m\}$ (varijabilnost uzorkovanja unutar grupe)
 $\theta_1, \dots, \theta_m | \mu, \tau^2$ j.s.u. iz $\mathcal{N}(\mu, \tau^2)$ (varijabilnost uzorkovanja među grupama)

Možemo identificirati $\phi_j = (\theta_j, \sigma^2)$ i $\psi = (\mu, \tau^2)$. Pri tome su (fiksni) nepoznati parametri $\sigma^2 > 0, \mu \in \mathbb{R}, \tau^2 > 0$. Za njih prepostaje postaviti apriorne distribucije kako bi kompletirali Bayesovski hijerarhijski model.

Ukoliko želimo kreirati aproksimaciju posteriorne distribucije parametara koristeći Gibbsovo uzorkovanje, postaviti ćemo sljedeće (semi)konjugatne apriorne distribucije:

$$\begin{aligned} 1/\sigma^2 &\sim \Gamma(\nu_0/2, \nu_0\sigma_0^2/2) \\ 1/\tau^2 &\sim \Gamma(\eta_0/2, \eta_0\tau_0^2/2) \\ \mu &\sim \mathcal{N}(\mu_0, \gamma_0^2). \end{aligned}$$

U konačnici nas zanima MCMC aproksimacija posteriorne distribucije

$$f(\theta_1, \dots, \theta_m, \mu, \tau^2, \sigma^2 | \mathbf{x}_1, \dots, \mathbf{x}_m) \propto f(\mu)f(\tau^2)f(\sigma^2) \left[\prod_{j=1}^m f(\theta_j | \mu, \tau^2) \right] \cdot \left[\prod_{j=1}^m \prod_{i=1}^{n_j} f(x_{i,j} | \theta_j, \sigma^2) \right].$$

- Uvjetne posteriorne distribucije za μ i τ^2 :

$$\begin{aligned} \{\mu | \theta_1, \dots, \theta_m, \tau^2\} &\sim \mathcal{N}\left(\frac{m\bar{\theta}/\tau^2 + \mu_0/\gamma_0^2}{m/\tau^2 + 1/\gamma_0^2}, \frac{1}{m/\tau^2 + 1/\gamma_0^2}\right) \\ \{\frac{1}{\tau^2} | \theta_1, \dots, \theta_m, \mu\} &\sim \Gamma\left(\frac{\eta_0 + m}{2}, \frac{\eta_0\tau_0^2 + \sum_j (\theta_j - \mu)^2}{2}\right) \end{aligned}$$

- Uvjetne distribucije za θ_j :

$$\{\theta_j | x_{1,j}, \dots, x_{n_j,j}, \sigma^2\} \sim \mathcal{N} \left(\frac{n_j \bar{x}_j / \sigma^2 + \mu / \tau^2}{n_j / \sigma^2 + 1 / \tau^2}, \frac{1}{n_j / \sigma^2 + 1 / \tau^2} \right)$$

- Uvjetna distribucija od σ^2 (koncentrirana oko zajedničke uzoračke procjene varijance):

$$\{1/\sigma^2 | \theta_1, \dots, \theta_m, x_1, \dots, x_m\} \sim \Gamma \left(\frac{1}{2}(\nu_0 + \sum_{j=1}^m n_j), \frac{1}{2} \left(\nu_0 \sigma_0^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (x_{i,j} - \theta_j)^2 \right) \right)$$

Konačno, Gibbsovo uzorkovanje za $\phi^{(s)} = (\theta_1^{(s)}, \dots, \theta_m^{(s)}, \mu^{(s)}, \tau^{2(s)}, \sigma^{2(s)})$:

- 1) Generiraj $\mu^{(s+1)} \sim f(\mu | \theta_1^{(s)}, \dots, \theta_m^{(s)}, \tau^{2(s)})$
- 2) Generiraj $\tau^{2(s+1)} \sim f(\tau^2 | \theta_1^{(s)}, \dots, \theta_m^{(s)}, \mu^{s+1})$
- 3) Generiraj $\sigma^{2(s+1)} \sim f(\sigma^2 | \theta_1^{(s)}, \dots, \theta_m^{(s)}, x_1, \dots, x_m)$
- 4) $\forall j \in \{1, \dots, m\}$ generiraj $\theta_j^{(s+1)} \sim f(\theta_j | \mu^{(s+1)}, \tau^{2(s+1)}, \sigma^{2(s+1)}, x_j)$

95.1 Hijerarhijski normalni model (različite varijance)

Dodatno se pretpostavlja da osim populacijskih očekivanja, postoje razlike i u varijabilnosti (σ_j^2 je varijabilnost za grupu j).

i Note

96 Bayesovski normalni hijerarhijski model (različite varijance)

$X_{1,j}, \dots, X_{n_j,j} | \theta_j, \sigma_j^2$ j.s.u. iz $\mathcal{N}(\theta_j, \sigma_j^2), j \in \{1, \dots, m\}$ (varijabilnost uzorkovanja unutar grupe)

$\theta_1, \dots, \theta_m | \mu, \tau^2$ j.s.u. iz $\mathcal{N}(\mu, \tau^2)$ (varijabilnost uzorkovanja među grupama)

$\sigma_1^2, \dots, \sigma_m^2 | \nu_0, \sigma_0^2$ j.s.u. iz $\Gamma(\nu_0/2, \nu_0 \sigma_0^2/2)$ (varijabilnost uzorkovanja među grupama)

Slično kao i u prošlom modelu slijedi

$$\{\theta_j | X_{1,j} \dots, X_{n_j,j}, \sigma_j^2\} \sim \mathcal{N} \left(\frac{n_j \bar{x}_j / \sigma_j^2 + \mu / \tau^2}{n_j / \sigma_j^2 + 1 / \tau^2}, \frac{1}{n_j / \sigma_j^2 + 1 / \tau^2} \right)$$

Uz apriornu distribuciju $\sigma_0^2 \sim \Gamma(a, b)$ slijedi

$$\{\sigma_0^2 | \sigma_1^2, \dots, \sigma_m^2, \mu_0\} \sim \Gamma \left(a + \frac{1}{2} m \nu_0, b + \frac{1}{2} \sum_{j=1}^m (1/\sigma_j^2)^2 \right).$$

Za ν_0 nema jednostavnog konjugata.

 Tip

97 Primjer - hijerarhijski normalni model (poznata varijanca)

Podaci (<http://www2.stat.duke.edu/~pdh10/FCBS/Replication/nels.RData>)

Provedena je studija o rezultatima u matematici učenika istih uzrasta različitih škola u SAD-u. Želimo ustvrditi postoje li razlike u očekivanim rezultatima po školama.

```
## vrijednosti hiperparametara apriorne distribucije
```

```
load(file="C:/Users/Korisnik/Desktop/OneDrive - Sveučilište Josipa Jurja Strossmayera Os
```

```
#### tehnicki detalji - reorganizacija podataka u listu
```

```
Y<-list()
```

```
YM<-NULL
```

```
J<-max(Y.school.mathscore[,1])
```

```
n<-ybar<-ymed<-s2<-rep(0,J)
```

```
for(j in 1:J) {
```

```
  Y[[j]]<-Y.school.mathscore[ Y.school.mathscore[,1]==j,2]
```

```
  ybar[j]<-mean(Y[[j]])
```

```
  ymed[j]<-median(Y[[j]])
```

```
  n[j]<-length(Y[[j]])
```

```
  s2[j]<-var(Y[[j]])
```

```
  YM<-rbind( YM, cbind( rep(j,n[j]), Y[[j]] ))
```

```
}
```

```
colnames(YM)<-c("school","mathscore")
```

```
nu0<-1 ; s20<-100
```

```
eta0<-1 ; t20<-100
```

```
mu0<-50 ; g20<-25
```

```
## početne vrijednosti
```

```
m<-length(Y)
```

```
n<-sv<-ybar<-rep(NA,m)
```

```
for(j in 1:m)
```

```
{
```

```
  ybar[j]<-mean(Y[[j]])
```

```
  sv[j]<-var(Y[[j]])
```

```
  n[j]<-length(Y[[j]])
```

```
}
```

```
theta<-ybar
```

```
sigma2<-mean(sv)
```

```
mu<-mean(theta)
```

```
tau2<-var(theta)
```

```
## postavke za MCMC
```

```
set.seed(1)
```

```
S<-5000
```

```
THETA<-matrix( nrow=S,ncol=m)
```

```
MST<-matrix( nrow=S,ncol=3)
```

```
## MCMC algoritam
```

```
for(s in 1:S)
```

```
{
```

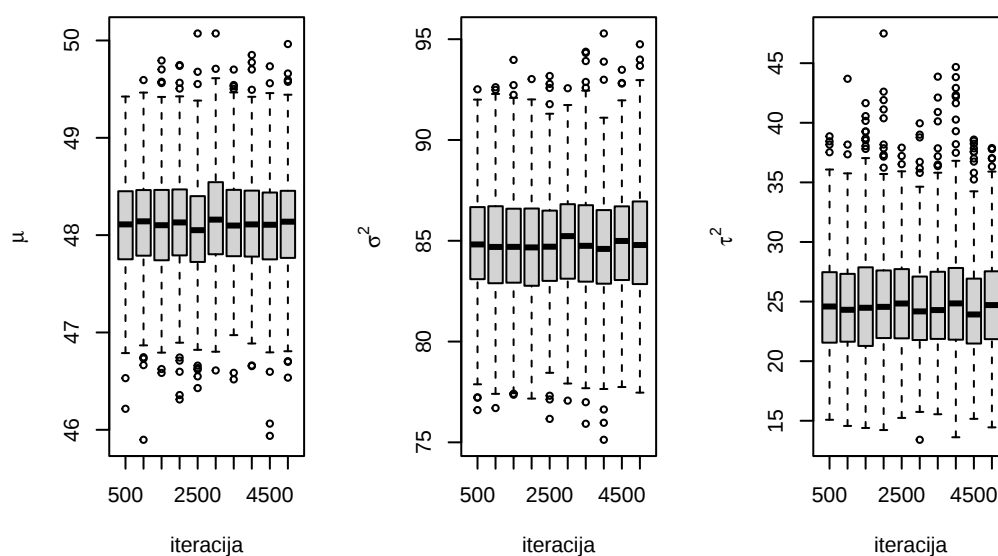
```
  # generiranje novih theta iz uvjetne distribucije uz ostale vrijednosti poznate
```

```
  for(j in 1:m)
```

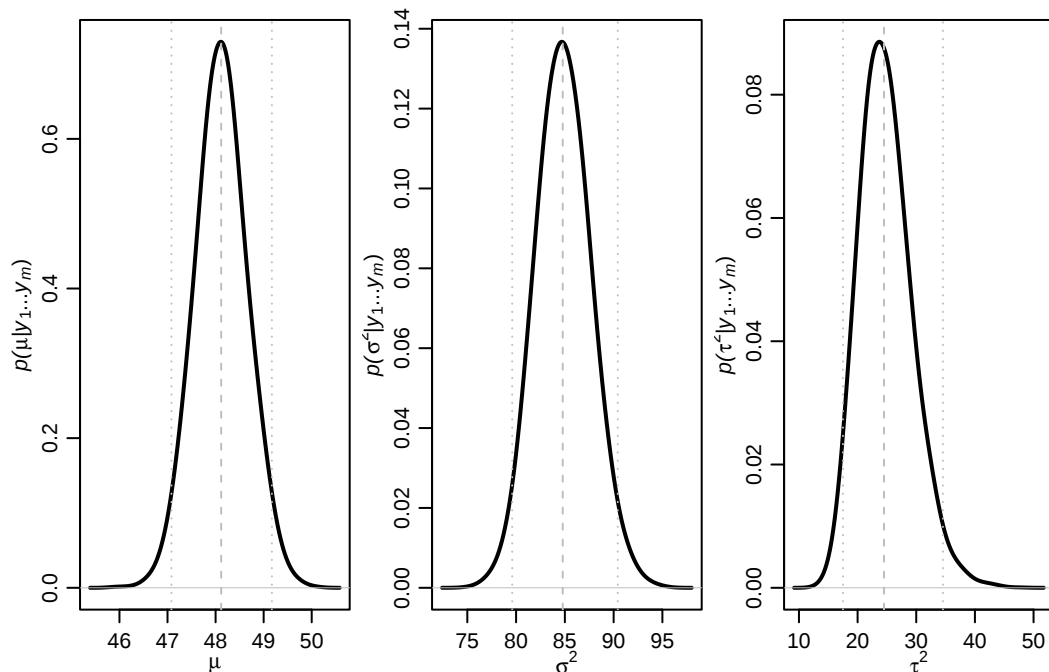
```
{
```

Sljedeći kod grafički ilustrira je li došlo do konvergencije pomoću kutijastog dijagrama za svakih 500 uzastopnih simuliranih vrijednosti parametra iz MCMC metode. U slučaju da nije došlo do konvergencije, uočili bi značajnije razlike u raspodjeli simuliranih vrijednosti kako broj iteracija raste. Graf sugerira da je došlo do konvergencije.

```
par(mfrow=c(1,3))
stationarity.plot(MST[,1],xlab="iteracija",ylab=expression(mu))
stationarity.plot(MST[,2],xlab="iteracija",ylab=expression(sigma^2))
stationarity.plot(MST[,3],xlab="iteracija",ylab=expression(tau^2))
```



```
par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.7,.7,0))
plot(density(MST[,1],adj=2),xlab=expression(mu),main="",lwd=2,
ylab=expression(paste(italic("p("),mu,"|",italic(y[1]),"...",italic(y[m]),")"))))
abline( v=quantile(MST[,1],c(.025,.5,.975)),col="gray",lty=c(3,2,3) )
plot(density(MST[,2],adj=2),xlab=expression(sigma^2),main="", lwd=2,
ylab=expression(paste(italic("p("),sigma^2,"|",italic(y[1]),"...",italic(y[m]),")"))))
abline( v=quantile(MST[,2],c(.025,.5,.975)),col="gray",lty=c(3,2,3) )
plot(density(MST[,3],adj=2),xlab=expression(tau^2),main="",lwd=2,
ylab=expression(paste(italic("p("),tau^2,"|",italic(y[1]),"...",italic(y[m]),")"))))
abline( v=quantile(MST[,3],c(.025,.5,.975)),col="gray",lty=c(3,2,3) )
```



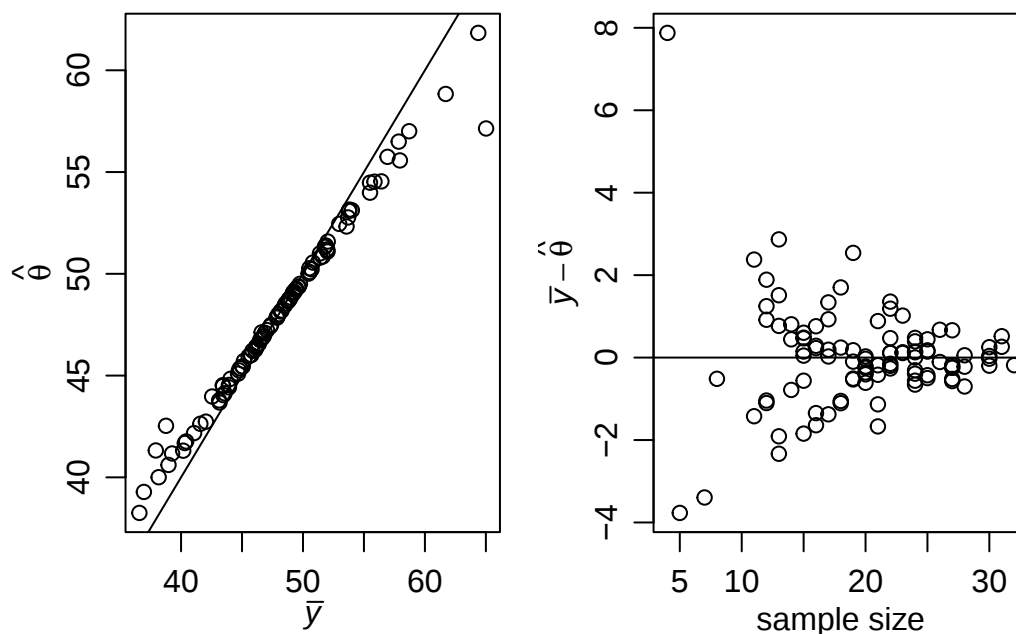
Hijerarhijski model implicira da su očekivane vrijednosti θ_j (uvjetno na podatke, μ, τ^2, σ^2) težinska sredina $\bar{x}_j$ i μ :

$$E[\theta_j | x_j, \mu, \tau, \sigma] = \frac{\bar{x}_j n_j / \sigma^2 + \mu / \tau^2}{n_j / \sigma^2 + 1 / \tau^2}$$

To implicira da je očekivana vrijednost θ_j pomaknuta od prosjeka $\bar{x}_j$ prema μ za iznos koji ovisi o broju podataka u grupi j - n_j . Taj efekt se naziva efekt skupljanja (eng. “shrinkage”) što ilustrira sljedeći grafički prikaz. Uočavamo da velike vrijednosti prosjeka odgovaraju nešto većim vrijednostima $\hat{\theta} = E[\theta_j | x_1, \dots, x_m]$, i obratno. Drugi dio grafičkog prikaza sugerira da je efekt veći što je broj podataka u grupi manji. Drugim riječima, ako imamo puno podataka, nećemo imati potrebu koristiti informacije iz drugih grupa, tj. one neće imati efekta.

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,2))

theta.hat<-apply(THETA,2,mean)
plot(ybar,theta.hat,xlab=expression(bar(italic(y))),ylab=expression(hat(theta)))
abline(0,1)
plot(n,ybar-theta.hat,ylab=expression( bar(italic(y))-hat(theta) ),xlab="sample size")
abline(h=0)
```



Ukoliko bismo htjeli rangirati rezultate škola prema očekivanim vrijednostima Bayesov pristup bi koristio posteriorna očekivanja

$$E[\theta_1 | \mathbf{x}_1, \dots, \mathbf{x}_m], \dots, E[\theta_m | \mathbf{x}_1, \dots, \mathbf{x}_m]$$

dok bi ignorirajući hijerarhijski model koristili prosjeke pojedinih grupa kao procjene očekivanja

$$\bar{x}_1, \dots, \bar{x}_m.$$

Potonju metodu možemo doživjeti kao usporedbu očekivanja nezavisnih populacija (ANOVA).

Iako metode daju slične rezultate, ne u potpunosti iste. Primjerice za škole 46 i 82 bismo dobili drugačije rezultate, što ilustrira sljedeći grafički prikaz (male točke prikazuju podatke, dok velike točke prosjeke pripadnih škola). Ideja je da za školu 46 imamo više podataka te smo uvjereniji u rezultat prosjeka te se posljedično manje skuplja prema zajedničkom populacijskom očekivanju $E[\mu | \mathbf{x}_1, \dots, \mathbf{x}_n]$. S druge strane za školu 82 imamo malo podataka te smo manje uvjereni u dobiveni prosjek (moguće je da se radi o djeci s lošijim rezultatima).

```

par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))

theta.order<-order(theta.hat)

idx<-c(46,82)

ybar.order<-order(ybar)

par(mfrow=c(1,1))
plot(density(THETA[,46],adj=2),col="black",xlim=
      range(c(Y[[46]],Y[[82]],THETA[,c(46,82)])),lwd=2,
      main="",xlab="math score",ylim=c(-.05,.2),ylab="",yaxt="n")
axis(side=2,at=c(0,0.10,0.20) )
lines(density(THETA[,82],adj=2),col="gray",lwd=2)
abline(h=0)

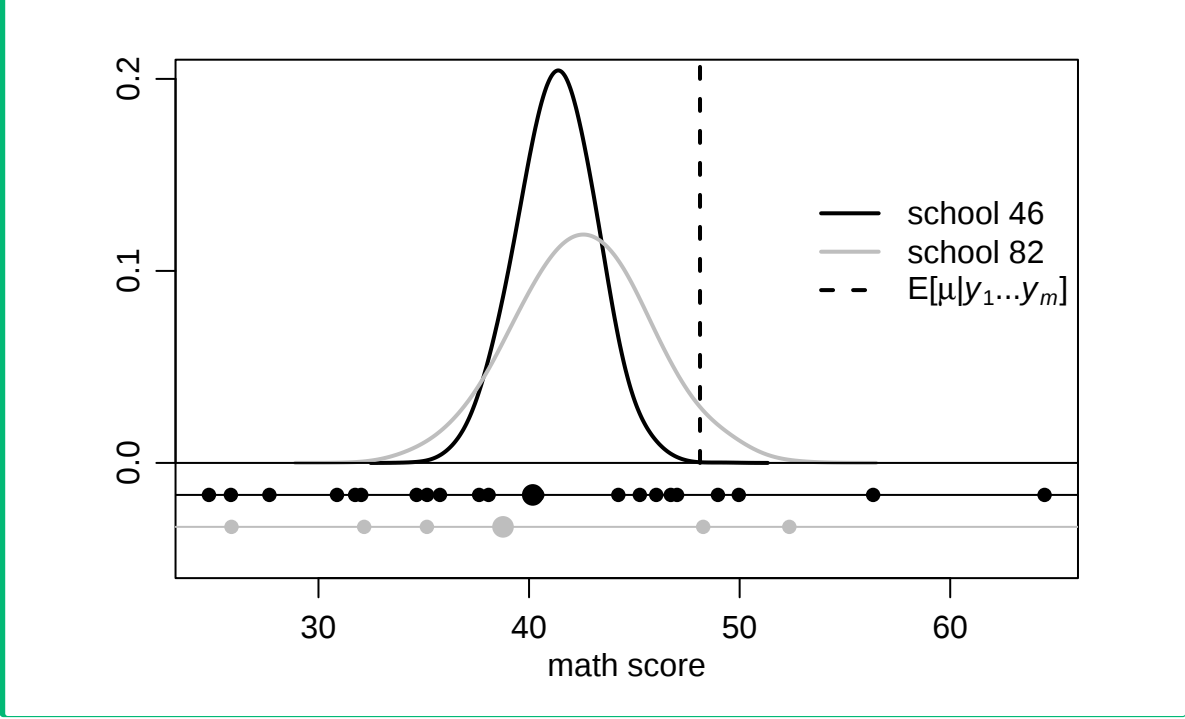
points( Y[[46]],rep(-0.01666,n[46]), col="black",pch=16)
points( ybar[46],-.01666,col="black",pch=16 ,cex=1.5)
abline( h=-.01666,col="black")

points( Y[[82]],rep(-0.0333,n[82]), col="gray",pch=16)
points( ybar[82],-.0333,col="gray",pch=16 ,cex=1.5)
abline( h=-.0333,col="gray")

segments(mean(MST[,1]), 0,mean(MST[,1]),1,lwd=2,lty=2 )

legend(52.5,.15,legend=c("school 46","school 82",
      expression(paste("E[" , mu,"|",italic(y[1]),"...",italic(y[m]),"]"))),
      lwd=c(2,2),lty=c(1,1,2),col=c("black","gray"),bty="n")

```



98

Zadatak 98.1. U osam različitih škola proveden je jednokratni program instrukcija u svrhu poboljšanja rezultata studenata na standardiziranim (SAT) testovima. Rezultati su prikazani u sljedećoj tablici. Pretpostavka je da podaci dolaze iz normalne distribucije. Koristite platformu STAN.

Škola	Izmjereni efekt	Standardna greška
1	28	15
2	8	10
3	-3	16
4	7	11
5	-1	9
6	1	11
7	18	10
8	12	18

- Implementirajte Bayesov model koji tretira svaku školu zasebno (no pooling).
- Implementirajte Bayesov model koji tretira sve škole zajedno (complete pooling).
- Implementirajte hijerarhijski Bayesov model za isti problem.
- Usporedite posteriorna očekivanja parametara za sva 3 modela.

 Danger

99 Rješenje

a)

```
model.a <-'data {  
  int<lower=0> J; // broj škola  
  real y[J]; // izmjereni efekt za školu  
  real<lower=0> sigma[J]; // standardna greška izmjerenog efekta  
}  
parameters {  
  real theta[J]; // očekivani efekt škole  
}  
model {  
  y ~ normal(theta, sigma); //pretpostavljeni normalni model  
}'
```

b)

```
model.b <-'data {  
  int<lower=0> J; // broj škola  
  real y[J]; // izmjereni efekt za školu  
  real<lower=0> sigma[J]; // standardna greška izmjerenog efekta  
}  
parameters {  
  real theta; // zajednički očekivani efekt škola  
}  
model {  
  y ~ normal(theta, sigma); //pretpostavljeni normalni model  
}'
```

```

model.c <-'data {
  int<lower=0> J; // broj škola
  real y[J]; // izmjereni efekt za školu
  real<lower=0> sigma[J]; // standardna greška izmjerenog efekta
}
parameters {
  real theta[J]; // očekivani efekt škole
  real mu; // zajedničko očekivanje škola
  real<lower=0> tau; // varijabilnost među školama
}
model {
  theta ~ normal(mu, tau);
  y ~ normal(theta, sigma); //pretpostavljeni normalni model
}'

```

d)

```

podaci<- list(J=8, y=c(28,8,-3,7,-1,1,18,12), sigma = c(15,10,16,11,9,11,10,18))

model.a.sim <-
  stan(model_code = model.a,
    data = list(J=8, y=c(28,8,-3,7,-1,1,18,12), sigma = c(15,10,16,11,9,11,10,18)),
    chains = 4, iter = 5000*2, cores=1, refresh=0, seed = 934)

model.b.sim <-
  stan(model_code = model.b,
    data = list(J=8, y=c(28,8,-3,7,-1,1,18,12), sigma = c(15,10,16,11,9,11,10,18)),
    chains = 4, iter = 5000*2, cores=1, refresh=0, seed = 934)

model.c.sim <-
  stan(model_code = model.c,
    data = list(J=8, y=c(28,8,-3,7,-1,1,18,12), sigma = c(15,10,16,11,9,11,10,18)),
    chains = 4, iter = 5000*2, cores=1, refresh=0, seed = 934)

```

Warning: There were 466 divergent transitions after warmup. See <https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup> to find out why this is a problem and how to eliminate them.

Warning: Examine the pairs() plot to diagnose sampling problems

Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and t
Running the chains for more iterations may help. See
<https://mc-stan.org/misc/warnings.html#tail-ess>

```
print(model.a.sim)
```

Inference for Stan model: anon_model.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta[1]	27.91	0.08	15.14	-1.69	17.56	28.01	38.21	57.12	35627	1
theta[2]	8.09	0.05	10.03	-11.59	1.27	8.05	14.91	27.80	34670	1
theta[3]	-3.08	0.09	15.86	-33.73	-13.83	-3.18	7.65	28.07	33742	1
theta[4]	6.99	0.06	11.04	-14.83	-0.36	6.98	14.41	28.66	35686	1
theta[5]	-0.99	0.05	9.08	-18.52	-7.20	-1.02	5.18	16.57	34417	1
theta[6]	1.02	0.06	10.97	-20.84	-6.26	1.03	8.35	22.61	35220	1
theta[7]	18.02	0.05	10.01	-1.59	11.22	18.00	24.84	37.68	35682	1
theta[8]	11.96	0.10	17.72	-22.76	-0.02	11.86	23.93	46.33	33297	1
lp__	-4.00	0.02	1.98	-8.66	-5.10	-3.68	-2.54	-1.11	9388	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:47:15 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```
print(model.b.sim)
```

Inference for Stan model: anon_model.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta	7.63	0.05	4.07	-0.27	4.83	7.61	10.43	15.52	7737	1
lp__	-2.85	0.01	0.68	-4.80	-3.03	-2.59	-2.41	-2.35	8799	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:47:59 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

```
print(model.c.sim)
```

```
Inference for Stan model: anon_model.  
4 chains, each with iter=10000; warmup=5000; thin=1;  
post-warmup draws per chain=5000, total post-warmup draws=20000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta[1]	11.88	0.15	8.63	-2.11	6.18	10.74	16.34	32.23	3321	1.00
theta[2]	7.87	0.09	6.45	-4.95	3.79	7.80	11.88	21.23	5672	1.00
theta[3]	5.90	0.10	8.08	-12.44	1.49	6.37	10.85	20.98	6366	1.00
theta[4]	7.67	0.09	6.77	-6.02	3.56	7.61	11.80	21.45	5709	1.00
theta[5]	4.87	0.10	6.41	-9.10	1.03	5.26	9.18	16.38	4199	1.00
theta[6]	6.04	0.09	6.80	-8.55	1.90	6.35	10.42	18.96	5185	1.00
theta[7]	11.05	0.13	7.00	-1.03	6.35	10.47	15.14	26.65	3114	1.00
theta[8]	8.53	0.10	8.23	-7.45	3.65	8.27	13.10	26.09	6486	1.00
mu	8.02	0.10	5.22	-1.75	4.70	7.98	11.22	18.55	2938	1.00
tau	7.30	0.14	5.69	1.09	3.20	5.94	9.76	21.64	1614	1.00
lp__	-18.13	0.21	5.48	-27.97	-22.01	-18.63	-14.50	-6.51	700	1.01

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:48:46 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

100

Zadatak 100.1. U bazi podataka “Skole8.csv” nalaze se podaci o vremenu koje učenici provedu rješavajući zadaću na tjednoj bazi u osam različitih škola.

- a) Predložite hijerarhijski normalni model (uz homogene varijance) za usporedbu očekivanih rezultata među školama. Pri tome koristite sljedeća apriorna uvjerenja:
 - Zajedničko očekivanje svih škola ima normalnu distribuciju s očekivanjem 7 i varijancom 5.
 - Varijabilnost (varijanca) među školama ima inverznu gama distribuciju s parametrima $\eta_0 = 2$ i $\tau_0^2 = 10$
 - Varijabilnost u podacima (varijanca) ima inverznu gama distribuciju s parametrima $\nu_0 = 2$ i $\sigma_0^2 = 15$.
- b) Implementirajte model na platformi STAN i Odredite aproksimacije posteriornih distribucija za varijabilnost (varijancu) među podacima, zajedničko očekivanje škola i očekivanje treće škole.
- c) Odredite aproksimaciju posteriornog očekivanog vremena učenika uloženog na rješavanje zadaće u ovisnosti o školi. Usporedite s prosječnim rezultatima po školama. Grafički usporedite. Postoje li razlike u rangiranju škola?
- d) Odredite posteriornu vjerojatnost da učenici sedme škole očekivano manje provode vremena rješavajući zadaću na tjednoj bazi u odnosu na šesti školu.
- e) Odredite posteriornu vjerojatnost da učenici sedme škole očekivano najmanje provode vremena rješavajući zadaću na tjednoj bazi.

 Danger

101 Rješenje

b)

```

skole <- 'data {
  int<lower=1> N; //ukupan broj podataka
  int<lower=1> j; // broj grupa
  vector[N] y; // podaci
  int n[j]; // velicina grupe
  real<lower=0> nu0; //hiperparametri
  real<lower=0> eta0;
  real<lower=0> tau20;
  real<lower=0> sigma20;
  real<lower=0> gama20;
  real<lower=0> mu0;
}
parameters {
  real<lower=0> tau2; // varijanca grupa
  real<lower=0> sigma2; // varijanca podataka
  real mu; // zajedničko očekivanje svih populacija
  real theta[j]; // očekivanje (efekt) grupa
}
model {
  tau2 ~ inv_gamma(eta0/2, eta0*tau20/2);
  sigma2 ~ inv_gamma(nu0/2, nu0*sigma20/2);
  mu ~ normal(mu0, sqrt(gama20));
  theta ~ normal(mu, sqrt(tau2));
  int pos;
  pos = 1;
  for (k in 1:j) {
    segment(y, pos, n[k]) ~ normal(theta[k], sqrt(sigma2));
    pos = pos + n[k];
  }
}
'

```

```

baza.skole<-read.csv2("/Users/Korisnik/Desktop/OneDrive - Sveučilište Josipa Jurja Strossmanna/
attach(baza.skole)
skole8.podaci <- list(
  j = 8,
  N=length(Vrijeme),
  n=c(25, 23, 20, 24, 24, 22, 22, 20),
  y = Vrijeme,
  mu0=7,
  gama20=5,
  tau20=10,
  eta0=2,
  sigma20=15,
  nu0=2)

```

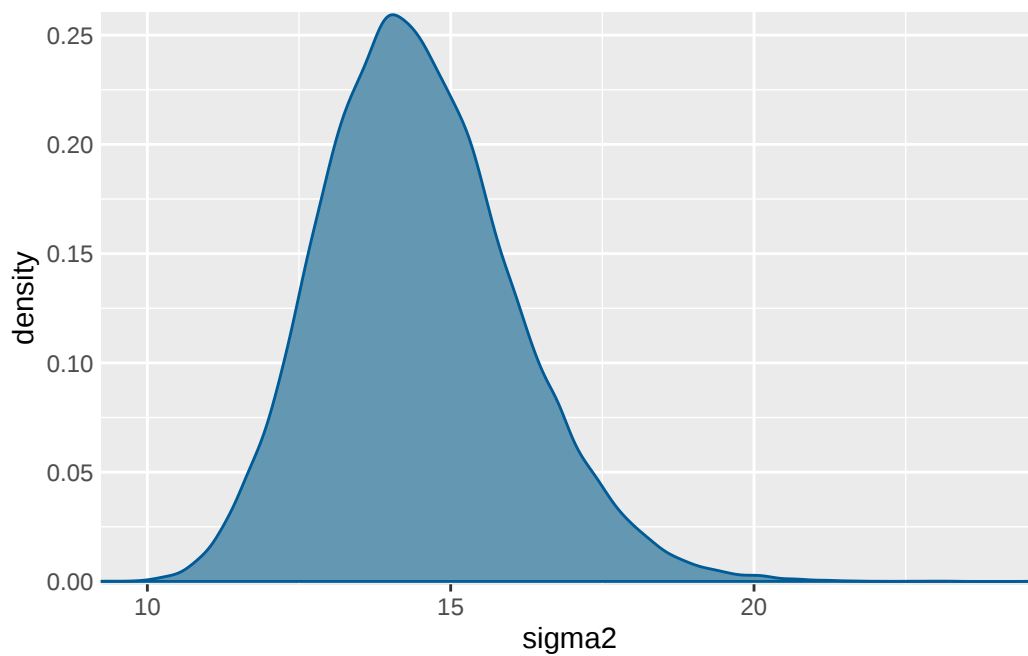
```

sim.MCMC.stan.skole <-stan(
  model_code = skole,
  data = skole8.podaci, # named list of data
  chains = 4, # broj korištenih Markovljevih lanaca

```

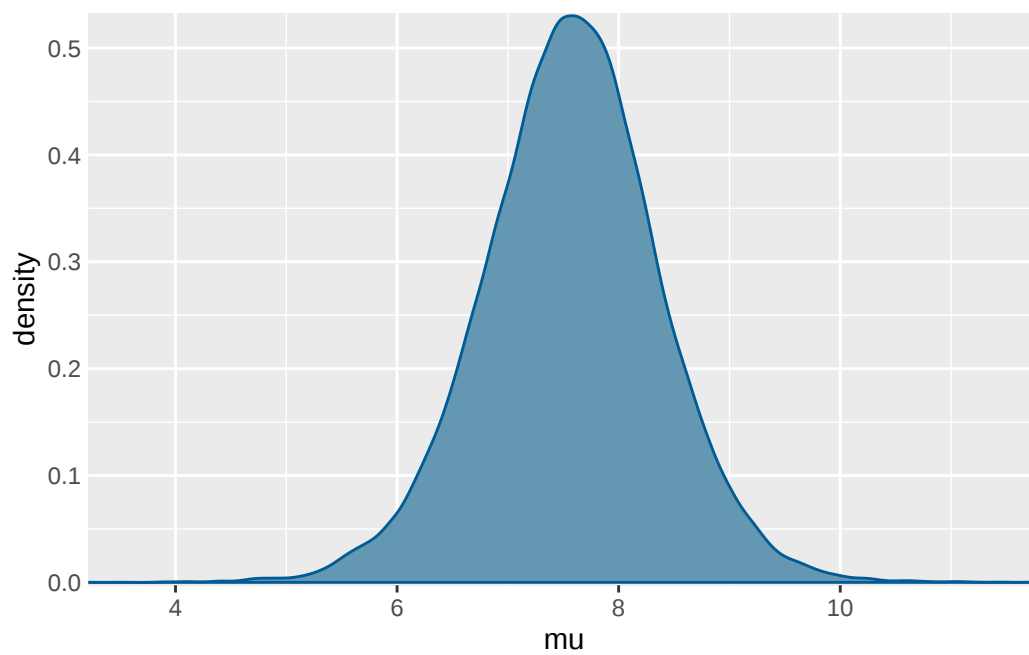
Aproksimacija posteriorne distribucije za σ^2 :

```
mcmc_dens(sim.MCMC.stan.skole, pars = "sigma2") +  
  yaxis_text(TRUE) +  
  ylab("density")
```



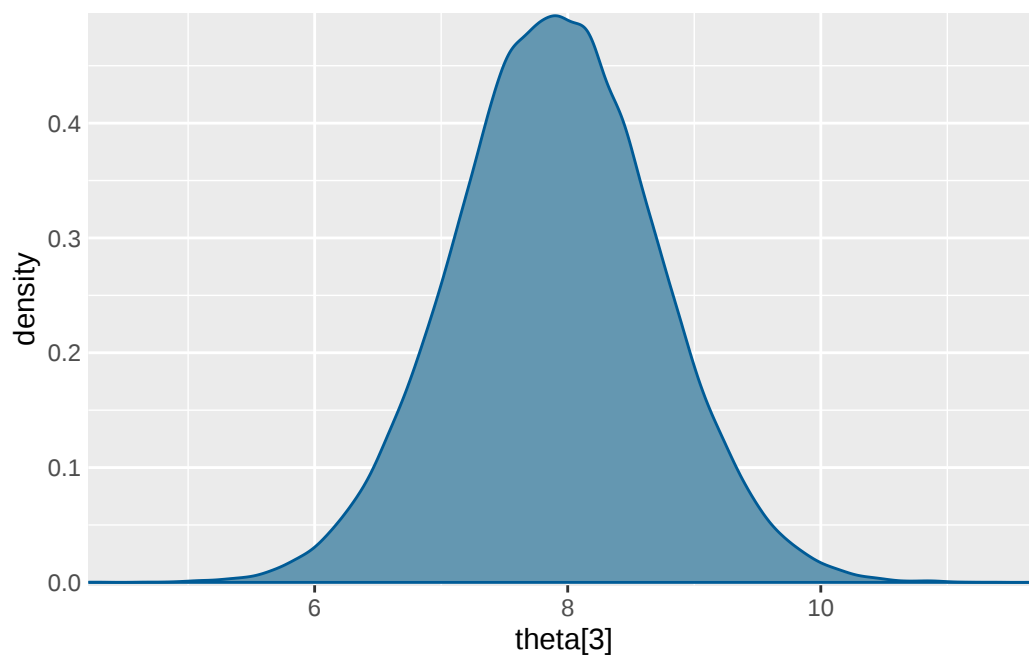
Aproksimacija posteriorne distribucije za μ :

```
mcmc_dens(sim.MCMC.stan.skole, pars = "mu") +  
  yaxis_text(TRUE) +  
  ylab("density")
```



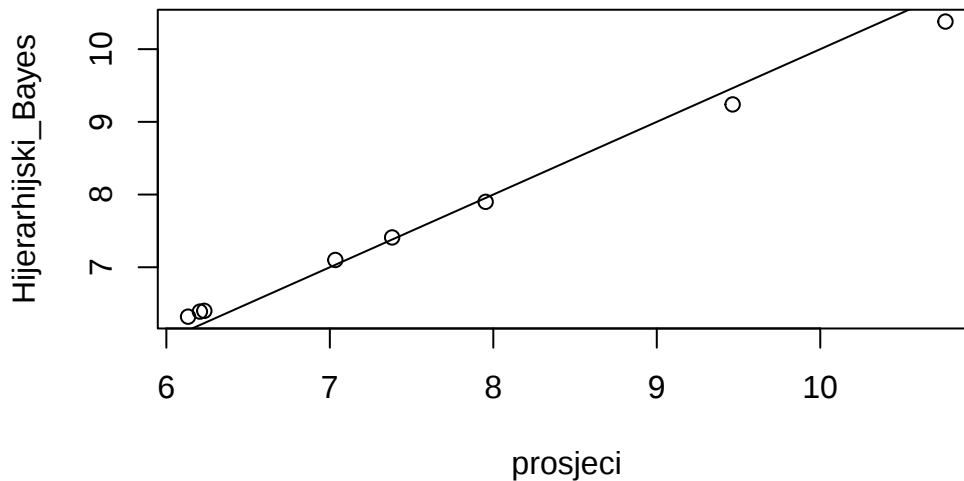
Aproksimacija posteriorne distribucije za θ_3 :

```
mcmc_dens(sim.MCMC.stan.skole, pars = "theta[3]") +  
  yaxis_text(TRUE) +  
  ylab("density")
```



c)

```
prosjeci<-c(9.464000, 7.033478, 7.953000, 6.232083, 10.765833, 6.205000, 6.132727,
Hijerarhijski_Bayes <- c(9.24, 7.1, 7.9, 6.4, 10.38, 6.39, 6.32, 7.41)
plot(prosjeci, Hijerarhijski_Bayes)
abline(a=0,b=1)
```



Ugrubo, možemo uočiti da Bayesovski hijerarhijski model školama s manjim prosjekom dodjeljuje veće efekte, dok onim s značajno većim prosjekom dodjeljuje manje efekte. Iako postoje razlike, prilikom rangiranja škola prema očekivanom vremenu uloženom na zadaće one ne bi postojale.

d)

$$P(\theta_7 < \theta_6 | \mathbf{y}_7, \mathbf{y}_6) \approx 0.52415$$

```
MCMC.skole <- rstan::extract(sim.MCMC.stan.skole)
mean(MCMC.skole$theta[,7] < MCMC.skole$theta[,6])
```

```
[1] 0.52415
```

e)

$$P(\theta_7 < \min\{\theta_1, \dots, \theta_6, \theta_8\} | \mathbf{y}_1, \dots, \mathbf{y}_8) \approx 0.32285$$

```
mean((MCMC.skole$theta[,7] < MCMC.skole$theta[,1]) & (MCMC.skole$theta[,7] < MCMC.skole$theta[,2]) & (MCMC.skole$theta[,7] < MCMC.skole$theta[,3]) & (MCMC.skole$theta[,7] < MCMC.skole$theta[,4]) & (MCMC.skole$theta[,7] < MCMC.skole$theta[,5]) & (MCMC.skole$theta[,7] < MCMC.skole$theta[,8]))
```

```
[1] 0.32285
```

102 Bayesov faktor, testiranje hipoteza i selekcija modela

Neka su M_1 , M_2 i E tri događaja, pri čemu M_1 i M_2 predstavljaju 2 različita modela, a E podatke (eng. “evidence”). Kako je

$$P(M_1|E) = \frac{P(E|M_1)P(M_1)}{P(E)}, \quad P(M_2|E) = \frac{P(E|M_2)P(M_2)}{P(E)}$$

slijedi

$$\frac{P(M_1|E)}{P(M_2|E)} = \frac{P(E|M_1) P(M_1)}{P(E|M_2) P(M_2)}.$$

Prethodnu jednakost interpretiramo kao

“Posteriorna šansa = Bayesov faktor x apriorna šansa”.

Bayesov faktor (BF) u omjer stavlja vjerojatnosnu “uvjerljivost” podataka (E) pod modelom M_1 i pod modelom M_2 .

i Note

103 Interpretacija Bayesovog faktora

- 1) $BF > 1$: model M_1 je uvjerljiviji od modela M_2 nakon što vidimo podatke E .
- 2) $BF < 1$: model M_1 je manje uvjerljiviji od modela M_2 nakon što vidimo podatke E :
- 3) $BF = 1$: modeli su podjednako uvjerljivi nakon što vidimo podatke.

U tom kontekstu, Bayesov faktor koristan je za selekciju modela. Zanimljivo, često korišteni Bayesov informacijski kriterij (BIC) u frekvencionističkoj statistici zapravo je svojevrsna aproksimacija Bayesovog faktora.



Tip

104 Primjer

Pretpostavimo da bacamo novčić 200 puta i 115 puta padne glava, a 85 puta pismo. Statistički model (vjerodostojnost) u problemu je $f(x|\theta) = \binom{200}{x} \theta^x (1-\theta)^{200-x}$, $\theta \in \langle 0, 1 \rangle$, pri čemu je θ nepoznata vjerojatnost da padne pismo na novčiću. Želimo usporediti dva modela

- $M_1 : \theta \geq 0.5$
- $M_2 : \theta < 0.5$

i pretpostavljamo da nemamo nikakvih apriornih informacija pa koristimo uniformnu apriornu distribuciju. U tu svrhu izračunamo Bayesov faktor (BF):

$$BF = \frac{P(M_1|E)}{P(M_2|E)} = \frac{P(\theta \geq 0.5|X = 85)}{P(\theta < 0.5|X = 85)} \bigg/ \frac{P(\theta \geq 0.5)}{P(\theta < 0.5)} \approx 0.01734.$$

Dobiveni BF značajno je manji od 1 što sugerira da je vjerojatniji model M_2 , odnosno da je novčić nepravilan i da češće pada glava.

Frekvencionistički bi prethodni problem sveo na testiranje hipoteza

$$\begin{aligned} H_0 : \theta &\geq 0.5 \\ H_1 : \theta &< 0.5. \end{aligned}$$

Provođenjem binomnog testa dobili bismo p-vrijednost

$$p = P(X \leq 85|M_1) = 0.02 < 0.05$$

što je manje od razine značajnosti $\alpha = 0.05$ te bismo odbacili nul-hipotezu i prihvatili alternativnu (model M_2). Možemo primijetiti da su u ovom problemu Bayesovski i frekvencionistički pristup dali jednake zaključke. Ipak, treba napomenuti da Bayesovski pristup ovisi i o izboru apriorne distribucije, tj. drugačiji odabir apriorne distribucije može dati zaključke drugačije onima kao kod frekvencionističkog pristupa.

Bayesovski pristup “testiranje hipoteza” svodi se na jednostavni izračun vjerojatnosti iz posteriorne distribucije. To je razlog zašto se koncept “testiranja hipoteza” u Bayesovskog paradigmi često preskače.

105 p-vrijednost

Uočimo:

- Bayesovska “p-vrijednost”: nakon što vidimo podatke koliko je uvjerljiv model (hipoteza)
- Frekvencionistička “p-vrijednost”: ako je model (hipoteza) istinita, koliko su podaci uvjerljivi.

Prethodno razmatranje jasno ukazuje da je Bayesovski pristup prirodniji jer ne uzima u obzir nepostojeće podatke.

U slučaju jednostranih hipoteza prethodno opisani Bayesovski pristup funkcionira, ali ako bismo htjeli formalno testirati dvostranu hipotezu $\theta \neq 0.5$ tada prethodno provedni izračun za BF ne funkcionira.

U tom slučaju koristimo vjerodostojne intervale kao alternativu. U konkretnom slučaju

$$\theta | \{X = 85\} \sim \text{Beta}(86, 116) \Rightarrow P(\theta \in [0.36, 0.49] | X = 85) \approx 0.95$$

pa takvu hipotezu odbacujemo jer prethodni 0.05-vjerodostojni interval ne sadrži $= 0.5$.

106 Regresija

Zanima nas modeliranje slučajne varijable Y na temelju $p \geq 1$ slučajnih varijabli (prediktora) $X_1, \dots, X_p$. U slučaju $p = 1$ govorimo o jednostavnoj regresiji. Pri tome raspoložemo podacima u n mjerenja

$$(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n),$$

pri čemu je $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})$, $x_{i,j} \in \mathbb{R}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, p\}$.

Regresijski model je općenito oblika

$$Y_i = f(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n \quad (\text{regresijski model})$$

gdje su $\epsilon_1, \dots, \epsilon_n$ slučajne varijable takve da je $E\epsilon_i = 0$, $i \in \{1, \dots, n\}$ koje zovemo greške.

Obično koristimo dodatne pretpostavke na greške u svrhu modelskog zaključivanja. S druge strane, oblik funkcije f definira tip regresije (linearna, eksponencijalna regresija...).

Primijetimo da regresijski model možemo zapisati i u alternativnom obliku

$$E[Y_i | \mathbf{x}_i = \mathbf{x}_i] = f(\mathbf{x}_i), \quad i \in \{1, \dots, n\}.$$

Umjesto očekivanja, moguće je modelirati neku njegovu funkciju (koju zovemo funkcija veze, eng. link function), a tada govorimo o generaliziranim regresijama, od kojih izdvajamo generalizirane linearne modele (GLM) koji podrazumijevaju linearan izbor funkcije f . Svi modeli koje ćemo obraditi su specijali slučajevi GLM modela.

106.1 Linearna regresija

Linearna regresija je specijalni slučaj GLM metoda kod kojih izravno modeliramo očekivanje kao linearnu kombinaciju prediktora, tj. za funkciju f biramo

$$f(\mathbf{x}_i) = \beta^T \mathbf{x}_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p}.$$

U Bayesovskoj paradigmi parametar β je slučajna varijabla, a u tom kontekstu nas zanima distribucija $f(y|\mathbf{x})$. Linearna regresija podrazumijeva linearan oblik distribucije $f(y|\mathbf{x})$, tj.

$$\int y f(y|\mathbf{x}) dy = E[Y|\mathbf{x}] = \beta_1 x_1 + \beta_2 x_2 + \dots \beta_p x_p = \beta^T \mathbf{x}.$$

Međutim, radi potpunosti Bayesovskog modela potrebno je specificirati vjerodostojnost, a ne samo jednu numeričku karakteristiku. U tu svrhu koristimo pretpostavku o normalnosti i nezavisnosti grešaka:

$$(\epsilon_1, \dots, \epsilon_n) \text{ j.s.u. iz } \mathcal{N}(0, \sigma^2).$$

Time je vjerodostojnost zadana s

$$\begin{aligned} f(y_1, \dots, y_n | \mathbf{x}_1, \dots, \mathbf{x}_n, \beta, \sigma^2) &= \prod_{i=1}^n f(y_i | \mathbf{x}_i, \beta, \sigma^2) \\ &= \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta^T \mathbf{x}_i)^2\right\} \end{aligned}$$

Prepostaje definirati apriorne distribucije za β i σ^2 kako bi Bayesovski model bio potpun.

U frekvencionističkoj linearnoj regresiji β i σ^2 su fiksne veličine, a vjerodostojnost će biti maksimizirana kada je suma kvadrata reziduala $SSR(\beta) = \sum_{i=1}^n (y_i - \beta^T \mathbf{x}_i)^2$ minimizirana. Pokazuje se da je MLE procjenitelj

$$\hat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Definiranu vjerodostojnosti možemo zapisati i u multivarijatnom obliku, tj. imamo

$$\{\mathbf{y} | \mathbf{X}, \beta, \sigma^2\} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$$

Ovaj Bayesovski model dopušta semikonjugate:

$$\begin{aligned} \beta &\sim \mathcal{N}(\beta_0, \Sigma_0) \Rightarrow \{\beta | \mathbf{y} = \mathbf{y}, \mathbf{X} = \mathbf{X}, \sigma^2\} \sim \mathcal{N}\left((\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1} (\Sigma_0^{-1} \beta_0 + \mathbf{X}^T \mathbf{y} / \sigma^2), (\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1}\right) \\ 1/\sigma^2 &\sim \Gamma(\nu_0/2, \nu_0 \sigma_0^2/2) \Rightarrow \{\sigma^2 | \mathbf{y} = \mathbf{y}, \mathbf{X} = \mathbf{X}, \beta\} \sim \text{InvGama}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + SSR(\beta)}{2}\right). \end{aligned}$$

Na temelju ovih semikonjugata moguće je napraviti posteriornu aproksimaciju koristeći Gibbsovo uzorkovanje.

Radi usporedbe s frekvencionistički procjeniteljem uočimo:

$$E[\beta|\mathbb{Y} = \mathbf{y}, \mathbb{X} = \mathbf{x}, \sigma^2] \approx (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X} \mathbf{y} = \hat{\beta}_{OLS}$$

ako je Σ_0^{-1} matrica male magnitude.

S druge strane, mogu se koristiti i slabo informativne apriorne distribucije, od kojih izdvajamo popularnu “g-apriornu distribuciju” koja osigurava da je procjena parametara invarijantna s obzirom na skalu regresora (primjerice jesu li podaci iskazani na mjesečnoj ili godišnjoj razini). Osim toga omogućuje da koristimo izravno MC metodu, umjesto korelirane MCMC metode:

$$\begin{aligned} \beta &\sim \mathcal{N}\left(\mathbb{0}, g\sigma^2 (\mathbb{X}^T \mathbb{X})^{-1}\right) \Rightarrow \{\beta|\mathbb{Y} = \mathbf{y}, \mathbb{X} = \mathbf{x}, \sigma^2\} \sim \mathcal{N}\left(\frac{g}{g+1}(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X} \mathbf{y}, \frac{g}{g+1}\sigma^2(\mathbb{X}^T \mathbb{X})^{-1}\right) \\ 1/\sigma^2 &\sim \Gamma(\nu_0/2, \nu_0\sigma_0^2/2) \Rightarrow \{\sigma^2|\mathbb{Y} = \mathbf{y}, \mathbb{X} = \mathbf{x}\} \sim \text{InvGama}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR_g}{2}\right). \end{aligned}$$

106.1.1 O selekciji modela u linearnoj regresiji

Ukoliko imamo velik broj mogućih prediktora na raspolaganju potrebno je odabrati razuman broj kvalitetnih prediktora za model. U Bayesovskoj paradigmi selekcija modela (u modelu linearne regresije koje prediktore uključiti) je jednostavna. Definiramo niz Bernoullijevih slučajnih varijabli kojim omogućujemo uključivanje/isključivanje pojedinog prediktora:

$$Y_i = z_1 b_1 x_{i,1} + \dots z_p b_p x_{i,p} + \epsilon_i,$$

gdje je

$$\beta_j = z_j \cdot b_j, \quad z_j \in \{0, 1\}, \quad b_j \in \mathbb{R}.$$

Uvođenjem $\mathbf{z} = (z_1, \dots, z_p)$ možemo uspoređivati modele s različitim prediktora na temelju Bayesovog faktora:

$$BF = \frac{P(\mathbf{y}|\mathbb{X} = \mathbf{x}, z_a)}{P(\mathbf{y}|\mathbb{X} = \mathbf{x}, z_b)} = \frac{P(z_a|\mathbf{y}, \mathbb{X} = \mathbf{x})}{P(z_b|\mathbf{y}, \mathbb{X} = \mathbf{x})} \cdot \frac{P(z_a)}{P(z_b)},$$

pri čemu nizovi nula i jedinica z_a i z_b specifikiraju koji su prediktori uključeni u dva različita modela (a i b).

Treba napomenuti kako nam je sada potrebna posteriorna distribucija $f(\mathbf{z}|\mathbf{y}, \mathbf{x})$ kako bismo mogli računati pripadni BF.

Analogno bismo napravili i za proizvoljnu regresiju.

106.2 Poissonova regresija

Raspoložemo podacima u n mjerenja

$$(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n), \quad y_i \in \mathbb{N}_0,$$

pri čemu je $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})$, $x_{i,j} \in \mathbb{R}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, p\}$.

Poissonova regresija podrazumijeva sljedeću strukturu:

i Note

107 Poissonova regresija

$$Y_i | \beta_0, \dots, \beta_p \sim \mathcal{P}(\lambda_i), \quad \lambda_i > 0, \beta_k \in \mathbb{R}, i \in \{1, \dots, n\}, k \in \{1, \dots, p\}$$

$$\ln(\lambda_i) = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}, \quad i \in \{1, \dots, n\}.$$

Interpretacija parametara:

- a) β_0 : kada su $X_{i,1} = X_{i,2} = \dots = X_{i,p} = 0$ tada je očekivana vrijednost od Y_i jednaka e^{β_0}
- b) β_k : Neka je λ_k očekivana vrijednost od Y_i kada je $X_{i,k} = x$ (uz ostale prediktore fiksne). Ako $X_{i,k}$ povećamo za jediničnu vrijednost (iz x u $x+1$) imamo

$$\beta_k = \ln(\lambda_{x+1}) - \ln(\lambda_x), \quad e^{\beta_k} = \frac{\lambda_{x+1}}{\lambda_x}.$$

Dakle, e^{β_k} je multiplikativna promjena u očekivanoj vrijednosti Y_i (β_k je promjena u logaritmiranoj očekivanoj vrijednosti Y_i).

S druge strane, usporedimo linearnu i Poissonovu regresiju:

- 1) Za razliku od linearne regresije ovdje nemamo “greške” (ϵ_i) u modelu jer Poissonova distribucija ima ugrađenu varijancu kroz očekivanje ($EY = Var(Y) = \lambda$).
- 2) Za razliku od linearne regresije ovdje ne modeliramo izravno uvjetno očekivanje $E[Y_i | \mathbf{x}_i = \mathbf{x}_i]$ već njegovu logaritamsku transformaciju. Dakle u Poissonovoj regresiji funkcija veze je logaritamska funkcija, dok je u linearnoj regresiji identiteta.
- 3) Linearna regresija podrazumijeva normalan model za podatke, dok Poissonova regresija Poissonov model za podatke

Prema tome i Poissonova regresija spada u GLM metode. U ovoj vrsti regresije nema jednostavnih semikonjugata koji vode prema posteriornoj distribuciji koja je poznata u zatvorenom obliku. U tom slučaju najjednostavnije je koristiti normalne apriorne distribucije za parametre budući da $\ln(\lambda_i)$ može biti bilo koji realan broj.

Note

108 Pretpostavke u Poissonovoj regresiji

- 1) **Struktura podataka** : uvjetno na prediktore $\mathbb{X}_i$, Y_i je nezavisan od ostalih mjerenja Y_j .
- 2) **Struktura (zavisne) varijable** Y : Y ima Poissonovu distribuciju i predstavlja diskretan brojač događaja unutar fiksnog intervala vremena ili prostora.
- 3) **Struktura veze**: logaritmirana očekivana vrijednost od Y može se zapisati kao linearna kombinacija prediktora ($\ln(\lambda_i) = \beta_0 + \beta_1 x_{i,1} + \dots \beta_p x_{i,p}$).
- 4) **Struktura varijabilnosti** u Y : $EY = VarY = \lambda \Rightarrow$ uvjetno na prediktore $\mathbb{X}$, tipična vrijednost od Y bi trebala približno biti jednaka varijabilnosti u Y (varijabilnost raste s očekivanjem).

108.1 Logistička regresija

Logističku regresiju koristimo kada želimo raditi binarnu klasifikaciju na temelju prediktora. Raspolažemo podacima u n mjerenja

$$(y_1, \mathbb{x}_1), \dots, (y_n, \mathbb{x}_n), \quad y_i \in \{0, 1\},$$

pri čemu je $\mathbb{x}_i = (x_{i,1}, \dots, x_{i,p})$, $x_{i,j} \in \mathbb{R}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, p\}$.

Logistička regresija podrazumijeva sljedeću strukturu:

109 Logistička regresija

$$Y_i | \beta_0, \dots, \beta_p \sim \text{Ber}(\pi_i), \quad \pi_i \in \langle 0, 1 \rangle, \beta_k \in \mathbb{R}, i \in \{1, \dots, n\}, k \in \{1, \dots, p\}$$

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}, \quad i \in \{1, \dots, n\}.$$

Interpretacija parametara:

- a) β_0 : kada su $X_{i,1} = X_{i,2} = \dots = X_{i,p} = 0$ tada šanse za događaj uspjeha (1) e^{β_0}
 b) β_k : Ukoliko sve prediktore osim $X_{i,k}$ držimo fiksnim i $X_{i,k}$ poraste za jediničnu veličinu (iz x u $x + 1$) imamo:

$$\beta_k = \ln(\text{šanse}_{x+1}) - \ln(\text{šanse}_x), \quad e^{\beta_k} = \frac{\text{šanse}_{x+1}}{\text{šanse}_x}.$$

Niti u ovom modelu nećemo proučavati semikonjugate, a za apriorne distribucije parametara možemo koristiti normalnu distribuciju koja reflektira naše subjektivno znanje o problemu (kroz interpretaciju parametara).

Uočimo kako je u ovom modelu funkcija veze logistička funkcija $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ jer je $\pi_i = E[Y_i | \mathbb{X}_i = \mathbb{x}_i]$.

109.0.1 Klasifikacija u logističkom modelu

Napravimo MCMC simulaciju N prediktivnih vrijednosti modela $\{\tilde{Y}_1, \dots, \tilde{Y}_N\}$ i izračunamo proporciju uspjeha (1) u simuliranim vrijednostima u oznaci p .

Ukoliko postavimo klasifikacijski “cut-off” $c \in [0, 1]$ imamo:

- Ako je $p \geq c$ klasificiraj Y kao 1.
- Ako je $p < c$ klasificiraj Y kao 0.

“Cut off” c subjektivno zadajemo i ovisi o problematici.

109.1 Zadaci



110

Zadatak 110.1. U bazi podataka **bikes** nalaze se podaci o korisnicima usluge dijeljenog korištenja bicikla u Washingtonu.

- a) Promotrite podatke o broju vožnji bicikala po danu. Predložite distribucijski model za podatke.
- b) Predložite i implementirajte model za broj vožnji bicikala po danu u ovisnosti o temperaturi koristeći STAN platformu. Prilikom definiranja apriornih distribucija koristite sljedeće informacije:
 - Na tipični dan (s temperaturom 70 F) očekujemo 5000 vožnji, a uzevši u obzir varijabilnost, očekujemo između 3000 i 7000 vožnji.
 - Za svako povećanje temperature u iznosu 1 stupanj, broj vožnji se tipično poveća za 100, a ako uzmemo u obzir neizvjesnost očekujemo povećanje između 20 i 180.
 - Za proizvoljnu temperaturu, broj dnevnih vožnji varira s tipičnom standardnom devijacijom od 1250 vožnji.
- c) Interpretirajte posteriorne distribucije regresijskih parametara β_0 i σ .
- d) Aproksimirajte posteriornu prediktivnu distribuciju broja vožnji bicikala na slučajno odabrani dan s temperaturom 75 F .
- e) Aproksimirajte posteriornu vjerojatnost da na slučajno odabrani dan s temperaturom od 75 F broj vožnji bicikala bude između 2500 i 5000.
- f) Ukoliko se temperatura poveća za 1 stupanj, koliku promjenu u očekivanom broju vožnji očekujete?

 Danger

111 Rješenje

b)

```
model.bicikli <- "  
data {  
  int<lower = 0> n;  
  vector[n] Y;  
  vector[n] X;  
}  
parameters {  
  real beta0;  
  real beta1;  
  real<lower = 0> sigma;  
}  
model {  
  Y ~ normal(beta0 + beta1 * X, sigma);  
  beta0 ~ normal(-2000, 1000);  
  beta1 ~ normal(100, 40);  
  sigma ~ exponential(0.0008);  
}  
  
generated quantities {  
  real y_tilda = normal_rng(beta0 + beta1 * 75, sigma);  
}  
"  
model.bicikli.sim <-  
  stan(model_code = model.bicikli,  
    data = list(n = nrow(bikes), Y = bikes$rides, X = bikes$temp_feel),  
    chains = 4, iter = 5000*2, cores=1, refresh=0, seed = 84735)  
  
print(model.bicikli.sim)
```

Inference for Stan model: anon_model.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
beta0	-2170.44	3.96	329.75	-2817.70	-2392.27	-2172.94	-1944.64	-1534.26
beta1	81.76	0.06	4.72	72.62	78.56	81.79	84.93	90.93
sigma	1283.40	0.44	41.18	1205.40	1254.96	1282.20	1310.99	1366.71
y_tilda	3962.69	9.08	1287.78	1426.45	3102.41	3960.33	4829.04	6504.58
lp__	-3822.26	0.01	1.18	-3825.29	-3822.81	-3821.96	-3821.39	-3820.89

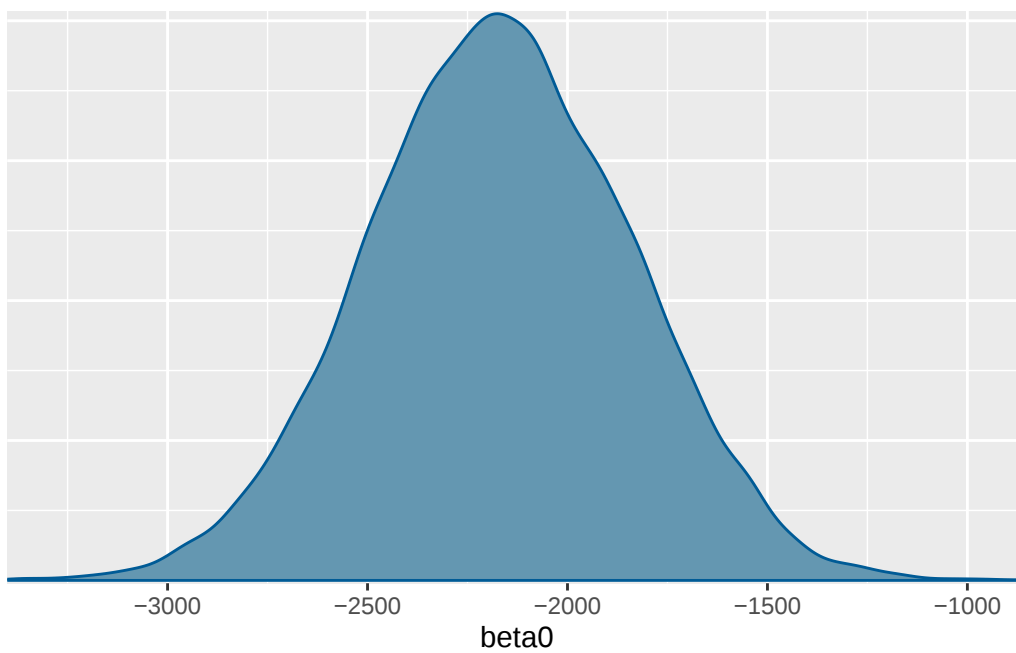
	n_eff	Rhat
beta0	6923	1
beta1	6927	1
sigma	8748	1
y_tilda	20125	1
lp__	6610	1

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:54:52 2024.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

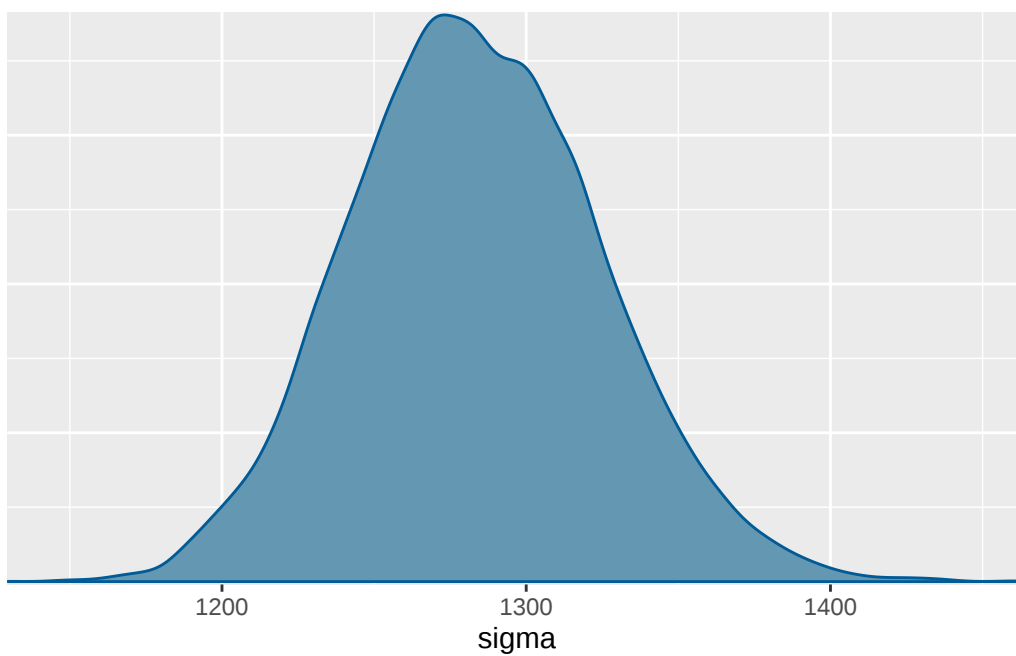
- c) Posteriorne distribucije parametara treba uspoređivati s početnim apriornim uvjerenjem koje smo postavili u modelu. Tako primjerice, za β_0 nakon što vidimo podatke na tipičan dan očekujemo ipak nešto manje vožnji, to jest približno 4830 vožnji, odnosno uzmemo li u obzir neizvjesnost očekujemo između 4183 i 5466 vožnji. Dakle, nakon što vidimo podatke očekujemo manje vožnji, ali uz manju neizvjesnost.

S druge strane za σ : očekujemo nešto veću varijabilnost u broju vožnji na proizvoljan dan (1283), uz 0.05-vjerodostojni interval [1205.4, 1366.71].

```
mcmc_dens(model.bicikli.sim, pars="beta0")
```

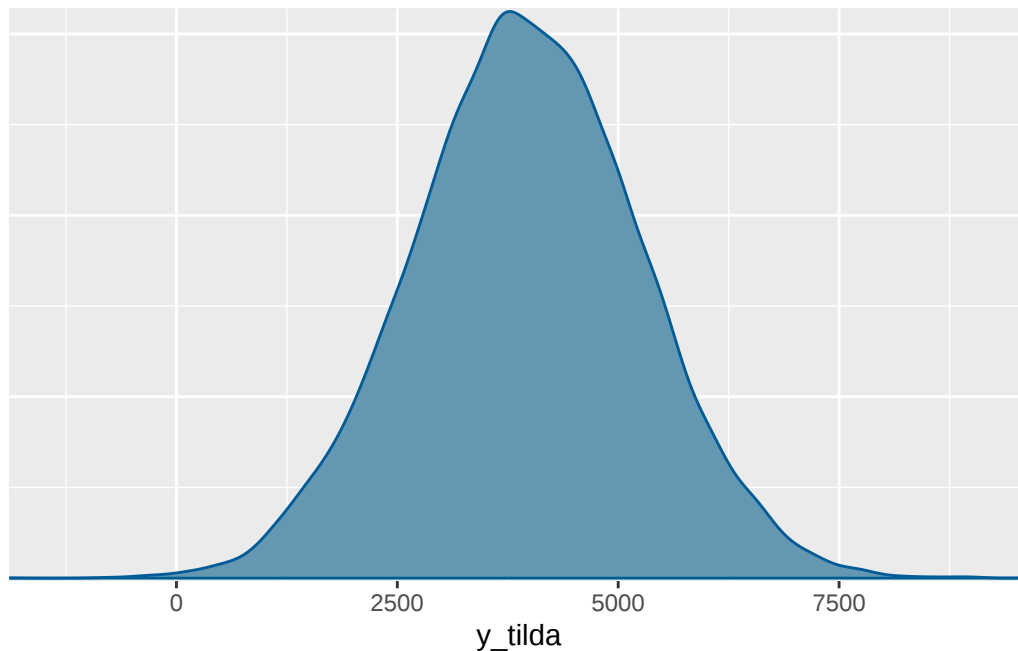


```
mcmc_dens(model.bicikli.sim, pars="sigma")
```



d)

```
mcmc_dens(model.bicikli.sim, pars="y_tilda")
```



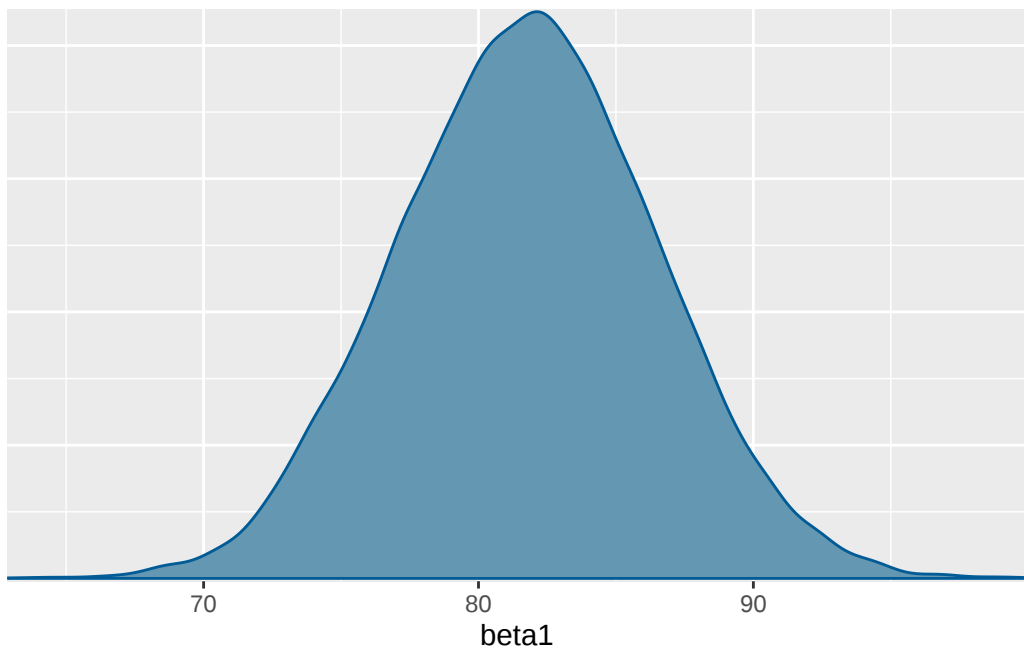
e)

```
MCMC.bicikli <- rstan::extract(model.bicikli.sim)
ecdf(MCMC.bicikli$y_tilda)(5000)-ecdf(MCMC.bicikli$y_tilda)(2500)
```

```
[1] 0.66195
```

f) Odgovor nam daje parametar β_1 , odnosno pripadna posteriorna distribucija. Prema modelu, očekujemo povećanje između 73 i 91 vožnji s 95% vjerojatnosti. Uzmemo li u obzir posteriorno očekivanje parametra β_1 , možemo reći da prema modelu očekujemo približno 82 nove vožnje za svako jedinično povećanje temperature.

```
mcmc_dens(model.bicikli.sim, pars="beta1")
```



112

Zadatak 112.1. U bazi podataka **airbnb_small** nalaze se podaci o smještaju izlistanom na platformi Airbnb u Chicagu iz 2016. godine.

- a) Promotrite podatke o broju recenzija po smještaju. Predložite distribucijski model za podatke.
- b) Predložite i implementirajte model za broj recenzija po smještaju u ovisnosti o ocjeni smještaja koristeći STAN platformu. Prilikom definiranja apriornih distribucija koristite sljedeće informacije:
 - Za tipični novi smještaj očekujemo 10 recenzija, a uzevši u obzir varijabilnost, očekujemo između 5 i 15 recenzija.
 - Za svako povećanje ocjene smještaja u jediničnom iznosu, broj recenzija se multiplikativno poveća za 0.5, a ako uzmemo u obzir neizvjesnost očekujemo multiplikativno povećanje između 0.1 i 0.9.
- c) Interpretirajte posteriornu distribuciju regresijskog parametra β_0 .
- d) Aproksimirajte posteriornu prediktivnu distribuciju broja recenzija po smještaju za slučajno odabrani smještaj ocjenom 5.
- e) Aproksimirajte posteriornu vjerojatnost da za slučajno odabrani smještaj s ocjenom 5 broj pripadnih recenzija bude veći od 30.
- f) Ukoliko se ocjena smještaja poveća za 0.5, koliku promjenu u očekivanom broju recenzija smještaja očekujete?

 Danger

113 Rješenje

b)

```
model.airbnb <- "  
data {  
  int<lower = 0> n;  
  array[n] int Y;  
  vector[n] X;  
  
}  
parameters {  
  real beta0;  
  real beta1;  
}  
model {  
  Y ~ poisson(exp(beta0 + beta1 * X));  
  beta0 ~ normal(2.3, 0.9);  
  beta1 ~ normal(-0.7, 2.56);  
}  
  
generated quantities {  
  real y_tilda = poisson_rng(exp(beta0 + beta1 * 5));  
}  
"  
  
model.airbnb.sim <-  
  stan(model_code = model.airbnb,  
    data = list(n = nrow(airbnb_small), Y = airbnb_small$reviews, X = airbnb_small$ra  
    chains = 4, iter = 5000*2, cores=1, refresh=0, seed = 934)  
  print(model.airbnb.sim)
```

Inference for Stan model: anon_model.

4 chains, each with iter=10000; warmup=5000; thin=1;

post-warmup draws per chain=5000, total post-warmup draws=20000.

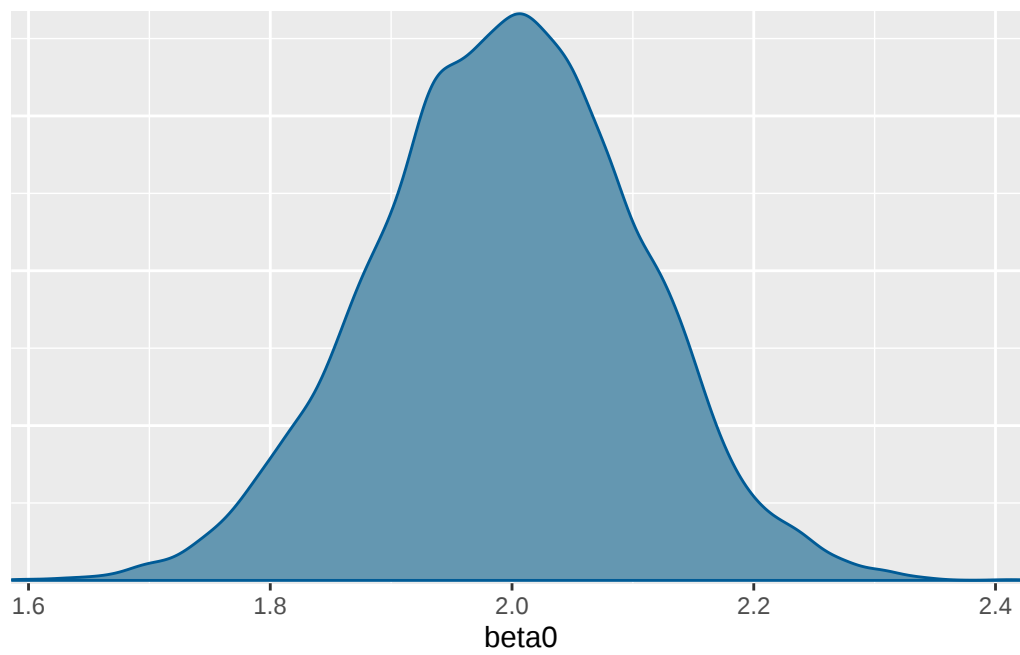
mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
------	---------	----	------	-----	-----	-----	-------

beta0	2.00	0.00	0.11	1.78	1.92	2.00	2.07	2.21
beta1	0.27	0.00	0.02	0.23	0.26	0.27	0.29	0.32
y_tilda	28.70	0.04	5.34	19.00	25.00	29.00	32.00	40.00
lp__	54687.25	0.01	0.97	54684.64	54686.87	54687.54	54687.95	54688.20
	n_eff	Rhat						
beta0	2599	1						
beta1	2599	1						
y_tilda	19754	1						
lp__	4159	1						

Samples were drawn using NUTS(diag_e) at Mon Jun 3 09:57:50 2024.
 For each parameter, n_eff is a crude measure of effective sample size,
 and Rhat is the potential scale reduction factor on split chains (at
 convergence, Rhat=1).

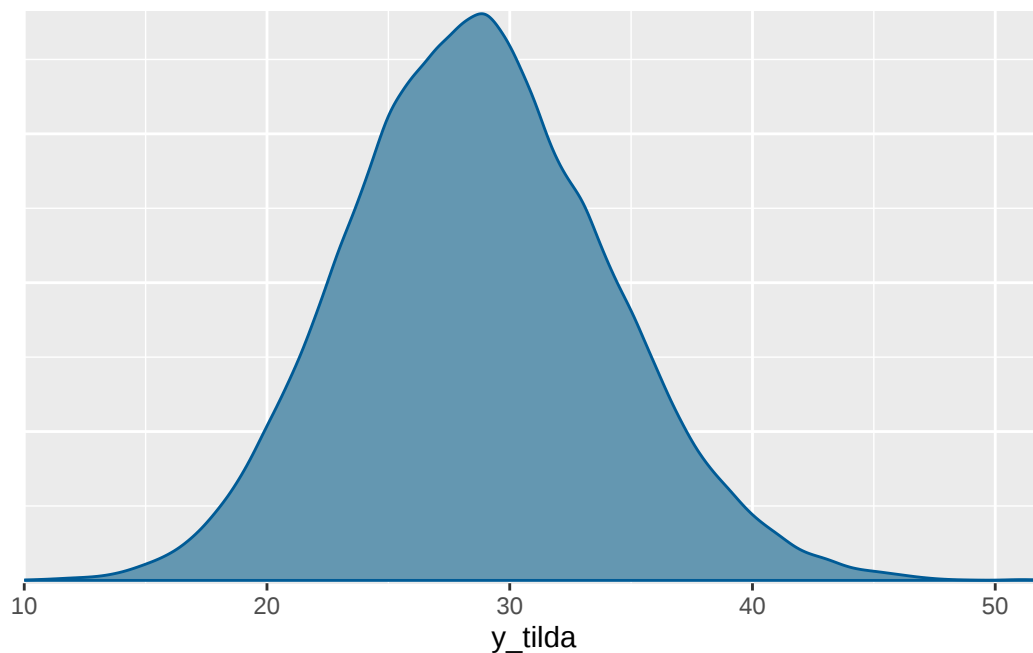
- c) Posteriornu distribucije parametra β_0 treba uspoređivati s početnim apriornim uvjerenjem koje smo postavili u modelu. Tako primjerice, za β_0 nakon što vidimo podatke za tipičan novi smještaj očekujemo ipak nešto manje recenzija, to jest približno 7 recenzija, odnosno uzmemo li u obzir neizvjesnost očekujemo između 6 i 9 recenzija. Dakle, nakon što vidimo podatke očekujemo manje recenzija, ali uz manju neizvjesnost.

```
mcmc_dens(model.airbnb.sim, pars="beta0")
```



d)

```
mcmc_dens(model.airbnb.sim, pars="y_tilda")
```



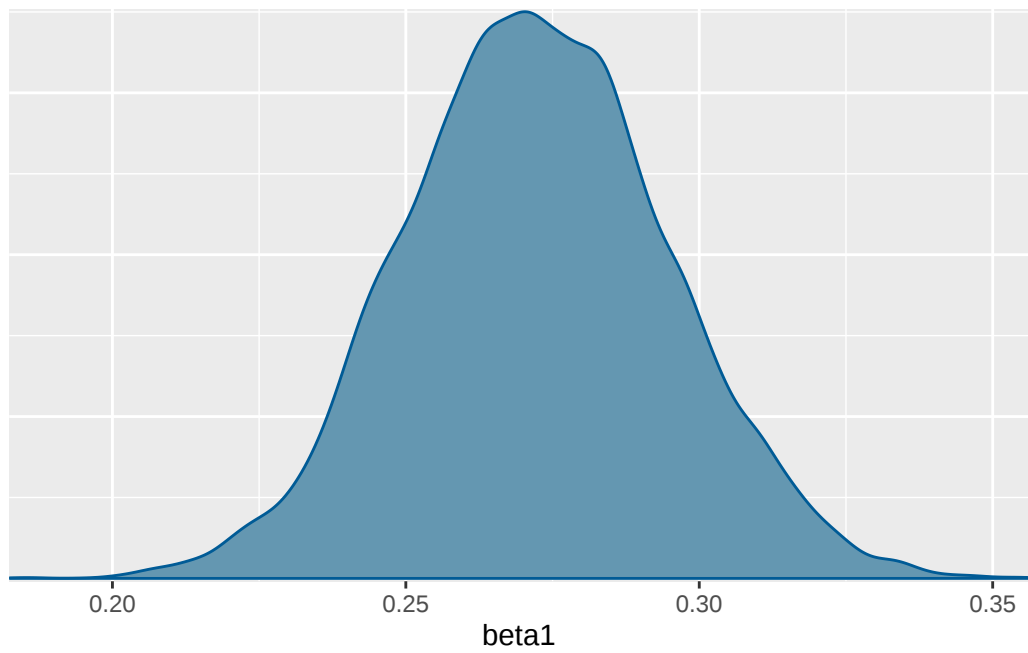
e)

```
MCMC.airbnb <- rstan:: extract(model.airbnb.sim)
1-ecdf(MCMC.airbnb$y_tilda)(30)
```

```
[1] 0.35515
```

f) Odgovor nam daje parametar β_1 , odnosno pripadna posteriorna distribucija. Prema modelu, očekujemo povećanje između 12% i 17% povećanja broja recenzija s 95% vjerojatnosti. Uzmemo li u obzir posteriorno očekivanje parametra β_1 , možemo reći da prema modelu očekujemo približno povećanje u iznosu 14% broja novih recenzija za svako polovično povećanje ocjena smještaja.

```
mcmc_dens(model.airbnb.sim, pars="beta1")
```



114 Vjerojatnosne distribucije

114.1 Beta distribucija

i Note

115 Beta distribucija

Za neprekidnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \langle 0, 1 \rangle$ i gustoćom

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{\langle 0,1 \rangle}(x),$$

gdje je

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

beta funkcija i $\alpha > 0$ i $\beta > 0$ parametri kažemo da je **beta slučajna varijabla** i koristimo oznaku $X \sim \text{Beta}(\alpha, \beta)$. Pripadnu distribuciju zovemo **beta distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dbeta(x, shape= $\alpha$ , rate= $\beta$ )`

Mjere centra i raspršenja:

- Očekivanje: $E[X] = \frac{\alpha}{\alpha + \beta}$
- Varijanca: $Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
- Uočimo: $Var(X) = \frac{EX(1-EX)}{\alpha + \beta + 1}$

115.1 Gama distribucija

 Note

116 Gama distribucija

Za neprekidnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \langle 0, \infty \rangle$ i gustoćom

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{(0, \infty)}(x),$$

gdje su $\alpha > 0$ i $\beta > 0$ parametri kažemo da je **gama slučajna varijabla** i koristimo oznaku $X \sim \Gamma(\alpha, \beta)$. Pripadnu distribuciju zovemo **gama distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dgamma(x, shape= $\alpha$ , rate= $\beta$ )`

Mjere centra i raspršenja:

- Očekivanje: $E[X] = \frac{\alpha}{\beta}$
- Varijanca: $Var(X) = \frac{\alpha}{\beta^2}$

116.1 Inverzna gama distribucija

117 Inverzna gama distribucija

Za neprekidnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \langle 0, \infty \rangle$ i gustoćom

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}} I_{\langle 0, \infty \rangle}(x),$$

gdje su $\alpha > 0$ i $\beta > 0$ parametri kažemo da je **inverzna gama slučajna varijabla** i koristimo oznaku $X \sim \text{InvGama}(\alpha, \beta)$. Pripadnu distribuciju zovemo **inverzna gama distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dinvgamma(x,  $\alpha$ ,  $\beta$ )`. S druge strane, ako $X \sim \Gamma(\alpha, \beta)$ tada $1/X \sim \text{InvGama}(\alpha, \beta)$.

Mjere centra i raspršenja:

- Očekivanje: $E[X] = \frac{\beta}{\alpha-1}$, $\alpha > 1$
- Varijanca: $Var(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, $\alpha > 2$.

117.1 Beta-binomna distribucija

118 Beta-binomna distribucija

Za diskretnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \{0, 1, \dots, n\}$ i gustoćom

$$f(x|n, \alpha, \beta) = \binom{n}{x} \frac{B(x + \alpha, n - x + \beta)}{B(\alpha, \beta)},$$

gdje su $n \in \mathbb{N}_0$, $\alpha > 0$ i $\beta > 0$ parametri kažemo da je **beta-binomna slučajna varijabla** i koristimo oznaku $X \sim \text{BetaBin}(n, \alpha, \beta)$. Pripadnu distribuciju zovemo **beta-binomna distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dbinom(x, alpha, beta)`.

Mjere centra i raspršenja:

- Očekivanje: $E[X] = \frac{n\alpha}{\alpha + \beta}$
- Varijanca: $Var(X) = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

118.1 Negativna-binomna distribucija

119 Negativna-binomna distribucija

Za diskretnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \mathbb{N}_0$ i gustoćom

$$f(x|r, \beta) = \binom{x+r-1}{x} \left(\frac{\beta}{1+\beta} \right)^x \cdot \left(\frac{1}{1+\beta} \right)^r,$$

gdje su $r > 0$ i $\beta > 0$ parametri kažemo da je **negativna-binomna slučajna varijabla** i koristimo oznaku $X \sim \text{NegBin}(r, \beta)$. Pripadnu distribuciju zovemo **negativna-binomna distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dnbinom(x, size=r, prob= $\frac{\beta}{1+\beta}$ )`.

Mjere centra i raspršenja:

- Očekivanje: $E[X] = r\beta$
- Varijanca: $\text{Var}(X) = r\beta(1 + \beta)$

119.1 Lomax distribucija

120 Lomax distribucija

Za neprekidnu slučajnu varijablu X sa slikom $\mathcal{R}(X) = \langle 0, \infty \rangle$ i gustoćom

$$f(x|\alpha, \lambda) = \frac{\alpha}{\lambda} \left(1 + \frac{x}{\lambda}\right)^{-(\alpha+1)} I_{\langle 0, \infty \rangle}(x),$$

gdje su $\alpha > 0$ i $\lambda > 0$ parametri kažemo da je **Lomax slučajna varijabla** i koristimo oznaku $X \sim Lomax(\alpha, \lambda)$. Pripadnu distribuciju zovemo **Lomax distribucija**.

U R-u pripadnu gustoću dohvaćamo s `dlomax(x,  $\frac{1}{\lambda}$ ,  $\alpha$ )`.

Mjere centra i raspršenja:

- Očekivanje: $E[X] = \frac{\lambda}{\alpha-1}$
- Varijanca: $Var(X) = \frac{\lambda^2 \alpha}{(\alpha-1)^2 (\alpha-2)}, \quad \alpha > 2$
- Medijan: $\lambda (\sqrt[3]{2} - 1)$

Reference

- Johnson, Ott, and Dogucu (2022)
- Hoff (2009)
- Bolstad (2007)
- Andrew Gelman (2004)
- Albert (2009)
- Albert, Jim. 2009. *Bayesian Computation with r*. 2nd ed. Use r! Springer.
- Andrew Gelman, Hal S. Stern, John B. Carlin. 2004. *Bayesian Data Analysis*. 2nd ed. Texts in Statistical Science. Chapman & Hall/CRC.
- Bolstad, William M. 2007. *Introduction to Bayesian Statistics*. 2nd ed. John Wiley.
- Dobrow, Robert P. 2016. *Introduction to Stochastic Processes with r*. John Wiley & Sons.
- Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods*. Vol. 580. Springer.
- Johnson, Alicia A, Miles Q Ott, and Mine Dogucu. 2022. *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. Chapman; Hall/CRC.