Bayesova statistika

Domača naloga 1

Sara Bizjak | 27202020

Oktober 2021

1. naloga

Na spodnjih grafih so preizkušani različni parametri α in β. Na prvem grafu lahko vidimo obe porazdelitvi za apriorno porazdelitev `Beta(1, 1)`, nadalje pa sta v isti vrsti prikazana grafa za oba para izbranih parametrov (α, β) in (β, α).

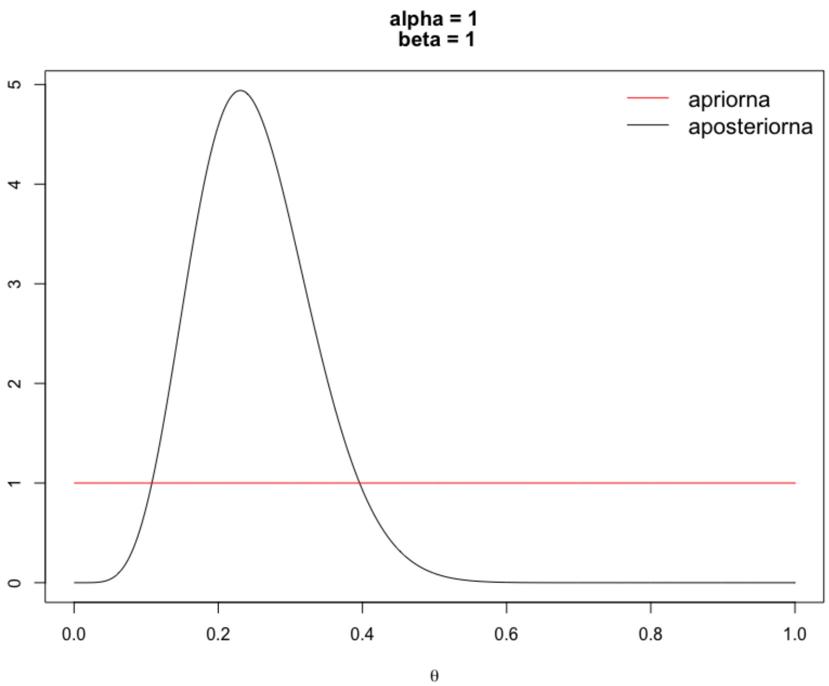

Figure 1: Graf z apriorno in aposteriorno porazdelitev za $\alpha = \beta = 1$.

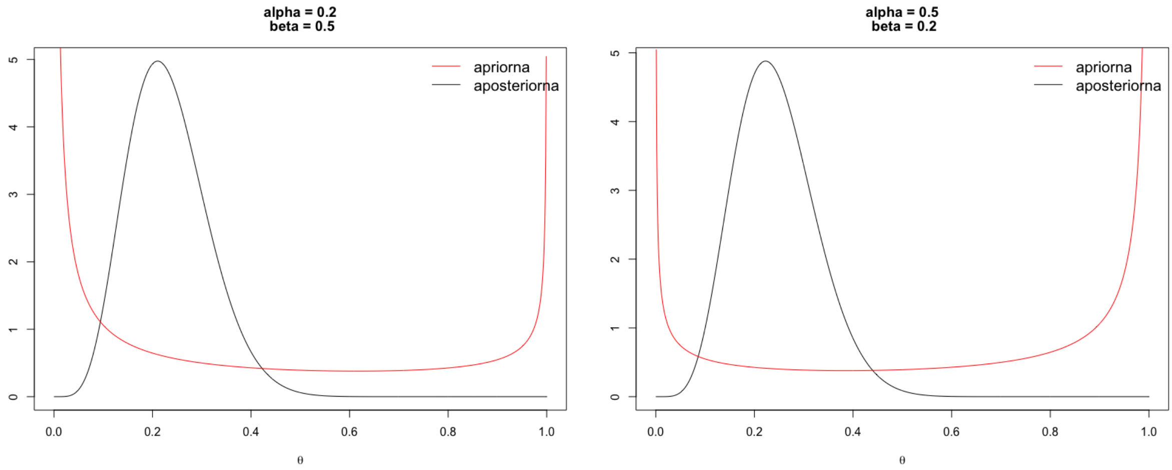

Figure 2: Grafa z apriorno in aposteriorno porazdelitvijo z vrednostma 0.2 in 0.5.

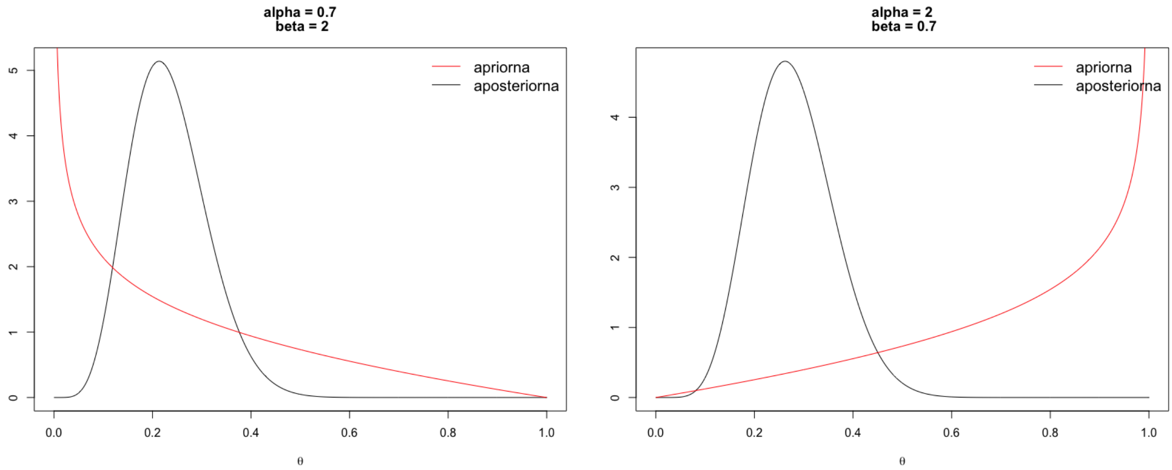

Figure 3: Grafa z apriorno in aposteriorno porazdelitvijo z vrednostma 0.7 in 2.

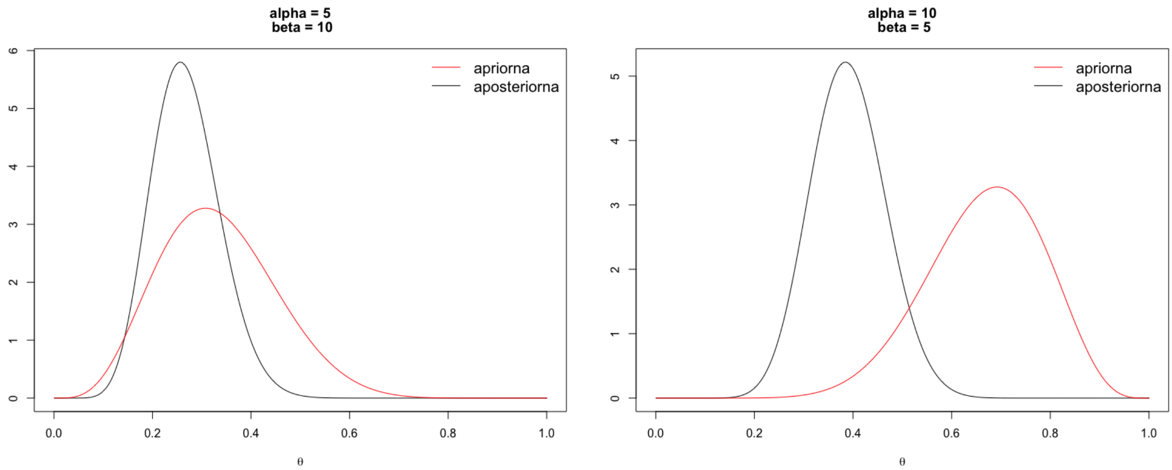

Figure 4: Grafa z apriorno in aposteriorno porazdelitvijo z vrednostma 5 in 10.

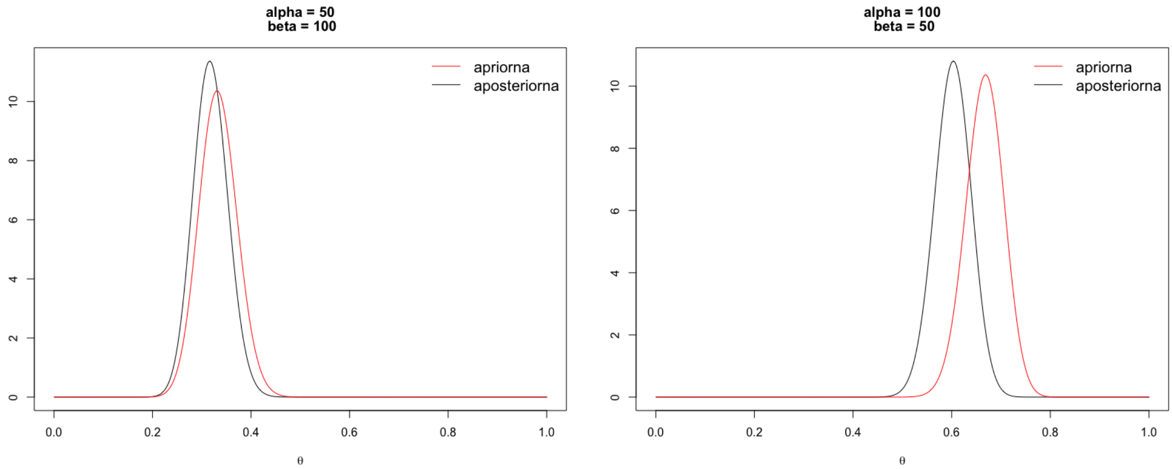

Figure 5: Grafa z apriorno in aposteriorno porazdelitvijo z vrednostma 50 in 100.

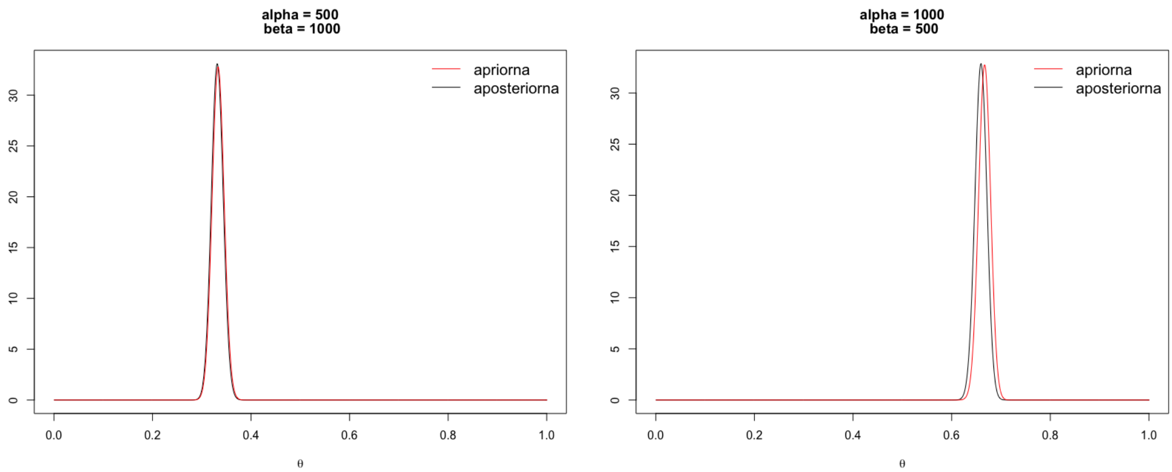

Figure 6: Grafa z apriorno in aposteriorno porazdelitvijo z vrednostma 500 in 1000.

Ugotovitve in opazke:

- Za α in β manjša od 1 je apriorna porazdelitev konveksne oblike.
- Če je ali α ali pa β manjši od 1, predstavlja graf apriorne funkcije "polovico konveksne oblike" iz prejšnjega primera.
- Za α in β večja od 1 je graf apriorne funkcije konkaven.
- Graf aposteriorne funkcije je v vseh primerih konkaven.
- Hitro opazimo, da sta si grafa funkcij apriorne porazdelitve na levi in desni sliki v vsaki vrstici simetrični glede na os $x = 0.5$ (torej če zamenjamo vrednosti za α in β), aposterirno porazdelitev pa se, zaradi premika apriorne, zamakne v isto smer, kamor se preslika aposteriorna, torej v vseh teh primerih bolj v desno.

- Z večanjem α in β hkrati opazimo, da sta si tudi obe porazdelitvi na grafu bližje (vrha sta vedno bolj "enotna"), saj na takšen način simuliramo večje zaupanje v predhodno znanje in bolj verjamemo apriorni porazdelitvi.

- Podobno kot prej sta si z manjšanjem α in β porazdelitvi na grafu različni in bolj narazen.

2. naloga

Vemo, da je pričakovana vrednost apriorne porazdlitve enaka 0.25 in v skladu s tem izberemo primerna α in β.
Rešujemo torej enačbo z dvema neznankama

$$\begin{aligned} E_{aprior} = \frac{1}{4} &= \frac{\alpha}{\alpha+\beta} \\ \frac{1}{4}\alpha + \frac{1}{4}\beta &= \alpha \\ \frac{1}{4}\beta &= \frac{3}{4}\alpha \\ \beta &= 3\alpha \end{aligned}$$

zato lahko parametra α in β izberemo na neskončno načinov tako, da ustrezata linearni zvezi $\beta = 3\alpha$.
Na spodnjih grafih je prikazanih nekaj takih možnosti za izbiro parametrov α in β.

Oceno pričakovane vrednosti lahko izrazimo kot pričakovano vrednost za aposteriorno porazdelitev:

$$\begin{aligned} \frac{\alpha+k}{\alpha+k+\beta+n-k} &= \frac{\alpha+k}{\alpha+\beta+n} \\ &= \frac{(\alpha+\beta)\frac{\alpha}{\alpha+\beta} + n\frac{k}{n}}{\alpha+\beta+n} \\ &= \frac{\alpha+\beta}{\alpha+\beta+n}\cdot\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n}\cdot\frac{k}{n}, \end{aligned}$$

kar je konveksna kombinacija števil $\frac{\alpha}{\alpha+\beta}$ in $\frac{k}{n}$.
Spomnimo se, da je $E_{aprior} = \frac{\alpha}{\alpha+\beta}$, izraz $\frac{k}{n}$ pa predstavlja frekventistično oceno. Označimo $\gamma(\alpha,\beta) = \alpha+\beta$. Tedaj lahko zgornjo enačbo zapišemo kot

$$E_{apost} = \frac{\gamma(\alpha,\beta)}{\gamma(\alpha,\beta)+n}\cdot E_{aprior} + \frac{n}{\gamma(\alpha,\beta)+n}\cdot\frac{k}{n}$$

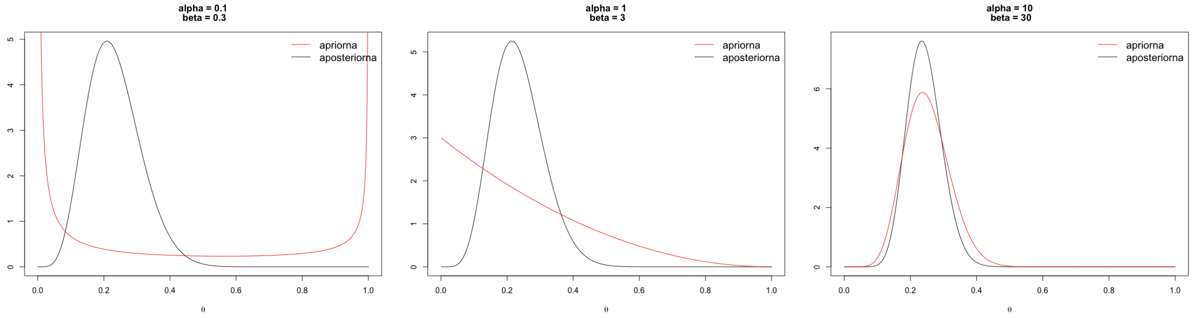

Figure 7: Grafi z apriorno in aposteriorno porazdelitvijo za parametra, ki zadoščata $\beta = 3\alpha$.

Ocene pričakovanih vrednosti za vse tri grafe, po vrsti:

```
[1] 0.2310606
[2] 0.2333333
[3] 0.2424242
```

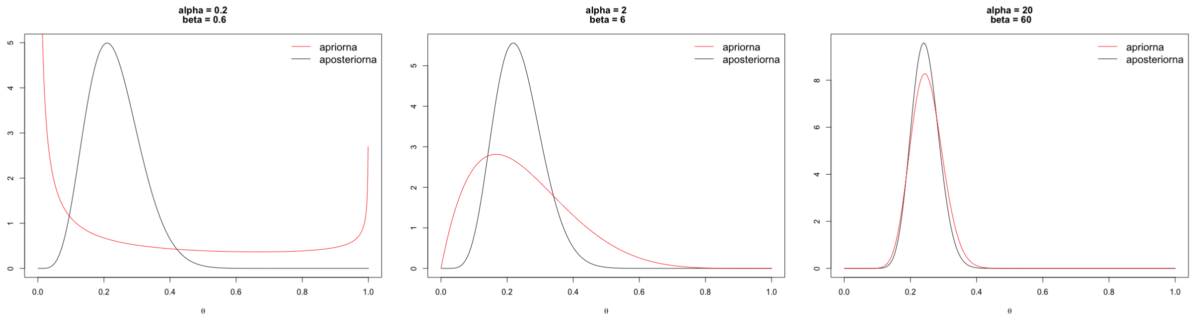

Figure 8: Grafi z apriorno in aposteriorno porazdelitvijo za parametra, ki zadoščata $\beta = 3\alpha$.

Ocene pričakovanih vrednosti za vse tri grafe, po vrsti:

```
[1] 0.2313433
[2] 0.2352941
[3] 0.245283
```

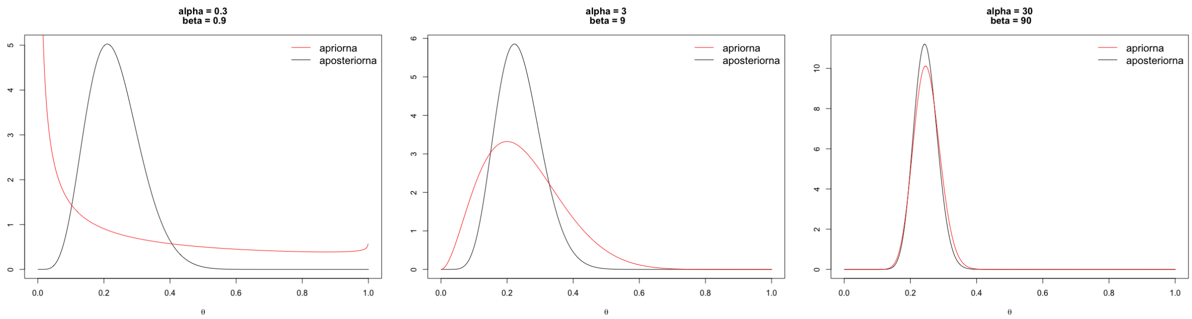

Figure 9: Grafi z apriorno in aposteriorno porazdelitvijo za parametra, ki zadoščata $\beta = 3\alpha$.

Ocene pričakovanih vrednosti za vse tri grafe, po vrsti:

```
[1] 0.2316176
[2] 0.2368421
```

[3] 0.2465753

Nadalje opazimo, da s funkcijo $\gamma(\alpha, \beta)$ kalibriramo verjetje v apriorno porazdelitev: z večanjem $\gamma(\alpha, \beta)$ bolj verjamemo apriorni porazdelitvi (takrat je manjša disperzija). Konkretno za ta primer: z večanjem $\gamma(\alpha, \beta)$ se ocena približuje vrednosti $\frac{1}{4} = 0.25$. To je razvidno tudi iz priloženih grafov.

3. naloga
Vzemimo sedaj vzorec študentov velikosti 30 izmed katerih jih je 21 odgovorilo pravilno. Privzamemo apriorno porazdelitev `Beta(1, 1)` in izračunajmo aposteriorno porazdelitev. Prikažimo jo na grafu.

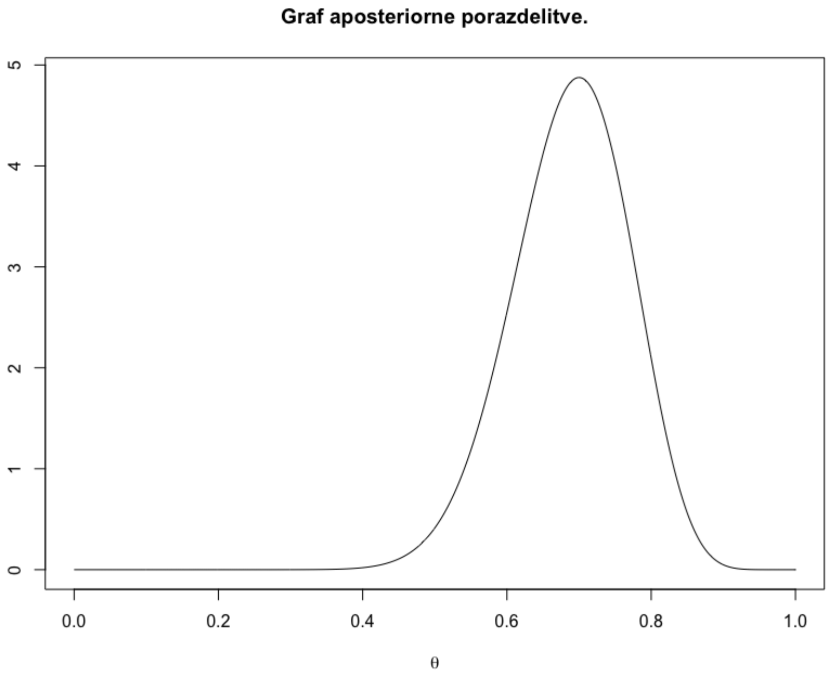

Figure 10: Graf aposteriorne porazdelitve pri danih podatkih.

4. naloga
Primerjajmo aposteriorno porazdelitev iz naloge 3, označimo jo Z_1, z aposteriorno porazdelitvijo porazdelitve `Beta(7, 21)`, ki smo jo izračunali na vajah. Ocenimo verjetnost $P(Z_2 < Z_1)$.
Ocenjeno verjetnost izračunamo s cenilko na tak način, da primerjamo Z_1 in Z_2 po elementih. Natančneje, za vsak istoležni element preverimo ali je element iz porazdelitve Z_2 manjši od elementa iz porazdelitve Z_1. Vsoto oz. seštevek za ta pogoj na koncu delimo s 10000 (toliko kot je generiranih vrednosti v obeh porazdelitvah).
Ocenjena verjetnost znaša:
[1] 0.9998
Na podlagi simulacije izračunajmo še 95% interval zaupanja, poročamo kar 2.5% in 97.5% kvan-

til simuliranih podatkov.
Izračunajmo najprej interval zaupanja za vsako od obeh aposteriornih porazdelitev.
Z_1:
```
2.5% 97.5%
0.5214762 0.8337051
```
Z_2:
```
2.5% 97.5%
0.1110932 0.4164294
```

Za interval zaupanja na podlagi simulacije za primerjanje obeh porazdelite vzamemo kar razliko (po komponentah) vektorjev s 10000 generiranimi vrednostmi iz obeh aposteriornih porazdelitev, kar je sorodno zgornji cenilki, s pomočjo katere smo prišli do ocene verjetnosti. Dobimo:
```
2.5% 97.5%
0.2024990 0.6452275
```