1. DN - Bayesova Statistika

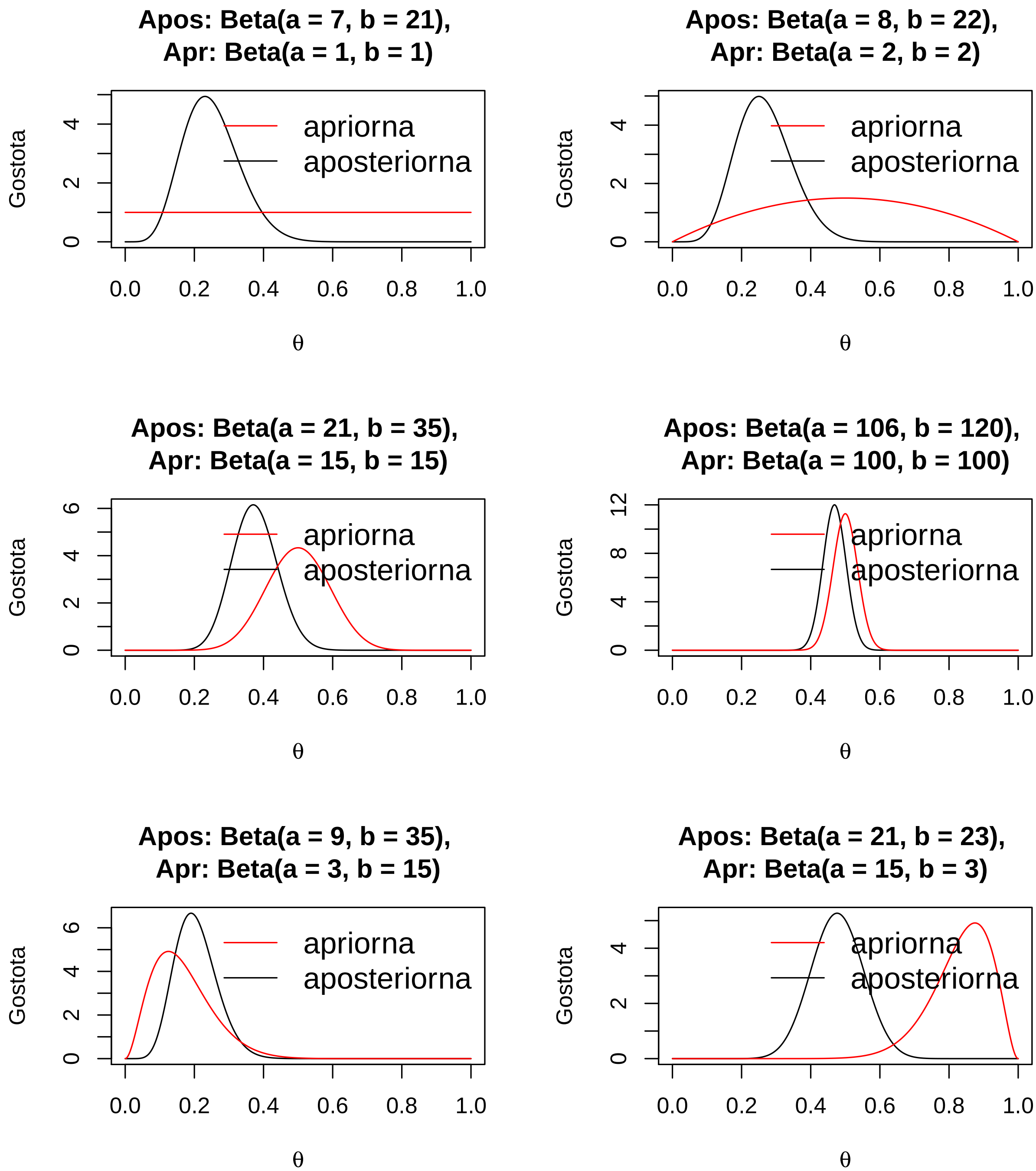
Jan Crne

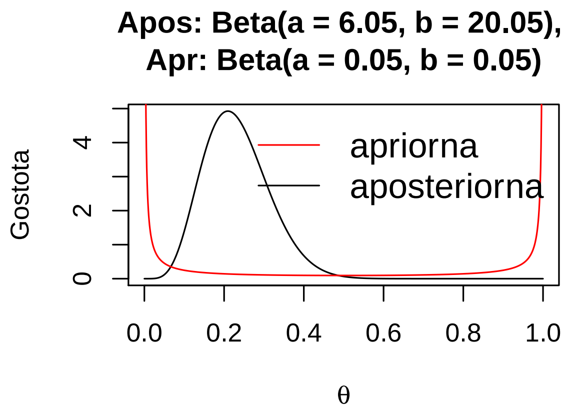
0.1 Prva Naloga

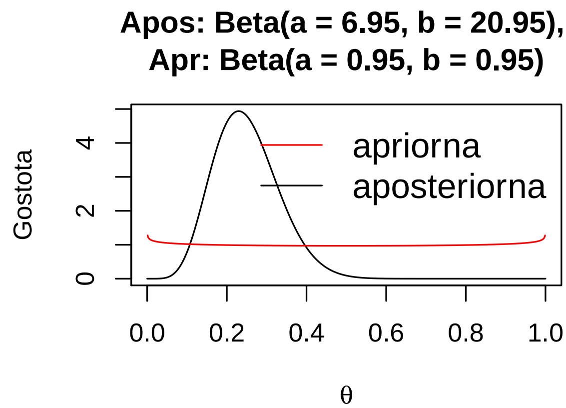
NAVODILO: Preizkusite razlicne informativne apriorne beta porazdelitve za primer, ki smo ga imeli na prvi vaji. Pri spreminjanju alfe in bete poskusite razlicne kombinacije glede na to, ali sta (oba) parametra vecja oziroma manjsa od 1. Kaj se zgodi, ce ju zamenjamo? Kaj se zgodi, ce oba povecamo? Pri vsaki razlicici apriorne porazdelitve narisite graf, na katerem sta narisani apriorna in aposteriorna porazdelitev. Povzemite obnasanje apriorne in aposteriorne porazdelitve v nekaj stavkih.

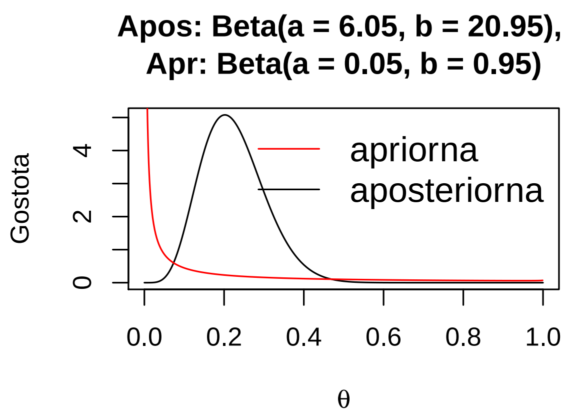
```
library(glue) # za uporabo potreben paket z imenom "glue"

n <- 26 # stevilo vprasanih studentov
k <- 6 # stevilo studentov, ki so odgovorili pravilno

theta <- seq(0, 1, 0.001)

alpha.apr <- c(1, 2, 15, 100, 3, 15, 0.05, 0.95, 0.05, 0.95, 0.5, 15)
beta.apr <- c(1, 2, 15, 100, 15, 3, 0.05, 0.95, 0.95, 0.05, 15, 0.5)
alpha.apos <- alpha.apr + rep(k, length(alpha.apr))
beta.apos <- beta.apr + rep(n-k, length(beta.apr))
```

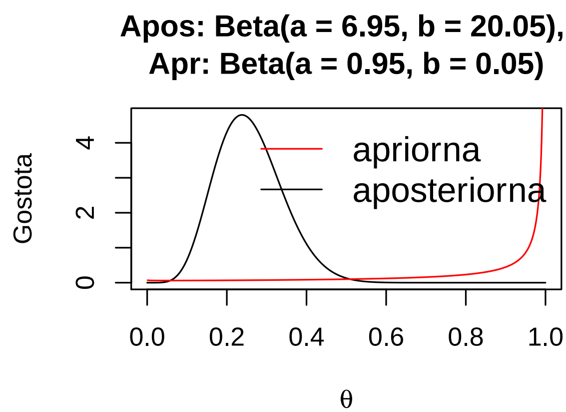
```
for (i in 1:length(alpha.apr)) {
  plot(theta, dbeta(theta, alpha.apos[i], beta.apos[i]),
       type='l', ylab = 'Gostota', xlab=expression(theta),
       main = glue("Apos: Beta(a = {alpha.apos[i]}, b = {beta.apos[i]}),
                   Apr: Beta(a = {alpha.apr[i]}, b = {beta.apr[i]})"))
  lines(theta, dbeta(theta, alpha.apr[i], beta.apr[i]),
        col = 'red')
  legend("topright", legend = c("apriorna","aposteriorna"),
       col = c("red","black"), lty = 1, bty = "n", cex = 1.3)
}
```

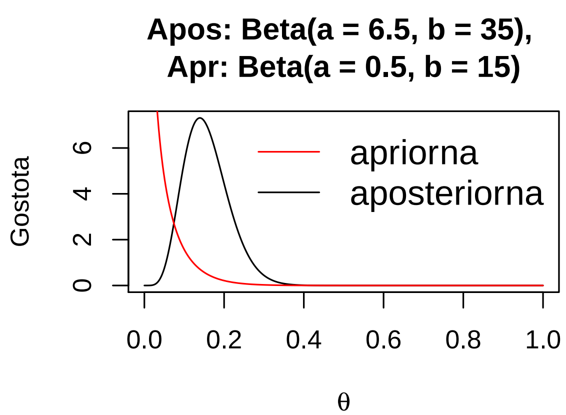

Apos: Beta(a = 7, b = 21),
Apr: Beta(a = 1, b = 1)
apriorna
aposteriorna
Gostota
θ
Apos: Beta(a = 8, b = 22),
Apr: Beta(a = 2, b = 2)
apriorna
aposteriorna
Gostota
θ
Apos: Beta(a = 21, b = 35),
Apr: Beta(a = 15, b = 15)
apriorna
aposteriorna
Gostota
θ
Apos: Beta(a = 106, b = 120),
Apr: Beta(a = 100, b = 100)
apriorna
aposteriorna
Gostota
θ
Apos: Beta(a = 9, b = 35),
Apr: Beta(a = 3, b = 15)
apriorna
aposteriorna
Gostota
θ
Apos: Beta(a = 21, b = 23),
Apr: Beta(a = 15, b = 3)
apriorna
aposteriorna
Gostota
θ

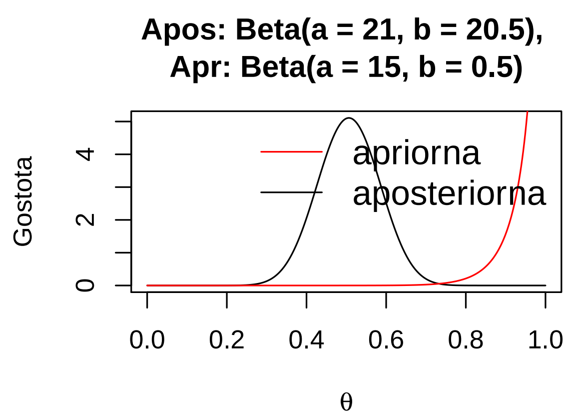












KOMENTAR: Z zvisevanjem parametrov (< 1), postajata tako apriorna kot aposteriorna porazdeitev manj razprseni, interval na katerem je zajeta vecina verjetnejsih moznih vrednosti parametra se skrci. Za res velike vrednosti parametrov npr. α=100, β=100, bosta tako apriorna, kot aposteriorna porazdelitev zelo podobni normalni porazdelitvi, prav tako sta si blizu skupaj.

Pri parametrih nizjih od ena, imamo v aposteriorni porazdelitvi vecjo verjetnost, za obstoj paramterov na robovih intervala, torej blizje 0 oz. 1, na aposteriorno verjetnost pa to ne vpliva pretirano, saj so vrednosti za katere je porazdelitev "premaknjena" zanemarljive v primerjavi z velikostjo nasega vzorca.

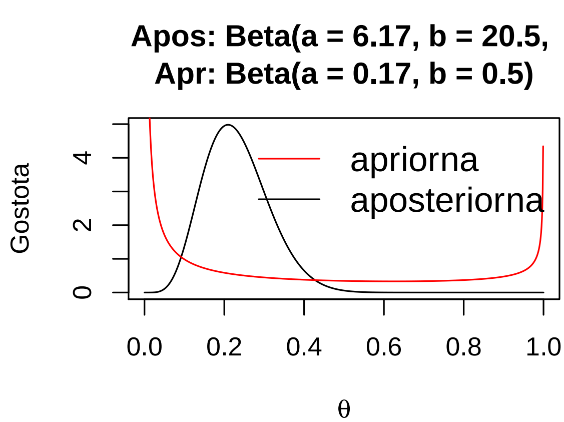
ce parametra α in β v apriorni porazdelitvi zamenjamo, dobimo za apriorno porazdelitev ravno zrcalno sliko, kar je razvidno ze iz formule za gostoto Beta porazdelitve. To pomeni, da bo za velike α in nizje β verjetnost, nasih uspehov visja in obratno. V enake smeri se premika tudi aposteriorna verjetnost.

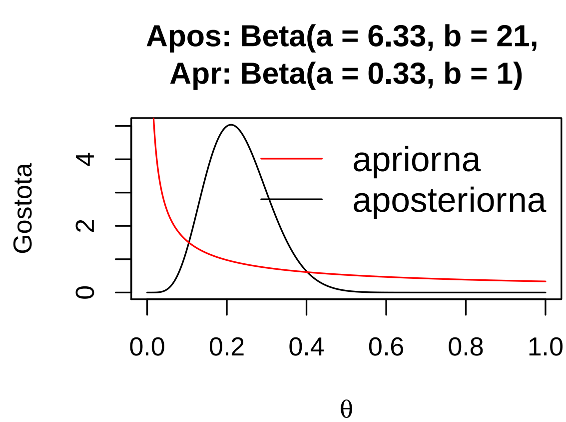
0.2 Druga naloga

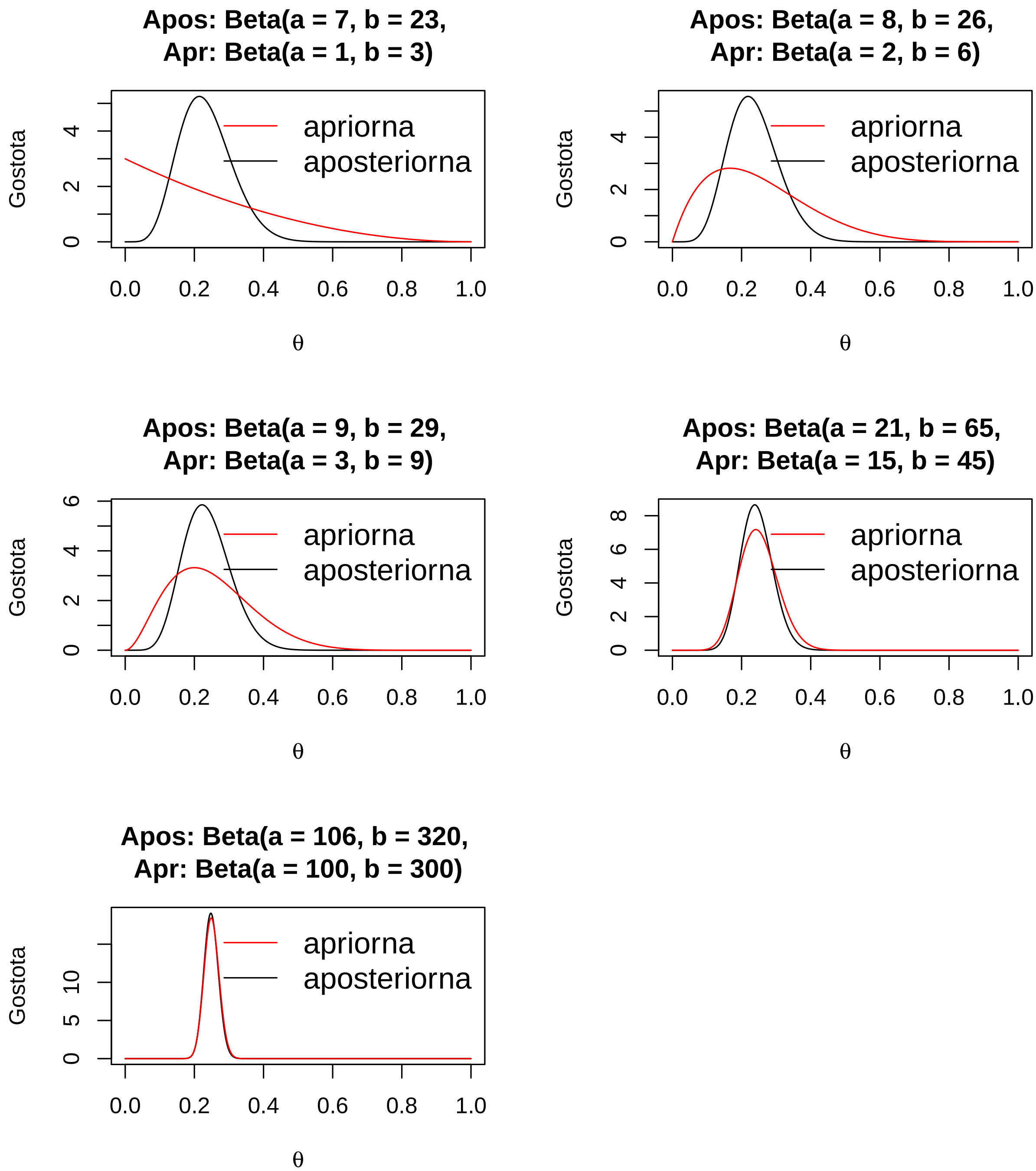
NAVODILO: Izberite alfa in beta tako, da bo pricakovana vrednost apriorne porazdelitve enaka 0,25. Ali lahko to naredite na vec nacinov? Ce lahko, potem preizkusite nekaj smiselnih moznosti (vsaj dve) glede na to, ali bolj ali manj verjamete vasemu prepricanju. Tudi tu pri vsaki razlicici apriorne porazdelitve narisite graf, na katerem sta narisani apriorna in aposteriorna porazdelitev. Poleg tega izracunajte se oceno pricakovane vrednosti. Napisite kratek komentar.

```
alpha.apr2 <- c(1/6, 1/3, 1, 2, 3, 15, 100)
beta.apr2 <- 3*alpha.apr2
alpha.apos2 <- alpha.apr2 + rep(k, length(alpha.apr2))
beta.apos2 <- beta.apr2 + rep(n-k, length(beta.apr2))
```

```
for (i in 1:length(alpha.apr2)) {
  plot(theta, dbeta(theta, alpha.apos2[i], beta.apos2[i]),
       type='l', ylab = 'Gostota', xlab=expression(theta),
       main = glue("Apos: Beta(a = {round(alpha.apos2[i], 2)}, b = {round(beta.apos2[i],
                        Apr: Beta(a = {round(alpha.apr2[i], 2)}, b = {round(beta.apr2[i], 2)}
  lines(theta, dbeta(theta, alpha.apr2[i], beta.apr2[i]),
        col = 'red')
  legend("topright", legend = c("apriorna","aposteriorna"),
       col = c("red","black"), lty = 1, bty = "n", cex = 1.3)
}
```








```
pric.vr.apr <- rep(0, length(alpha.apr2))
for (i in 1:length(alpha.apr2)) {
  pric.vr.apr[i] <- (alpha.apr2[i] + k) / (alpha.apr2[i] + beta.apr2[i] + n)
}

pric.vr <- data.frame("Alpha" = round(alpha.apr2, 2),
                      "Beta" = round(beta.apr2, 2),
```

```
                        "Pric. vrednost" = round(pric.vr.apr, 4))

knitr::kable(pric.vr, caption = "Pric. Vr. Aposteriorne Porazdelitve")
```

Table 1: Pric. Vr. Aposteriorne Porazdelitve

Alpha	Beta	Pric..vrednost
0.17	0.5	0.2312
0.33	1.0	0.2317
1.00	3.0	0.2333
2.00	6.0	0.2353
3.00	9.0	0.2368
15.00	45.0	0.2442
100.00	300.0	0.2488

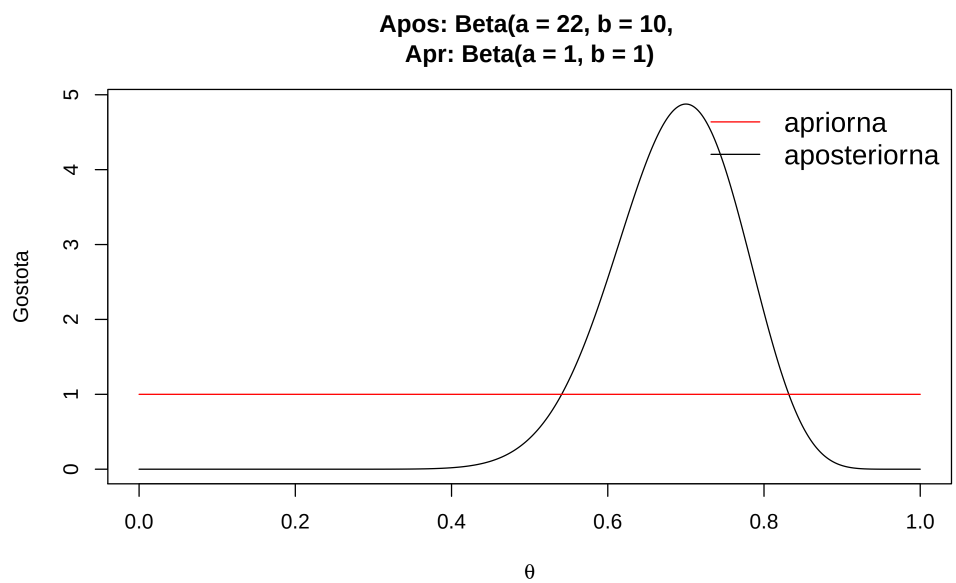
KOMENTAR: Iz podane informacije o upanju apriorne porazdelitve, dobimo za parametre naslednjo zvezo $3*\alpha = \beta$. Iz tabele matematicnih upanj v odvisnosti od parametrov, vidimo, da ob visanju parametra α in posledicno β, matematicno upanje cedalje bolj priblizuje 1/4, ki je ravno nasa podana zunanja informacija. Logika za tem je, da je ob vecjih parametrih Beta porazdelitev bolj podobna normalni porazdelitvi, katere upanje je kar enako upanju nase Beta porazdelitve.

0.3 Tretja naloga

NAVODILO: Denimo, da vzamemo nov vzorec studentov velikosti 30, ki so tudi odgovorili na prvotno vprasanje iz prve vaje. Izmed podanih 30 odgovorov je bilo 21 pravilnih. Privzemite neinformativno apriorno porazdelitev (Beta(1,1)) in izracunajte aposteriorno porazdelitev.

```
n2 <- 30 # novo stevilo vprasanih studentov
k2 <- 21 # stevilo studenotv, ki so odgovorili pravilno v novem vzorcu

alpha.apr3 <- 1
beta.apr3 <- 1
alpha.apos3 <- alpha.apr3 + k2
beta.apos3 <- beta.apr3 + n2 - k2

Z1 <- dbeta(theta, alpha.apos3, beta.apos3)

plot(theta, Z1, type='l', ylab = 'Gostota', xlab=expression(theta),
main = glue("Apos: Beta(a = {round(alpha.apos3, 2)}, b = {round(beta.apos3, 2)},
                   Apr: Beta(a = {round(alpha.apr3, 2)}, b = {round(beta.apr3, 2)})"))
```

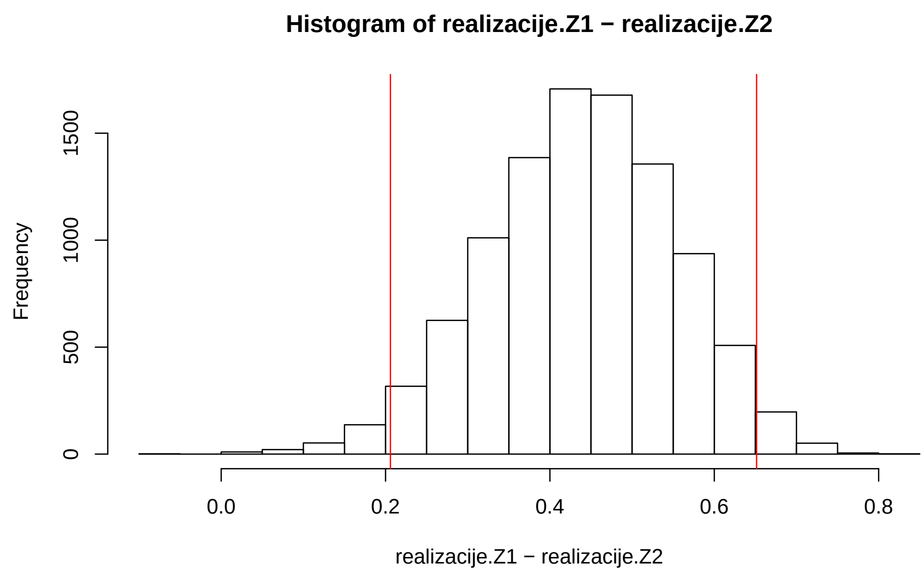
```
lines(theta, dbeta(theta, alpha.apr3, beta.apr3),
      col = 'red')
legend("topright", legend = c("apriorna","aposteriorna"),
     col = c("red","black"), lty = 1, bty = "n", cex = 1.3)
```



0.4 cetrta naloga

NAVODILO: zelimo primerjati aposteriorno porazdelitev iz tocke 3 (oznacimo jo Z1) z aposteriorno porazdelitvijo, ki smo jo izracunali na vajah (Beta(7, 21) oznacimo jo Z2). Za ta namen lahko izracunamo verjetnost P(Z2 < Z1). To verjetnost lahko izracunamo direktno, lahko pa jo tudi ocenimo s pomocjo simulacije. Generirajte 10 000 vrednosti iz obeh aposteriornih porazdelitev (pomagajte si s funkcijo rbeta) na podlagi katerih ocenite verjetnost P(Z2 < Z1). Koliko znasa ocenjena verjetnost? Zapisite cenilko, ki ste jo uporabili. Izracunajte se 95% interval zaupanja na podlagi simulacije - kot 95% interval zaupanja lahko porocate 2,5% in 97,5% kvantil simuliranih podatkov (funkcija quantile).

```
realizacije.Z1 <- rbeta(10000, alpha.apos3, beta.apos3)
realizacije.Z2 <- rbeta(10000, alpha.apos[1], beta.apos[1])

iskana.verjetnost <- sum(realizacije.Z2 < realizacije.Z1) / 10000

kvantila <- quantile(realizacije.Z1 - realizacije.Z2, c(0.025, 0.975))
hist(realizacije.Z1 - realizacije.Z2)
abline(v = kvantila[1], col = "red")
abline(v = kvantila[2], col = "red")
```



KOMENTAR: Iz simulacije ocenjena verjetnost znasa 0.9998.

Za cenilko sem uporabil kar primerjavo med realizacijami. Kjer je veljalo, da je bila i-ta realizacija $Z2$ nizja od i-te realizacije $Z1$, sem realizacijama priredil vrednost 1, nato sem enostavno prestel stevilo vseh enic in jih delil z celotnim stevilom opravljenih ponovitev poskusa torej z 10000.

Iz izracuna spodnjega in zgornjega kvantila sem dobil, da je 2.5% vseh realizacij slucajne spremenljivke $Z1 - Z2$ nizje od 0.2052573 in 2.5% visjih od 0.6456657. To ponazarja tudi histogram realizacij.