Domača naloga 1

Bayesova statistika

Alen Kahteran

29. 11. 2020

Naloga 1

Definicija funkcij in začetnih parametrov

Za podatke sem določil enake, kot smo jih imeli na vajah in to sta

$$n = 26 \quad in \quad k = 6\,,$$

kjer n predstavlja število poskusov in k predstavlja število uspehov. Sta vnaprej določena parametra s katerima, bomo risali funkcijo verjetja, apriorno ter aposteriorno porazdelitev za parameter θ kjer velja

$$\Pr(X = k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}\,.$$

Funkcija verjetja bo na vseh slikah enaka, saj je določena s prej napisano enačbo.

```
# initial parameters
n <- 26
k <- 6

# likelihood function
verjetje <- function(theta, k, n){
  dbinom(k, size = n, prob = theta)
}
```

Pomembno je da funkcijo `verjetje()` množimo s konstanto, tako da bo integral te funkcije vedno enak 1. Zato jo pomnozimo s funkcijo `konst()`.

```
# multiplying with konst we achieve that the integral of likelihood (on theta) is equal to
# 1
konst <- function(k, n){
  theta <- seq(0.001, 1, 0.001)
  1 / (0.001 * sum(verjetje(theta, k, n)))
}
```

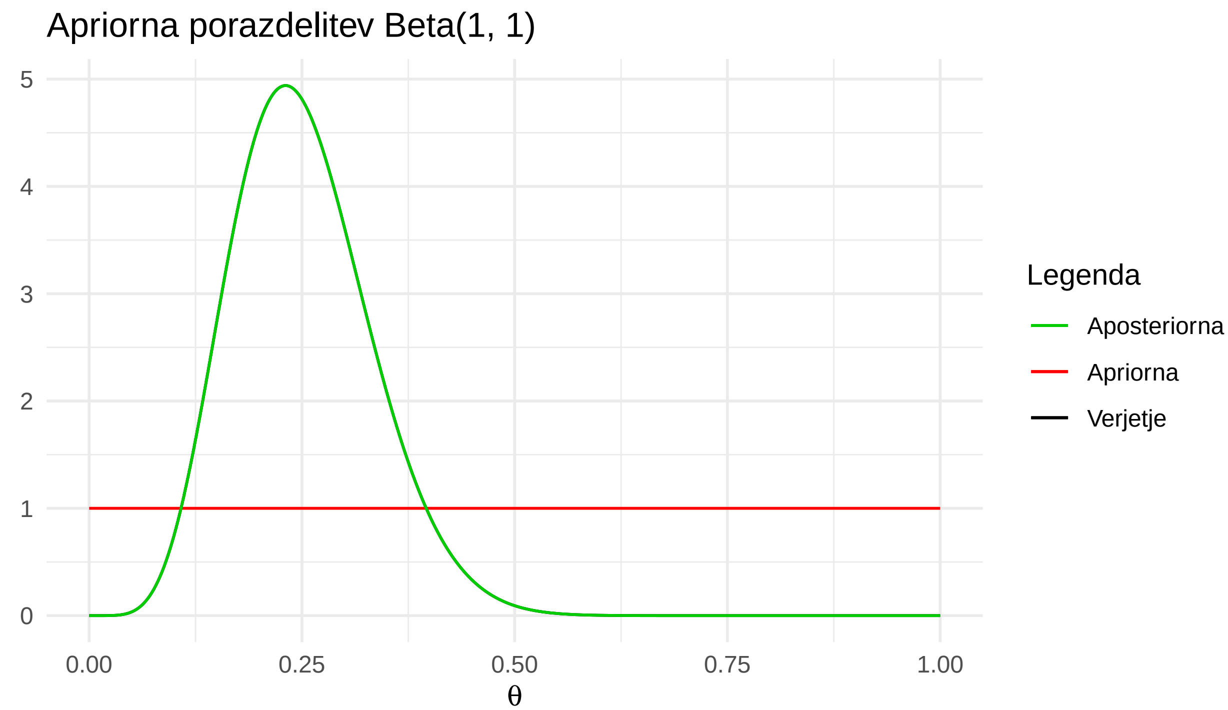
Vse skupaj zavijemo v funkcijo `plot_3()`, ki za podan `alpha` in `beta` izračuna funkcijo verjetja, apriorno ter aposteriorno porazdelitev, in jih izriše na eno sliko.

```
plot_3 <- function(alpha, beta) {

    # calculate alpha and beta posterior
    alpha.apost <- k + alpha
    beta.apost <- n - k + beta
```

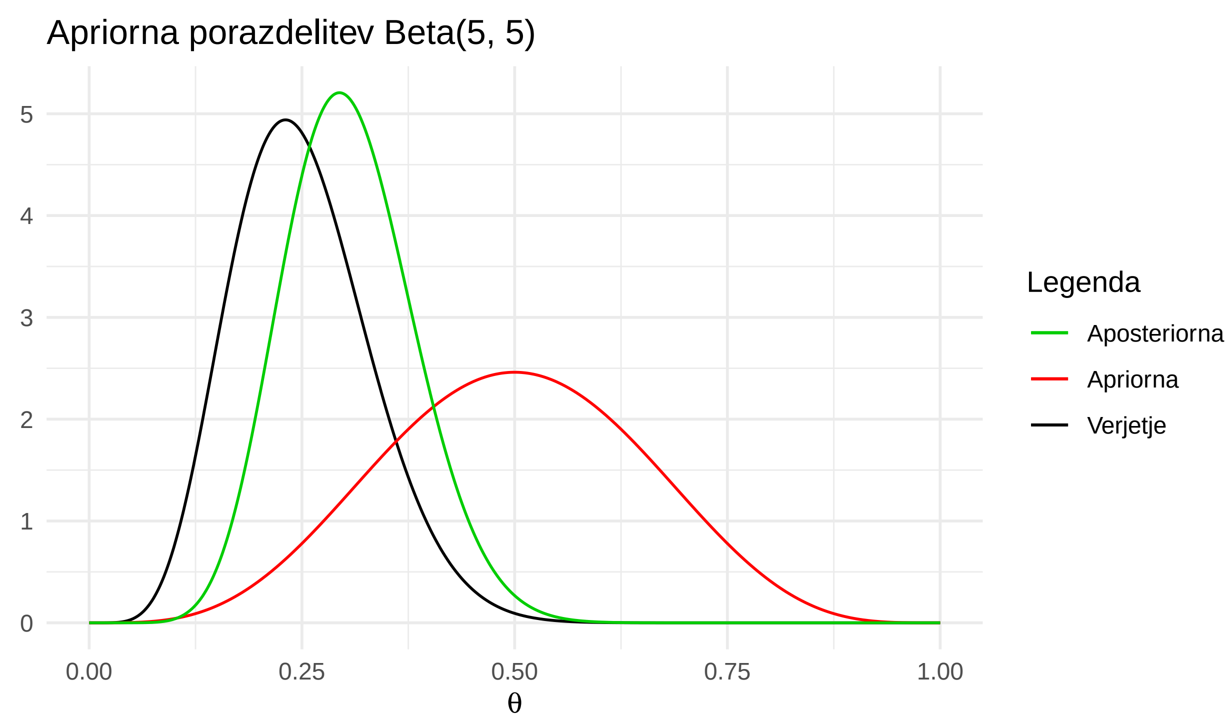
```r
    # generate theta values and calculate posterier, prior and likelihood
    theta <- seq(0, 1, 0.001)
    aposteriorna <- dbeta(theta, alpha.apost, beta.apost)
    konst.verjetje <- konst(k, n) * verjetje(theta, k, n)
    apriorna <- dbeta(theta, alpha, beta)

    # just getting max value from likelihood and posterior for proper plotting
    # as prior could have infinite values (when theta is 0 or 1)
    y.max <- max(c(konst.verjetje, aposteriorna))
    # defining colors
    colors <- c("Verjetje" = "black",
                "Apriorna" = "red",
                "Aposteriorna" = "green3")

    # plotting likelihood, prior and posterior distributions.
    ggplot(NULL, aes(x=theta)) +
        geom_line(aes(y=konst.verjetje, col="Verjetje")) +
        geom_line(aes(y=apriorna, col="Apriorna")) +
        geom_line(aes(y=aposteriorna, col="Aposteriorna")) +
        scale_color_manual(values = colors) +
        labs(x=expression(theta),
             color="Legenda",
             title=paste0("Apriorna porazdelitev Beta(", alpha, ", ", beta, ")")) +
        theme_minimal() +
        theme(axis.title.y=element_blank()) +
        ylim(c(0, y.max))
}
```

Izris

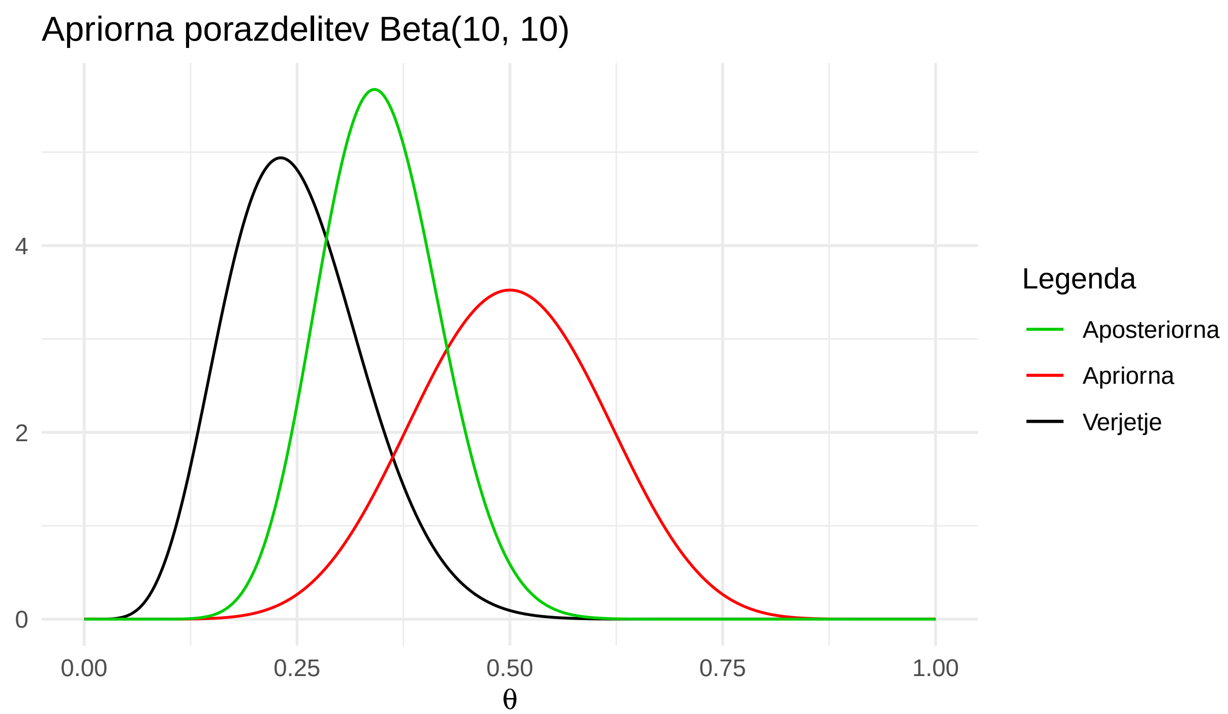
```r
plot_3(1, 1)
```



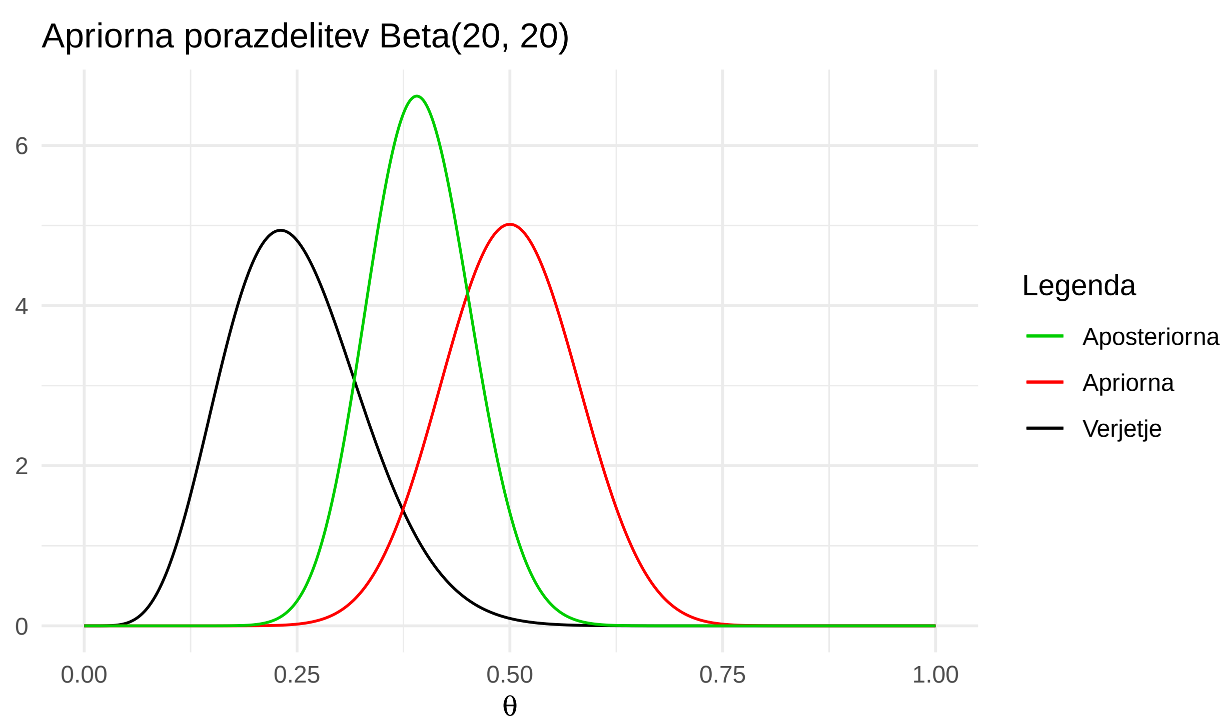

```r
plot_3(5, 5)
```



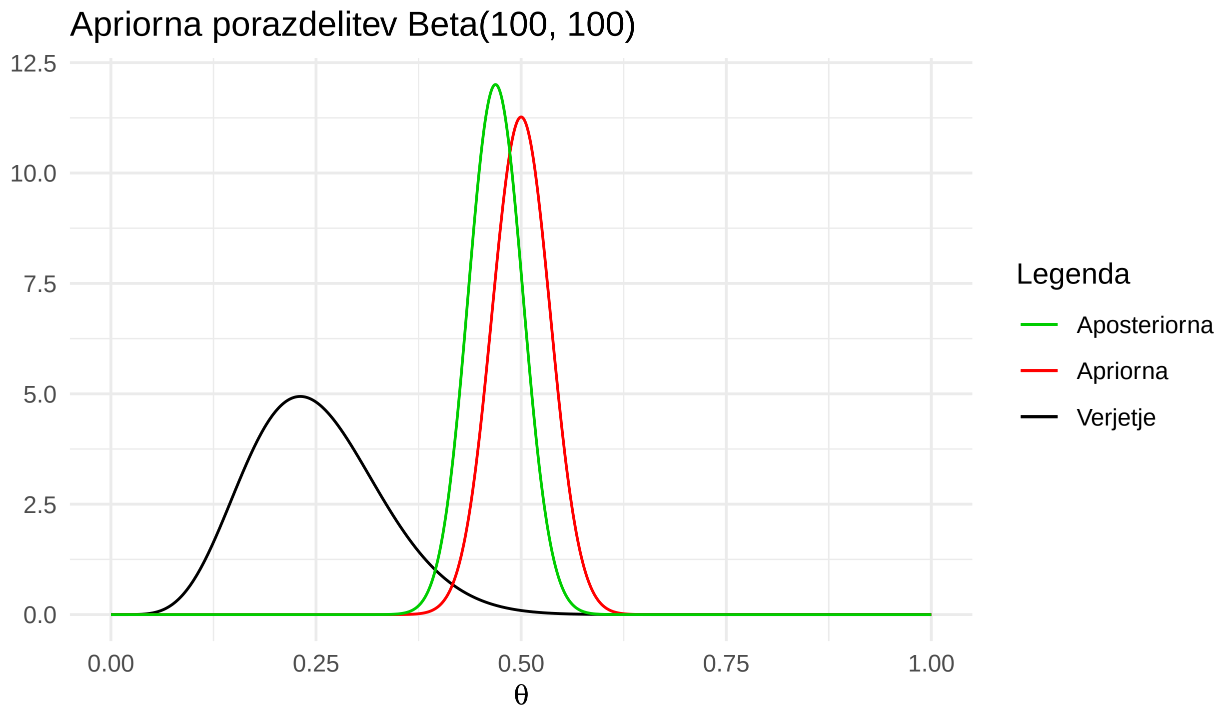
```r
plot_3(10, 10)
```



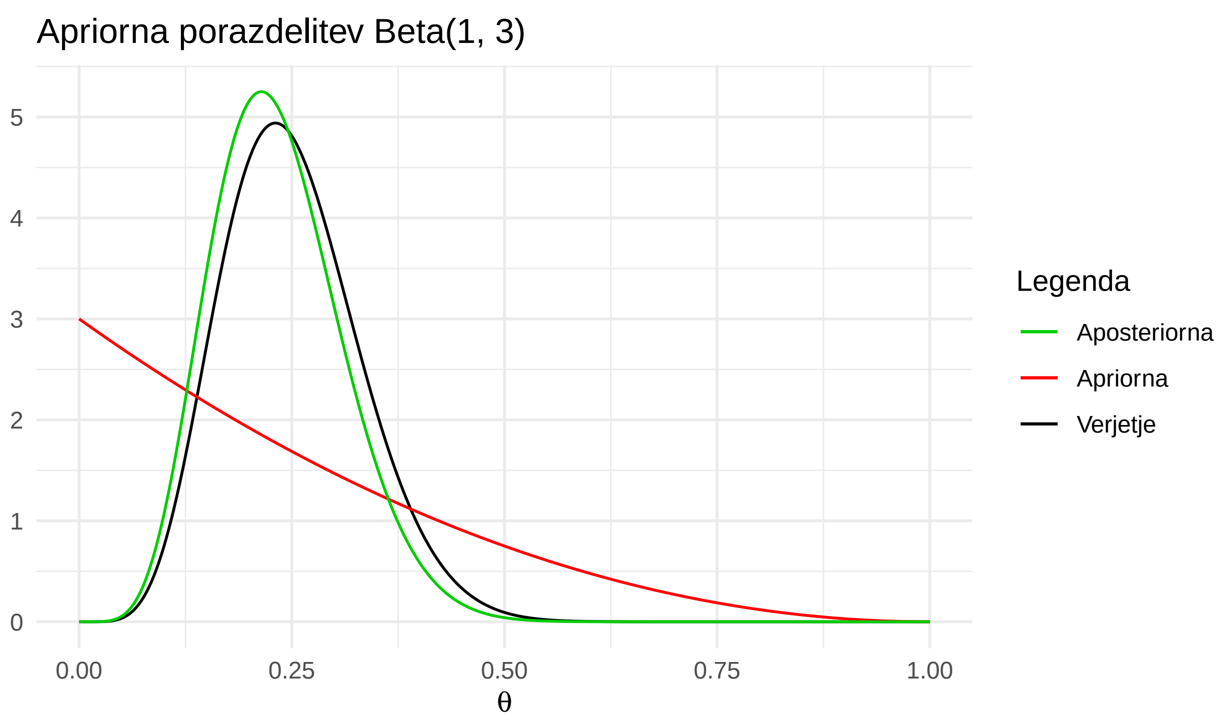

```r
plot_3(20, 20)
```



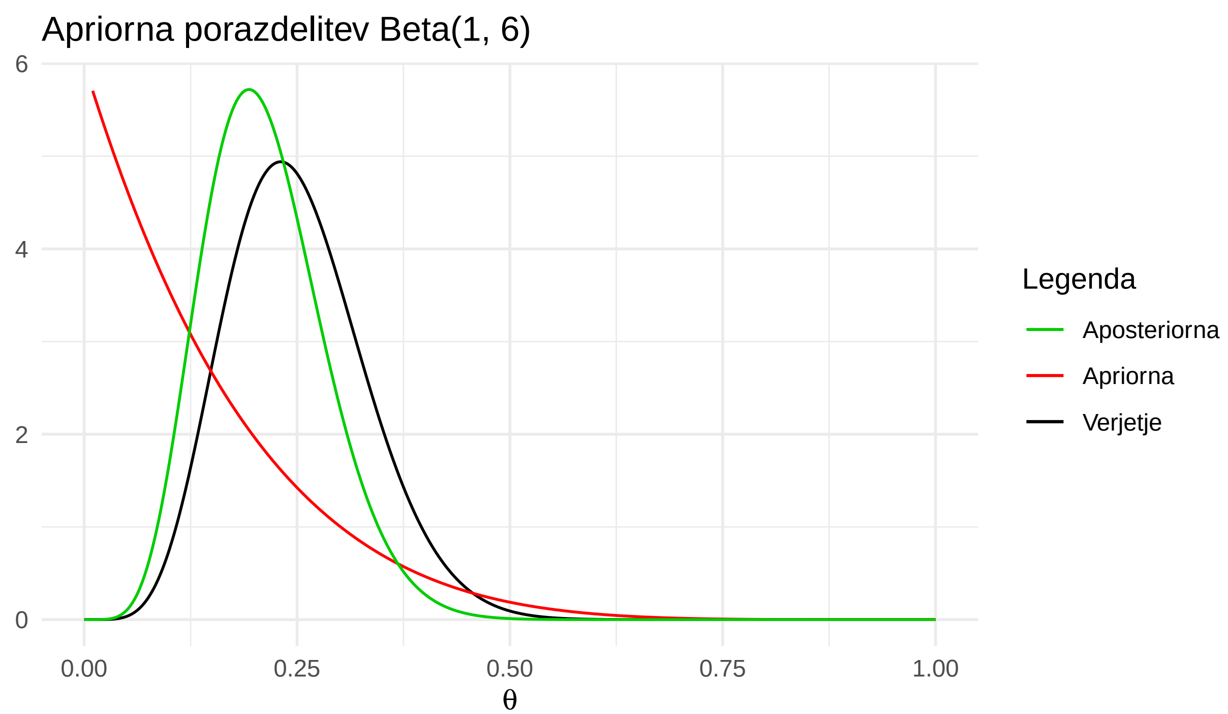
```r
plot_3(100, 100)
```



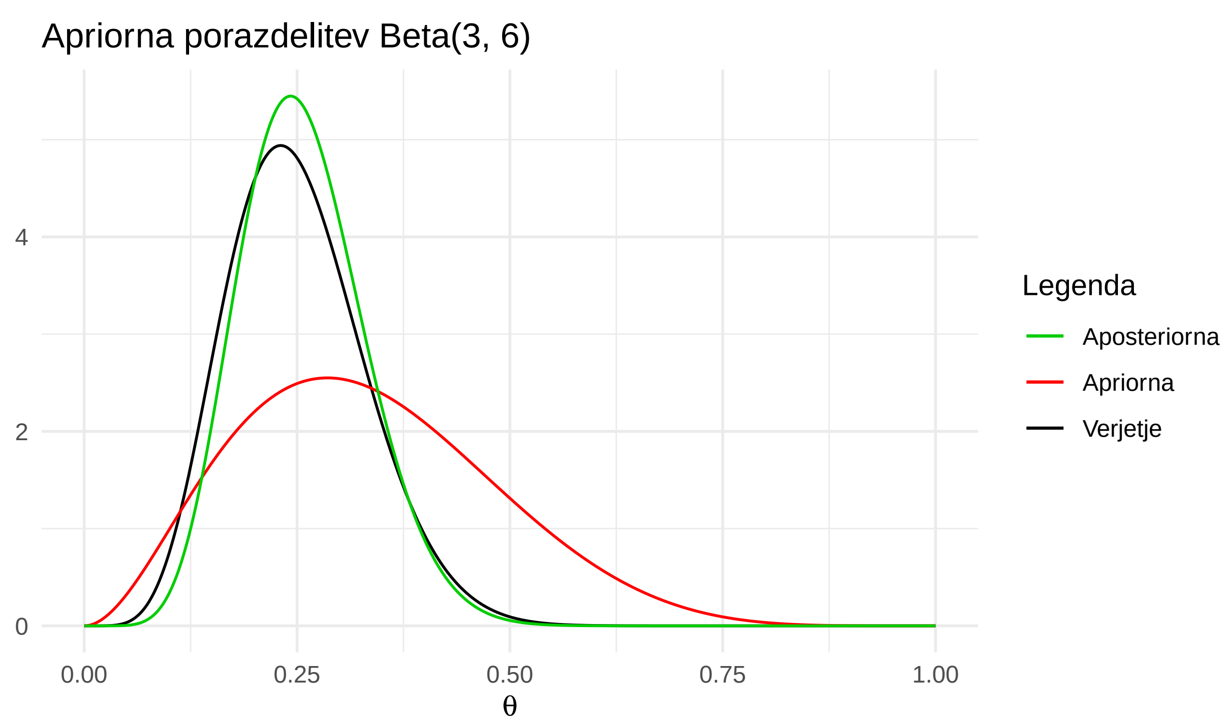

```r
plot_3(1, 3)
```



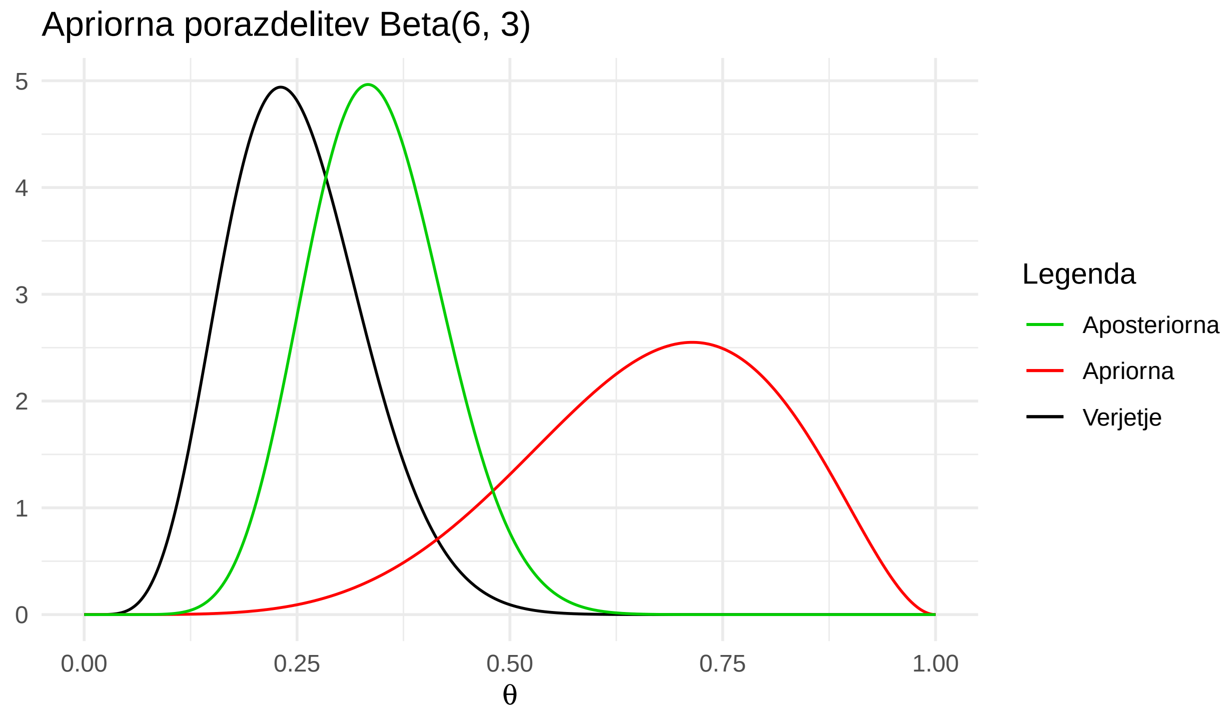
```r
plot_3(1, 6)
```



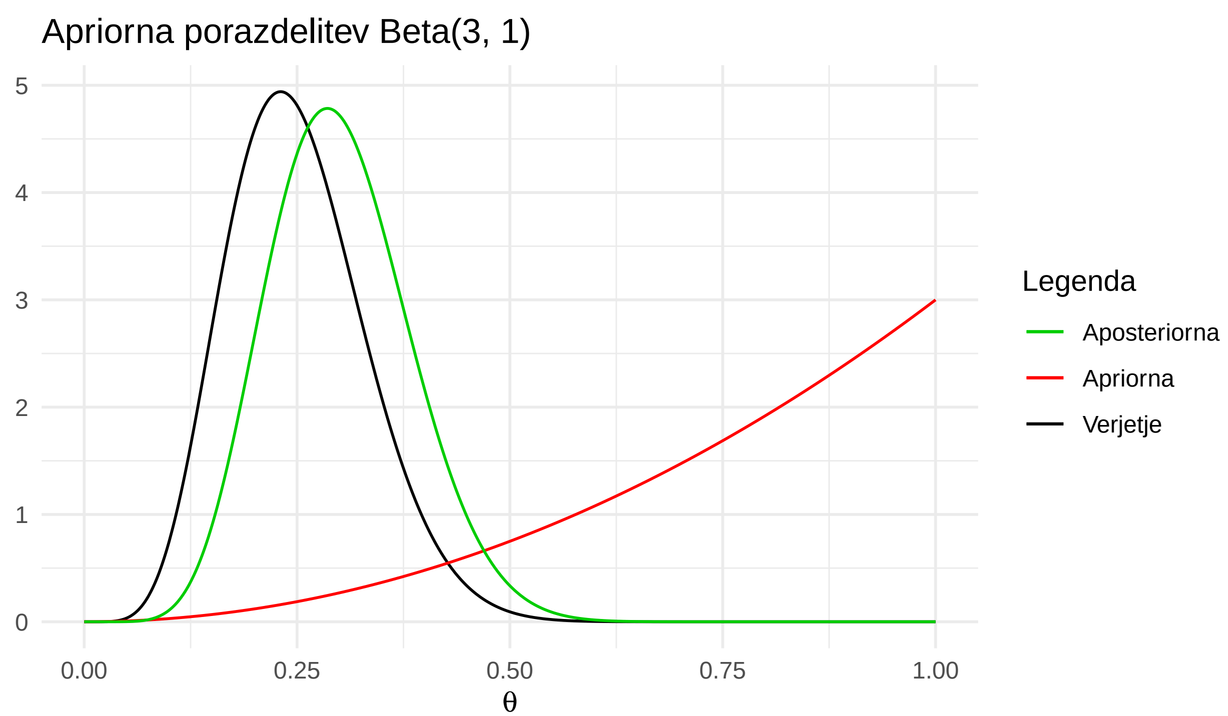

```r
plot_3(3, 6)
```



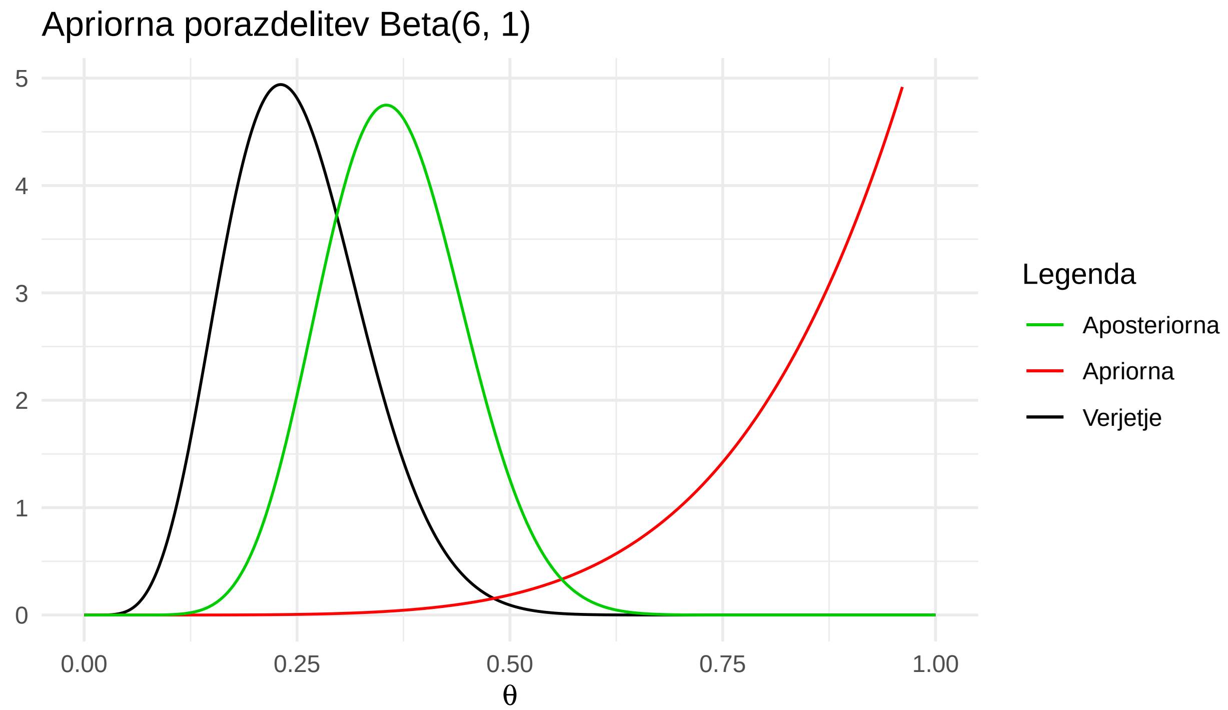
```r
plot_3(6, 3)
```



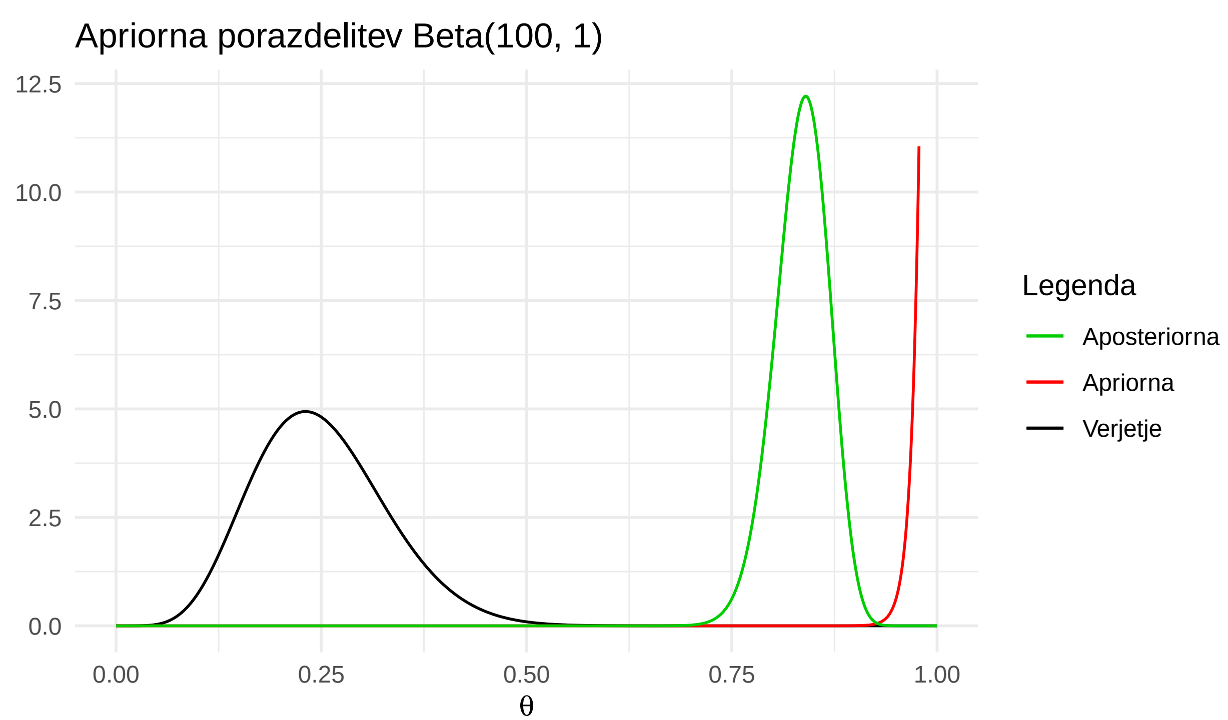

```r
plot_3(3, 1)
```



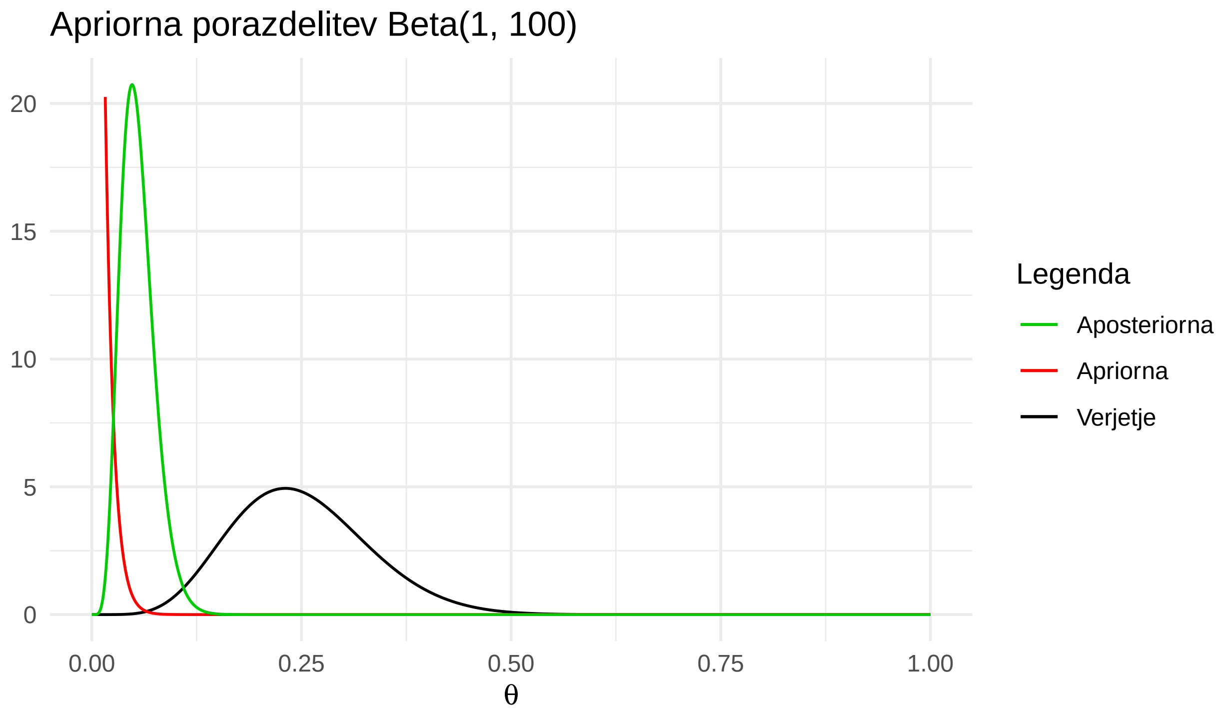
```
plot_3(6, 1)
```



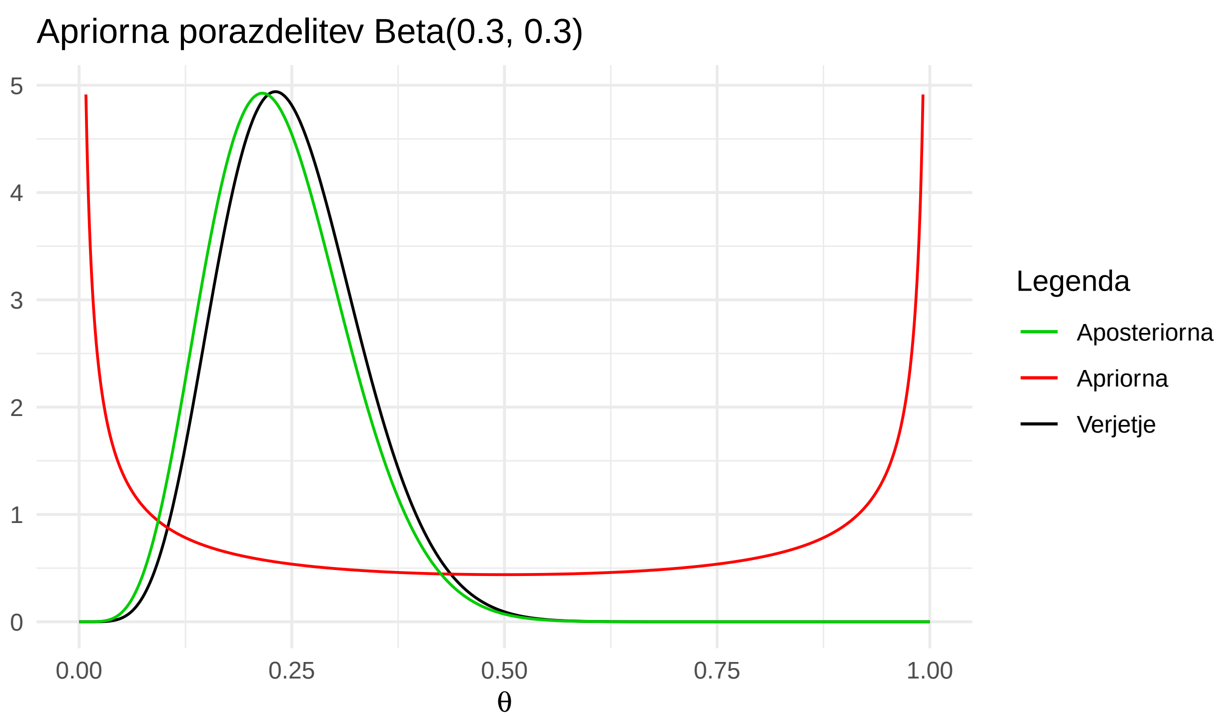

```
plot_3(100, 1)
```



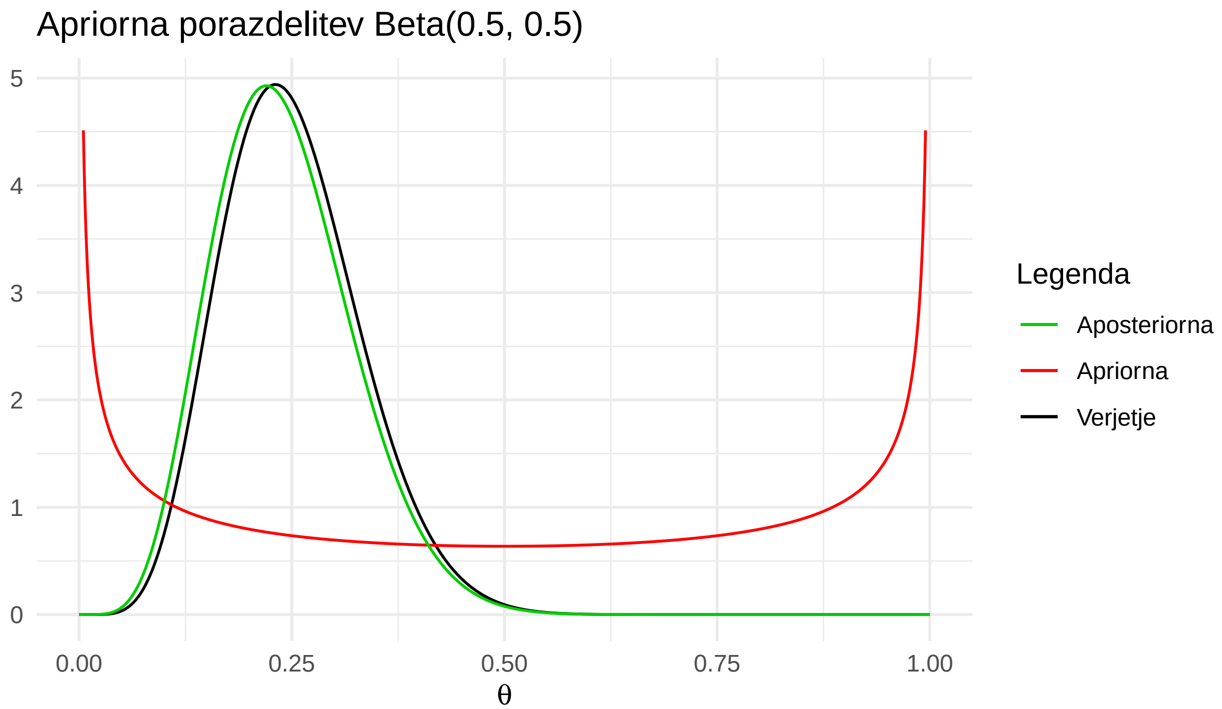
```
plot_3(1, 100)
```



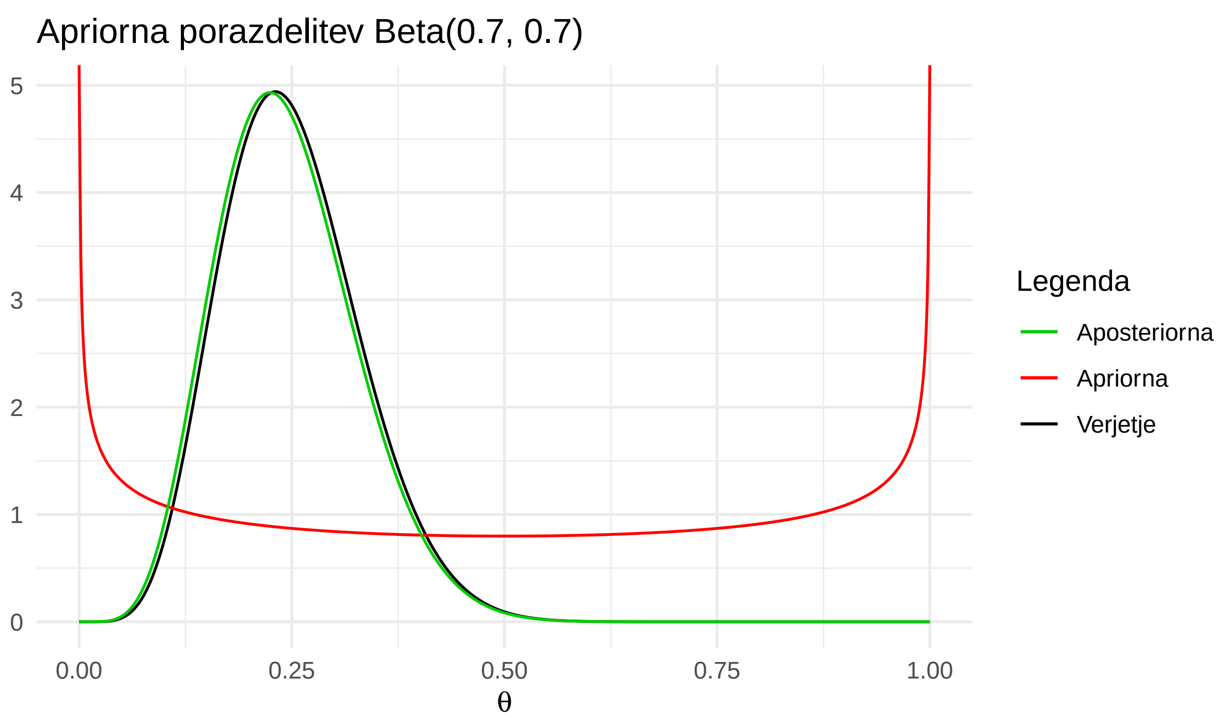

```
plot_3(0.3, 0.3)
```



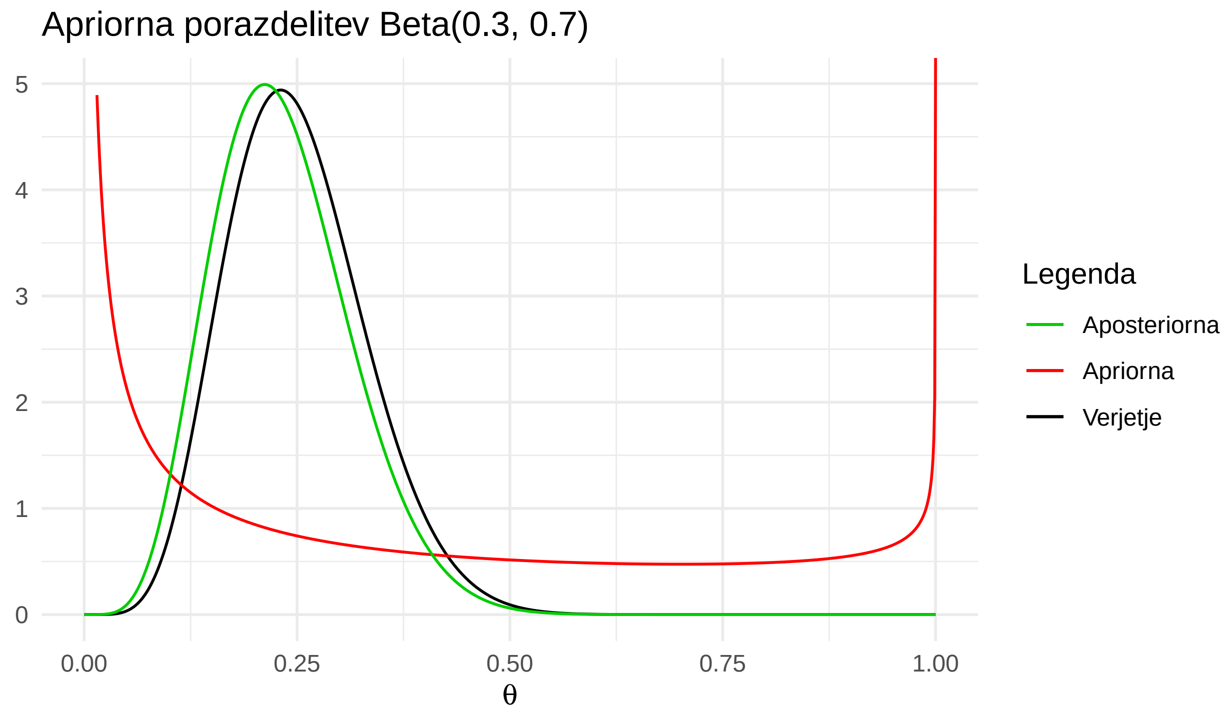
```r
plot_3(0.5, 0.5)
```



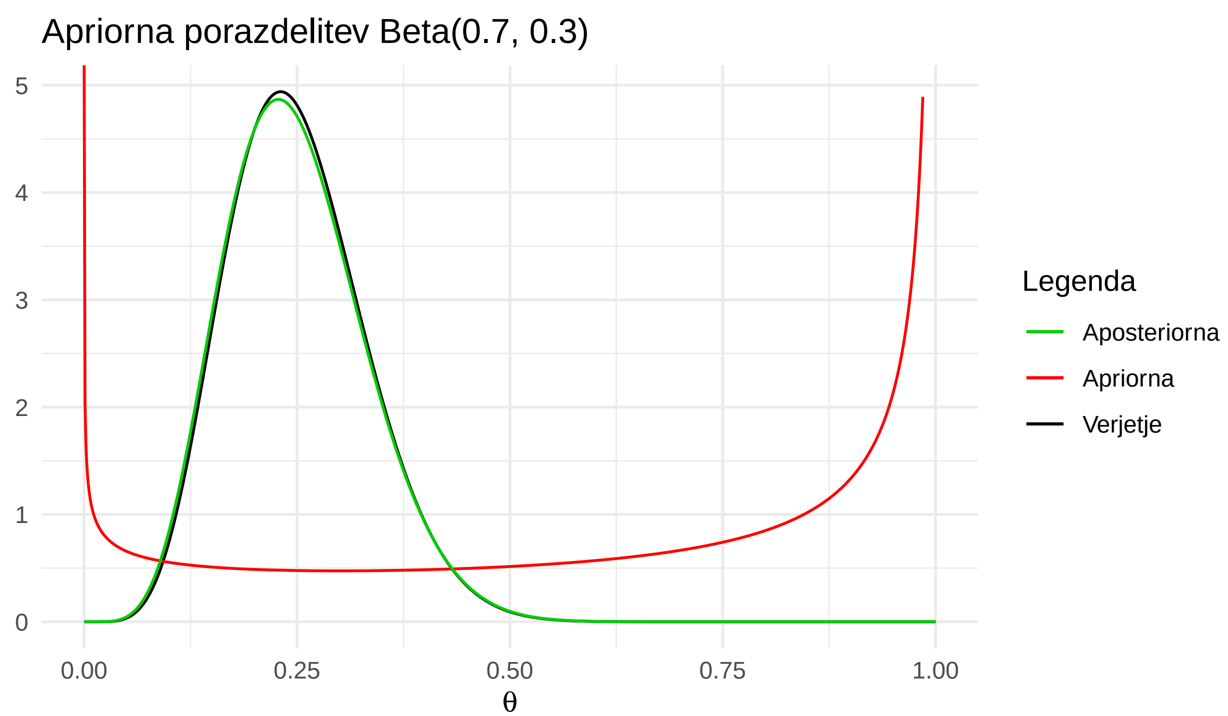

```r
plot_3(0.7, 0.7)
```



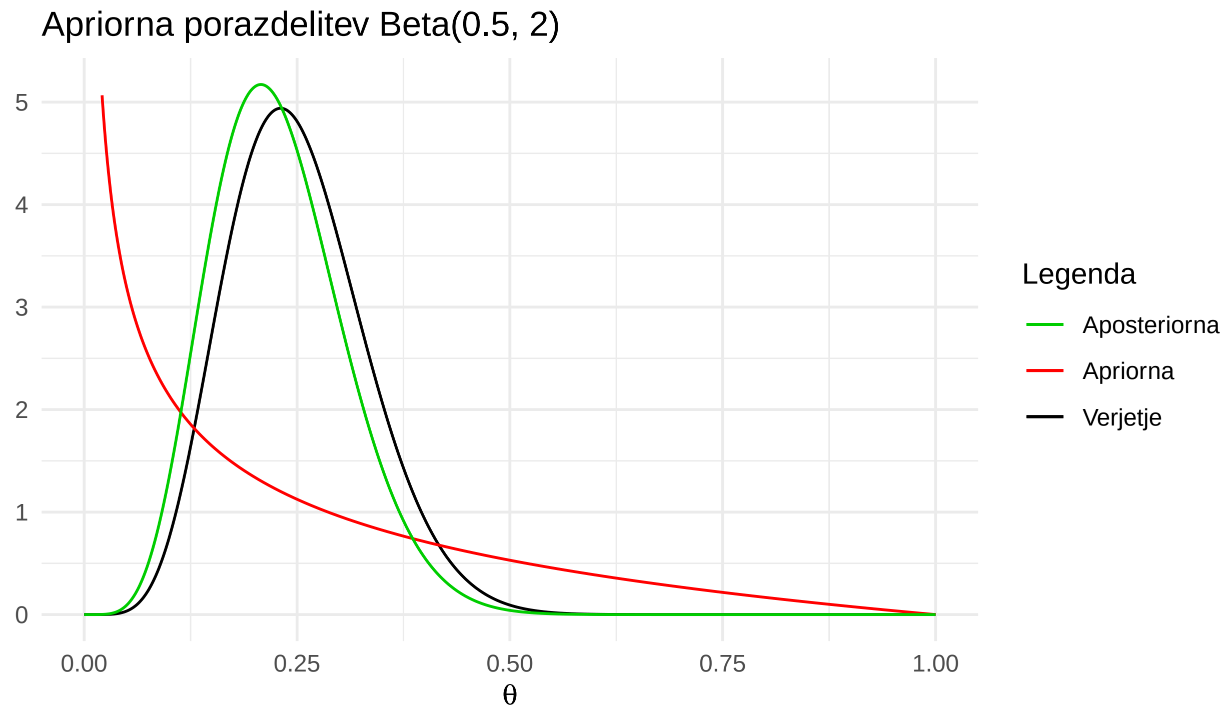
```r
plot_3(0.3, 0.7)
```



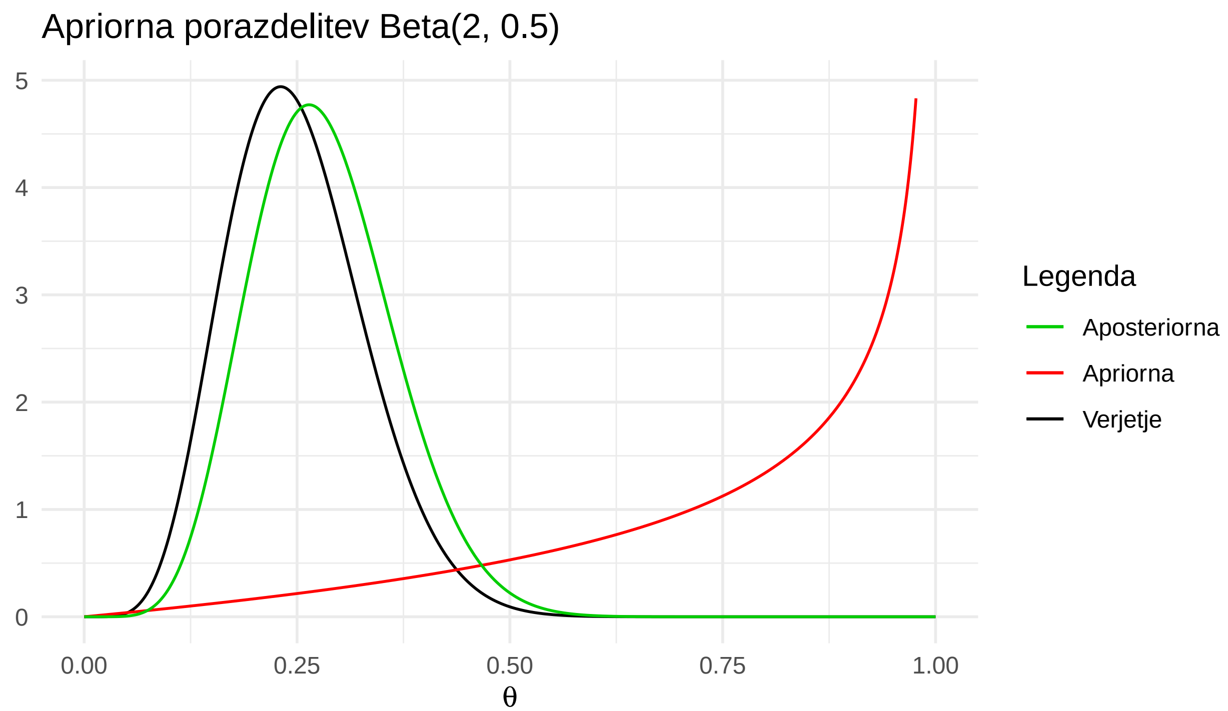

```r
plot_3(0.7, 0.3)
```



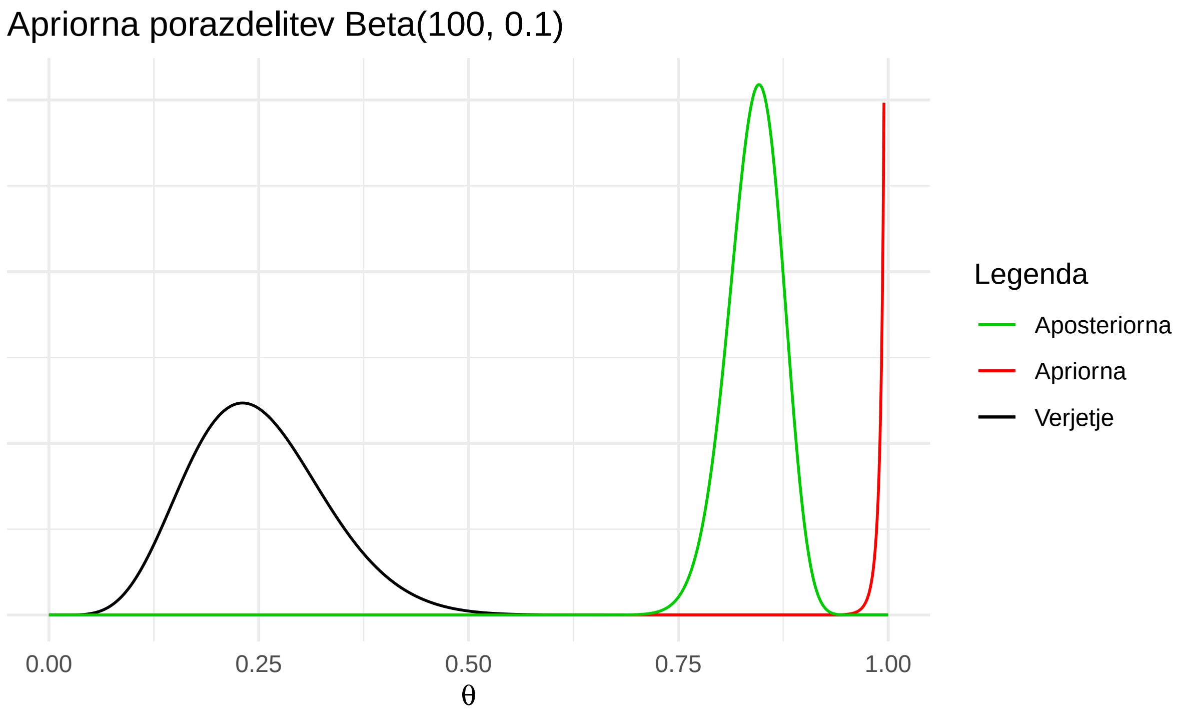
```r
plot_3(0.5, 2)
```



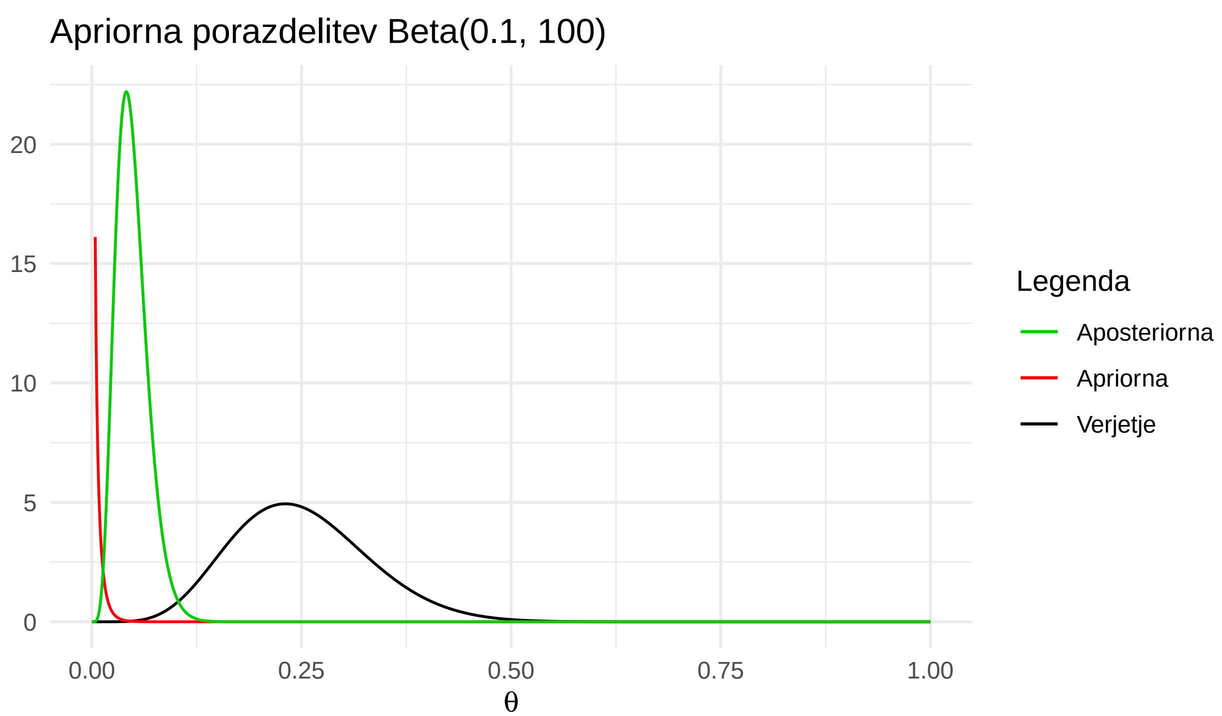

```r
plot_3(2, 0.5)
```



```r
plot_3(100, 0.1)
```




```r
plot_3(0.1, 100)
```



Ugotovitve

Funkcija verjetja so informacije, ki smo jih pridobili z eksperimentom (26 poskusov, 6 uspehov). Apriorna porazdelitev določa koliko vemo o parametru θ predhodno, mogoče iz drugih eksperimentov. Več kot vemo o predhodni porazdelitvi, bolj je aposteriorna porazdelitev blizu apriorni. To je videti iz zelo velikih α in β parametrov (primeri ko je vsaj eden izmed parametrov enak 100)

z α in β smo določali obliko apriorne porazdelitve. Večji kot je α bolj se nagibamo proti 1, večji kot je β bolj se nagibamo proti 0. Seveda mora veljati da sta tako α kot β večja od 0.

V primeru neinformativne porazdelitve (nimamo nobenih predhodnih informacij o parametru θ; $\alpha = \beta = 1$), praktično vse naše znanje o parametru θ izhaja iz funkcije verjetja.

Naloga 2

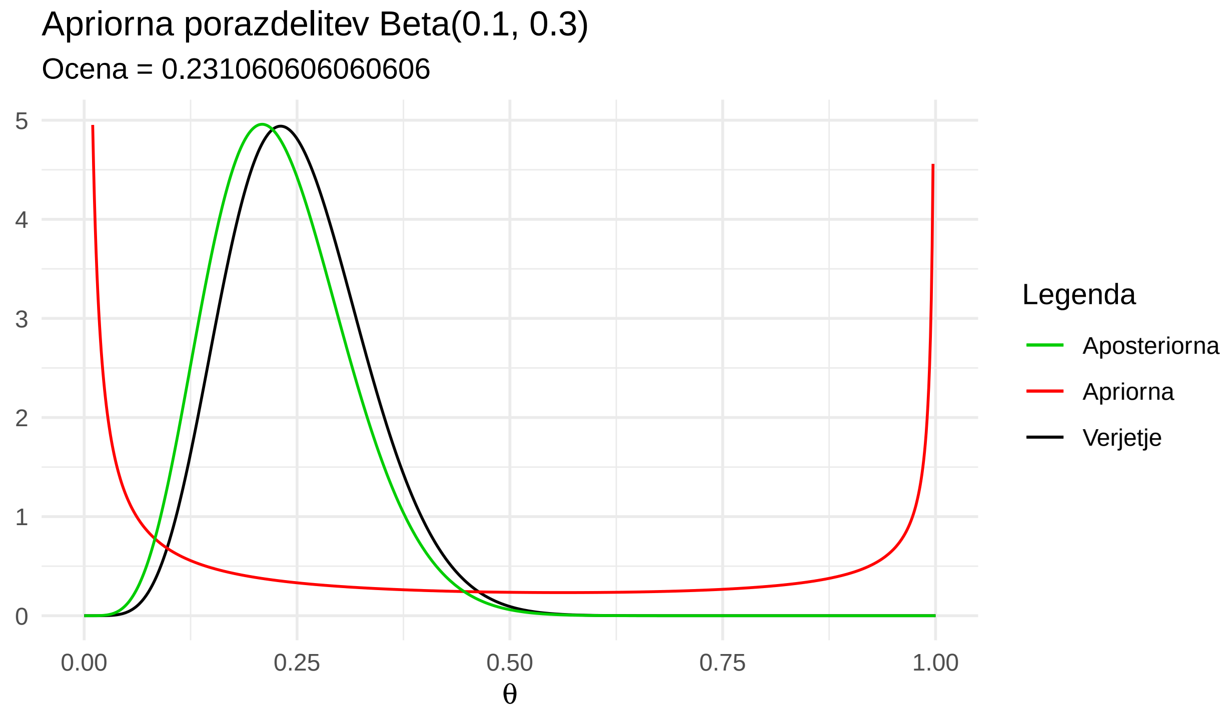
Izbrati moramo takšen α in β da bo pričakovana vrednost θ enak 0.25. Vemo da velja $E[B(\alpha,\beta)] = \frac{\alpha}{\alpha+\beta}$. Glede na to lahko izračunamo naslednje

$$\frac{\alpha}{\alpha+\beta} = \frac{1}{4} 4\alpha = \alpha + \beta 3\alpha = \beta\,.$$

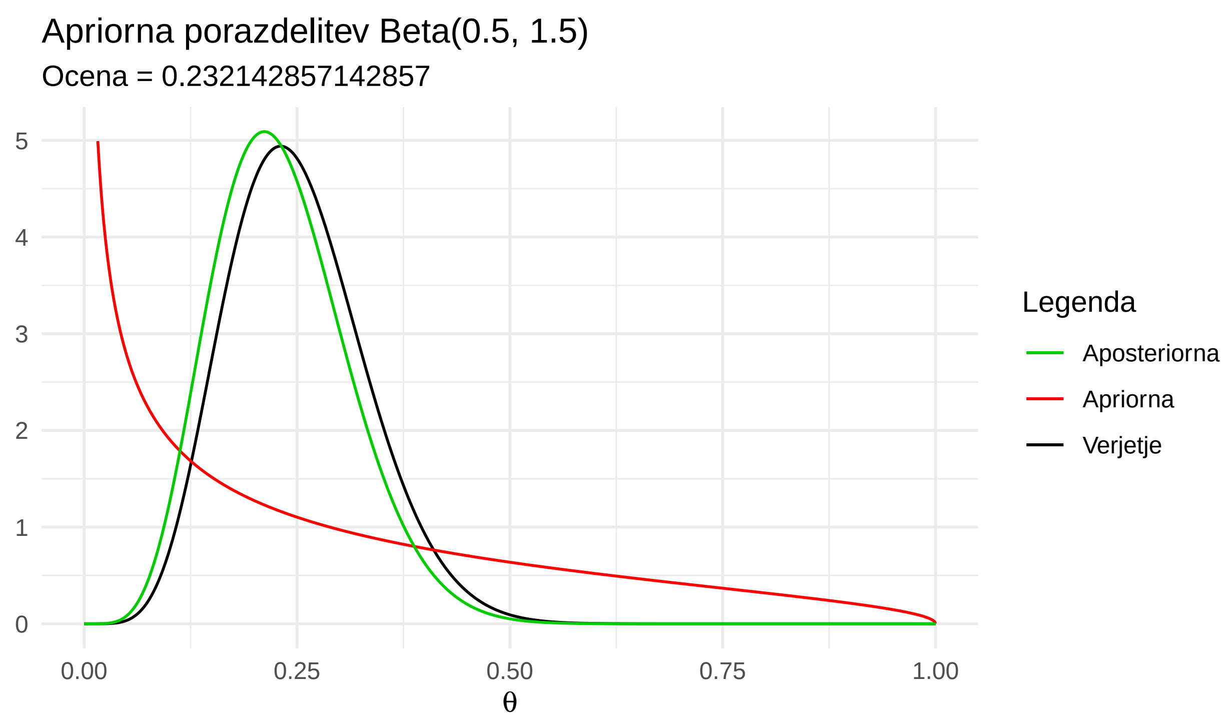
Se pravi lahko izberemo takšen par α in β, da velja $3\alpha = \beta$.

```
plot_3_estimate <- function(alpha, beta) {

    # calculate alpha and beta posterior
    alpha.apost <- k + alpha
    beta.apost <- n - k + beta

    # generate theta values and calculate posterier, prior and likelihood
    theta <- seq(0, 1, 0.001)
    aposteriorna <- dbeta(theta, alpha.apost, beta.apost)
    konst.verjetje <- konst(k, n) * verjetje(theta, k, n)
    apriorna <- dbeta(theta, alpha, beta)

    # just getting max value from likelihood and posterior for proper plotting
    # as prior could have infinite values (when theta is 0 or 1)
    y.max <- max(c(konst.verjetje, aposteriorna))
    # defining colors
    colors <- c("Verjetje" = "black",
                "Apriorna" = "red",
                "Aposteriorna" = "green3")

    # plotting likelihood, prior and posterior distributions.
    ggplot(NULL, aes(x=theta)) +
        geom_line(aes(y=konst.verjetje, col="Verjetje")) +
        geom_line(aes(y=apriorna, col="Apriorna")) +
        geom_line(aes(y=aposteriorna, col="Aposteriorna")) +
        scale_color_manual(values = colors) +
        labs(x=expression(theta),
             color="Legenda",
             title=paste0("Apriorna porazdelitev Beta(", alpha, ", ", beta, ")"),
             subtitle=paste0("Ocena = ", alpha.apost/(alpha.apost + beta.apost))) +
        theme_minimal() +
        theme(axis.title.y=element_blank()) +
```

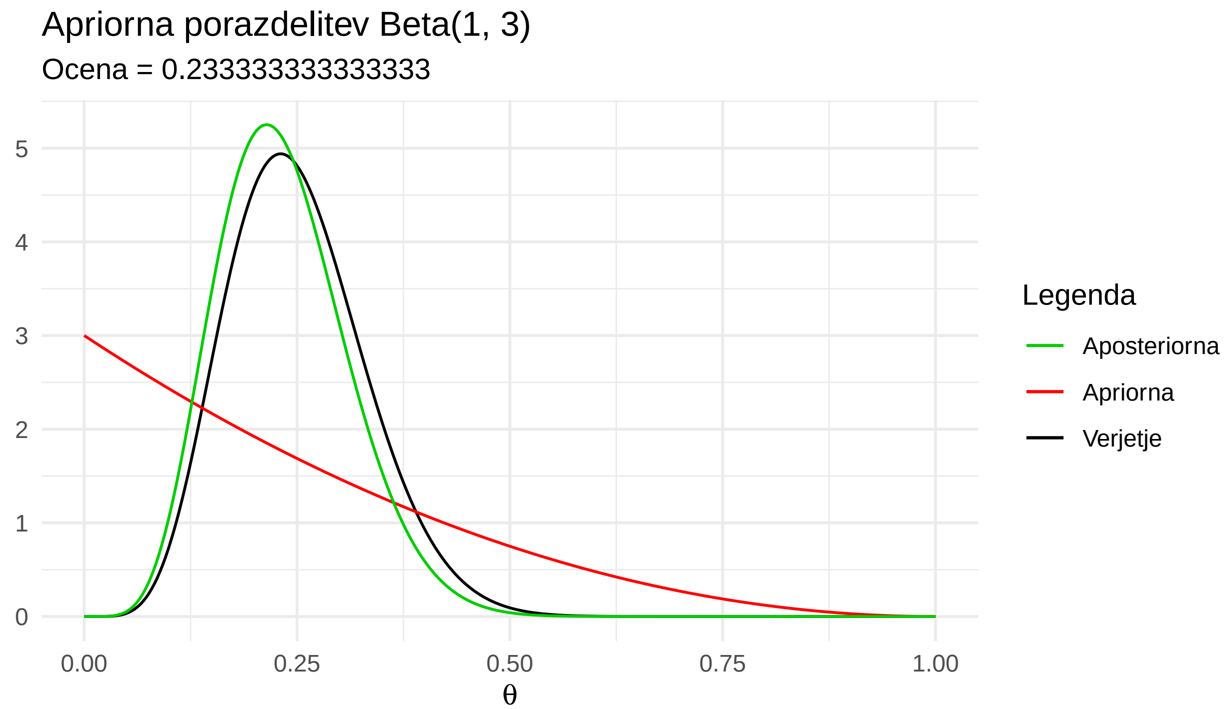
```
        ylim(c(0, y.max))
}

plot_3_estimate(0.1, 0.3)
```



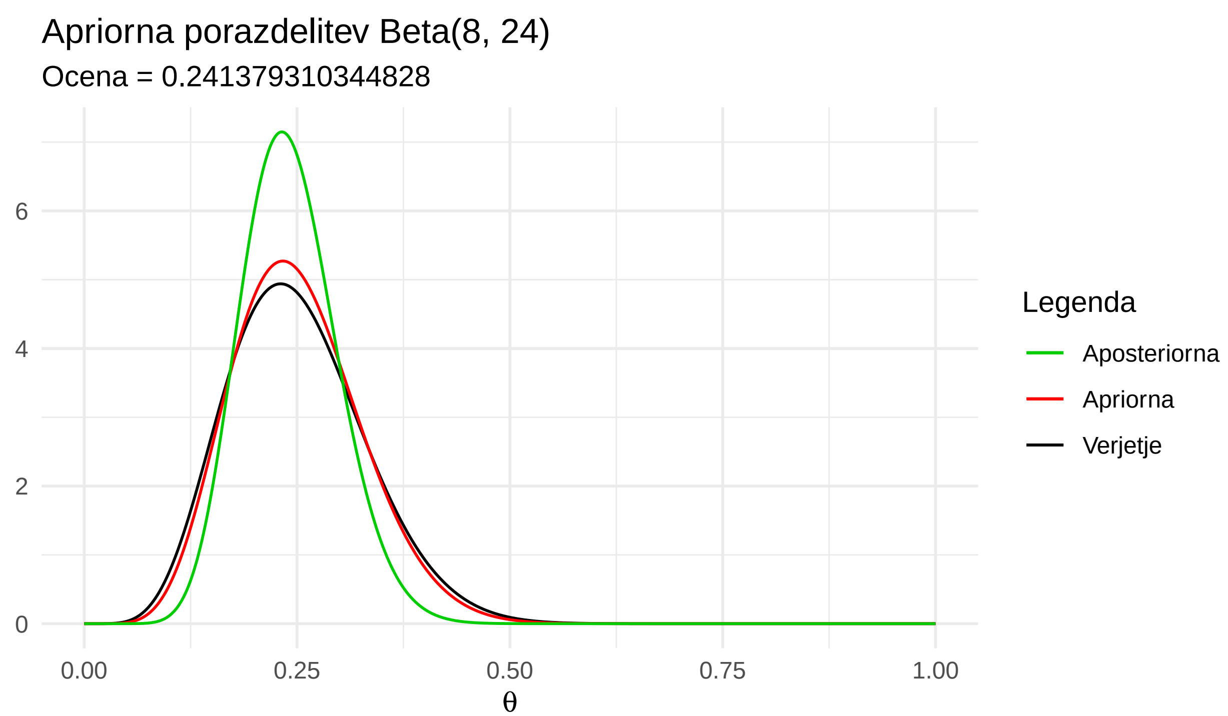

```
plot_3_estimate(0.5, 1.5)
```



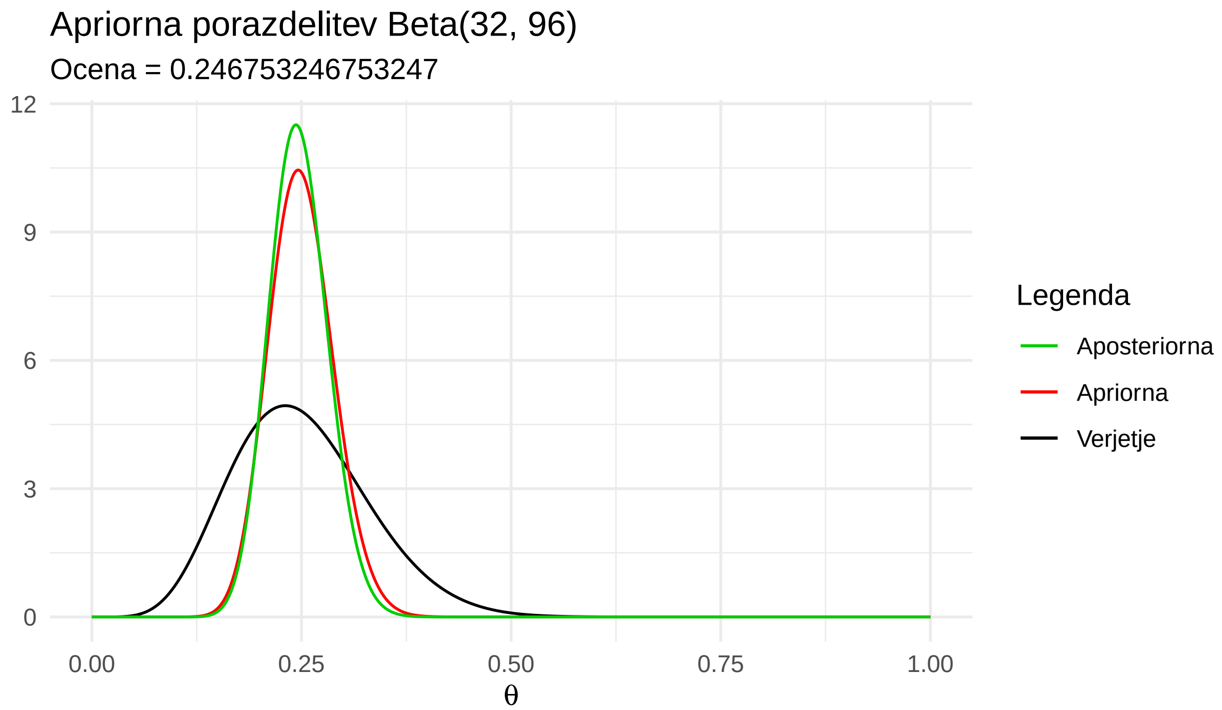
```r
plot_3_estimate(1, 3)
```



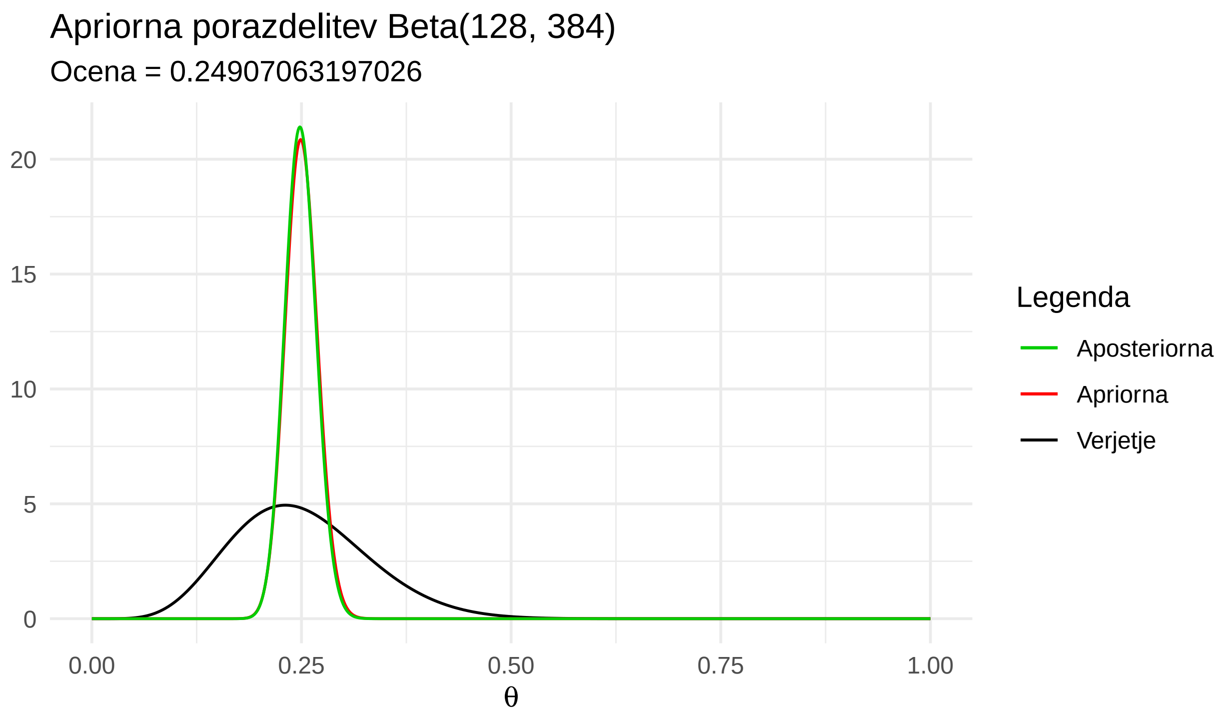

```r
plot_3_estimate(8, 24)
```



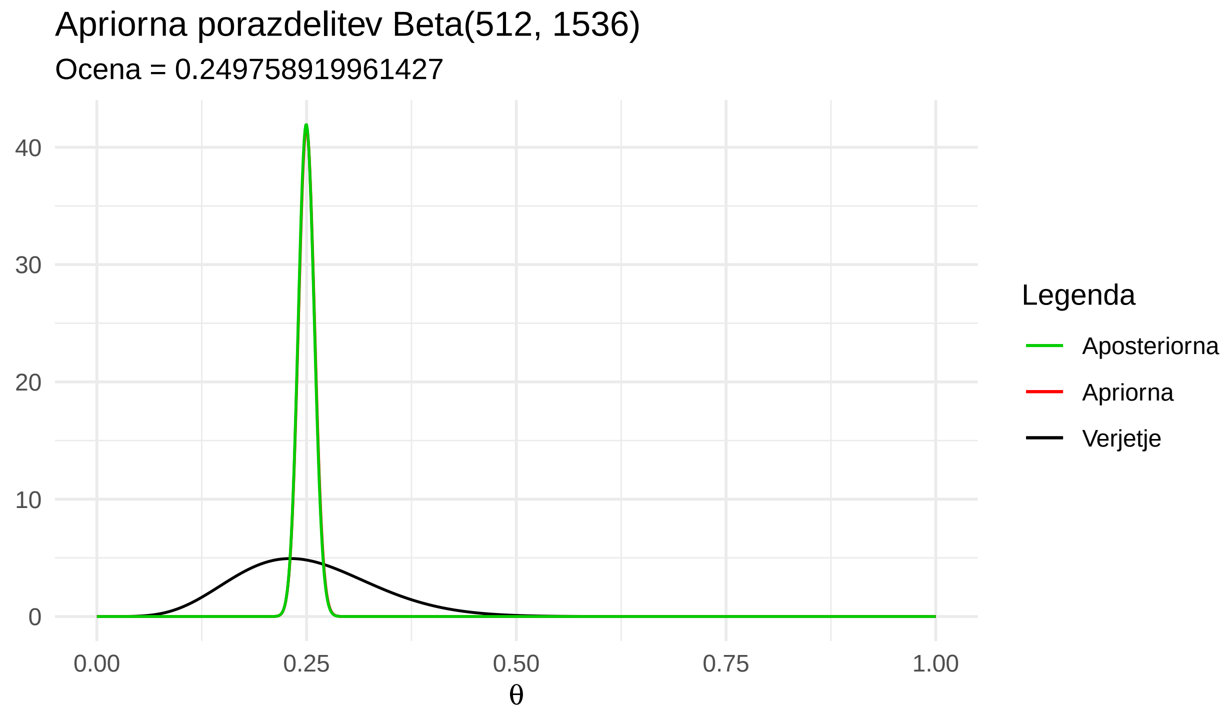
```r
plot_3_estimate(32, 96)
```




```r
plot_3_estimate(128, 384)
```



```r
plot_3_estimate(512, 1536)
```



Podobno kot pri nalogi 1, vidimo kako vpliva bodisi funkcija verjetja bodisi apriorna porazdelitev. Pri majhnih α in β, smo manj prepričani v porazdelitev apriorne porazdelitve, vendar kljub precej neinformativni apriorni porazdelitvi smo še vedno blizu pričakovani vrednosti 0.25, ker je tudi vrh funkcije verjetja zelo blizu naši pričakovani vrednosti. S tem ko večamo tako α kot β pravzaprav manjšamo varianco, in posledično je naša pričakovana vrednost aposteriorne porazdelitve, vedno bližje naši pričakovani vrednosti. Podobno je tudi aposteriorna porazdelitev vedno bližje apriorni porazdelitvi.