

Beta-Binomski Model

Matevž Gombač

2023-11-11

Uvod

V tej domači nalogi si bom ogledal določene lastnosti Binomskega modela, ki smo ga obravnavali tako na vajah kot na predavanjih. Preizkusil bom različne apriorne beta porazdelitve in vizualiziral svoje ugotovitve. Pri tem bom uporabljal podatke iz vaj.

Beta-Binomski model

Primer modela je ugotavljanje pravilnega odgovora izmed štirih možnosti. Naš vzorec je X_1, X_2, \dots, X_n , kjer je n število odgovorov, X_i pa pravilnost i -tega odgovora ($X_i = 1$, če odgovor pravilen, sicer 0). Zanimala nas bo verjetnost θ , da na vprašanje odgovorimo pravilno. Pri tem si bomo pomagali s skupnim številom pravih odgovorov, torej $X = X_1 + X_2 + \dots + X_n$.

Glede na θ je seveda $(X | \theta) \sim \text{Bin}(n, \theta)$. To nam ponudi idejo iz predavanj, da za apriorno porazdelitev vzamemo beta porazdelitev: $\pi(\theta) = \pi(\theta | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$, kjer je $B(\alpha, \beta)$ običajna beta funkcija za $\alpha, \beta > 0$. Naj bo $0 \leq k \leq n$ število pravih odgovorov oziroma $X = k$. Potem je po izpeljavi iz predavanj tudi aposteriorna porazdelitev pravilnosti odgovora v družini beta, in sicer s parametroma $\alpha_{\text{apost}} = \alpha + k$ ter $\beta_{\text{apost}} = n - k + \beta$.

1. Naloga

Za naš primer si vzemimo konkretne podatke $n = 26$ in $k = 6$.

```
n <- 26
k <- 6
```

Ogledal si bom spreminjanje apriorne beta porazdelitve glede na različne kombinacije parametrov α in β .

Če vzamemo $\alpha = \beta = 1$ dobimo za apriorno porazdelitev enakomerno porazdelitev $U(0, 1)$, ki ji rečemo tudi neinformativna apriorna porazdelitev. Oglejmo si sedaj primer, ko sta parametra strogo večja ali manjša od 1.

Spomnimo se še *verjetja*, ki je definirano kot $L(\theta | x) = L(\theta; x) := f(x | \theta)$, kjer je f začetna porazdelitev oz. gostota. V našem primeru je f binomska porazdelitev in sledi

$$L(\theta | x) = \theta^k (1 - \theta)^{n-k}$$

.

Primerjajmo apriorno, aposteriorno in verjetje:

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

#Verjetje se mora pointegrirati v 1 (rabimo konst.)
```

```

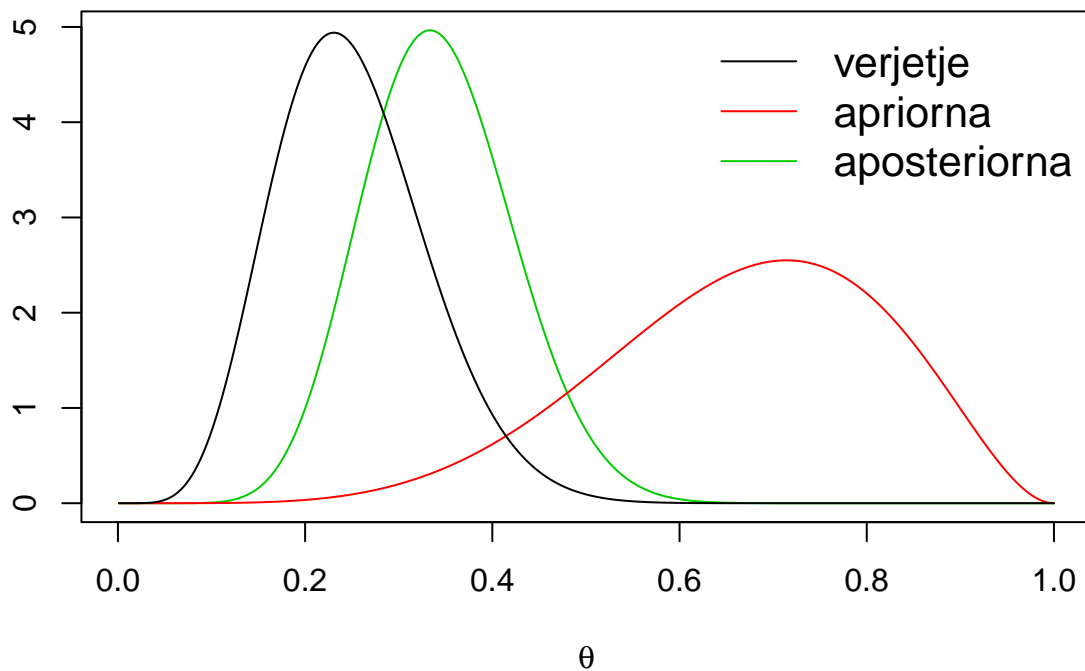
konst <- function(k,n) {
  theta <- seq(0.001, 1, 0.001)
  1 / (0.001 * sum(verjetje(theta,k,n)))
}

narisi_vse <- function(alpha, beta) {
  theta <- seq(0.001, 1, 0.001)
  alpha_apost = alpha + k
  beta_apost = beta + n - k

  apriorna <- dbeta(theta, alpha, beta)
  aposteriorna <- dbeta(theta, alpha_apost, beta_apost)
  konst_verjetje <- konst(k,n) * verjetje(theta, k, n)
  plot(theta, aposteriorna,type="l", col="green3", xlab=expression(theta), ylab="")
  lines(theta, apriorna, col = "red")
  lines(theta, konst_verjetje, col="black")
  legend("topright", legend= c("verjetje", "apriorna", "aposteriorna"), col = c("black","red", "green3"),
  }
#varianca od beta porazd
varianca_beta <- function(alpha, beta) {
  (alpha*beta) / ((alpha + beta)^2 * (alpha + beta +1 ))
}

narisi_vse(6,3)

```



Vidimo, da apriorna porazdelitev ni najbolje opisala dejanske verjetnosti pravilnega odgovora, poleg tega je pa imela tudi precej veliko varianco v primerjavi z apriorno porazdelitvijo.

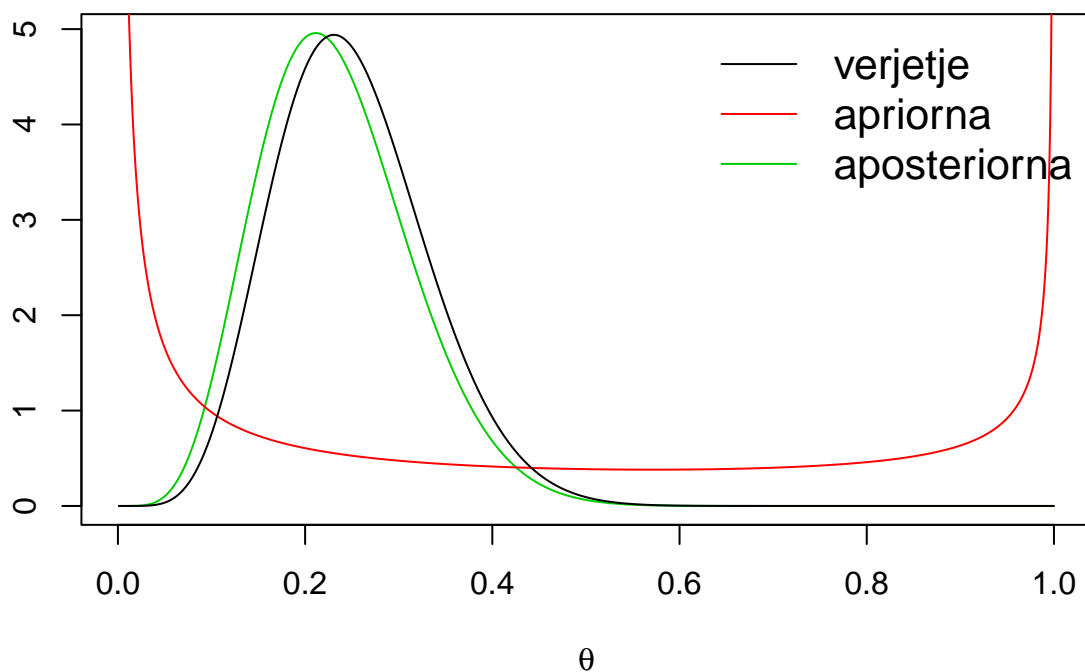
```
varianca_beta(3,6) #var apriorne
```

```
## [1] 0.02222222
```

```
varianca_beta(n + 3, n - k + 6) #var aposteriorne
```

```
## [1] 0.004451004
```

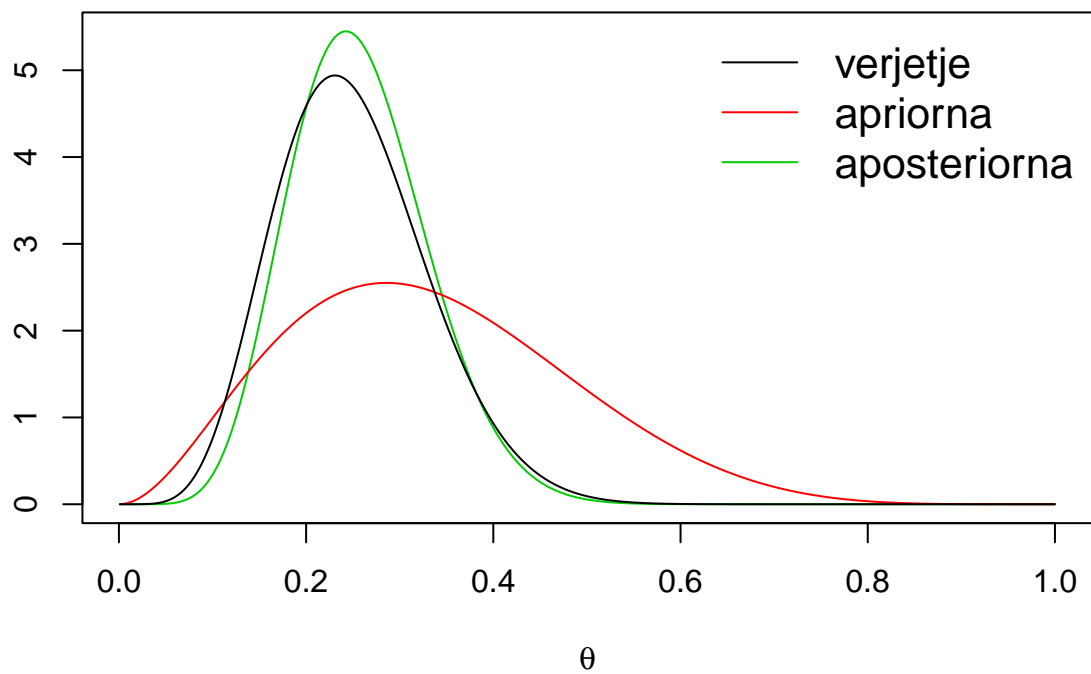
```
narisi_vse(0.2, 0.4)
```



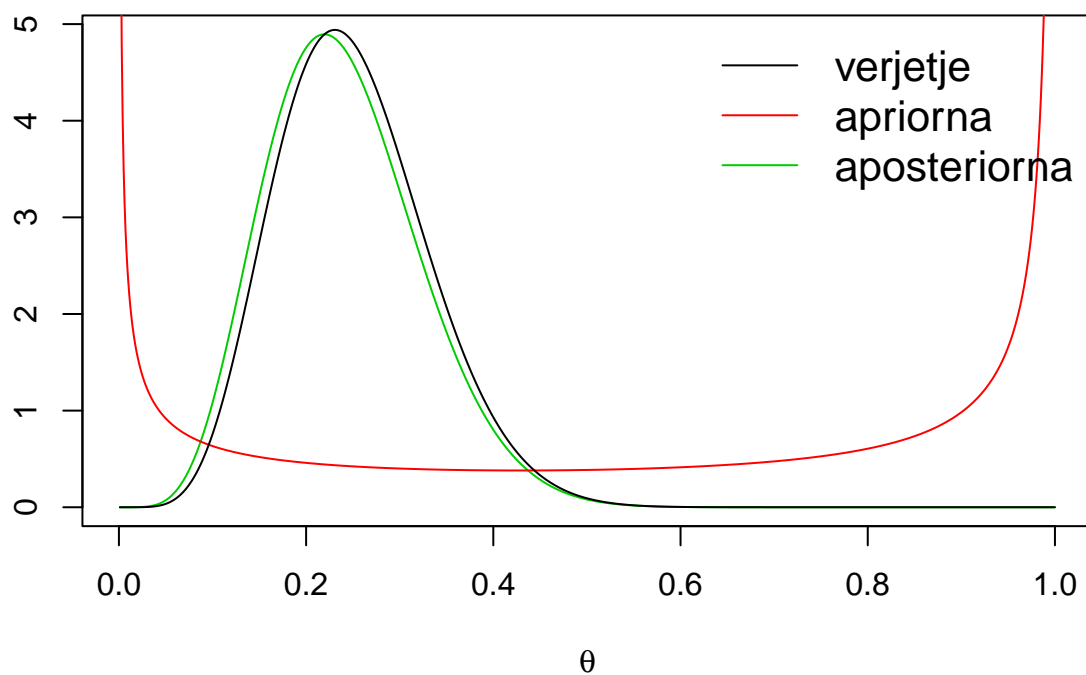
Za $\alpha, \beta < 1$ pa vidimo, da se apriorna porazdelitev skoraj ujema z Jeffreyovo, kar pa ni presenečenje, saj je ta določena z $B(0.5, 0.5)$ (za binomski model). Hkrati pa se aposteriorna veliko bolj prilega verjetju.

Poglejmo še, kaj se zgodi, ko parametra zamenjamo in kaj se zgodi, če ju povečamo.

```
narisi_vse(3,6)
```

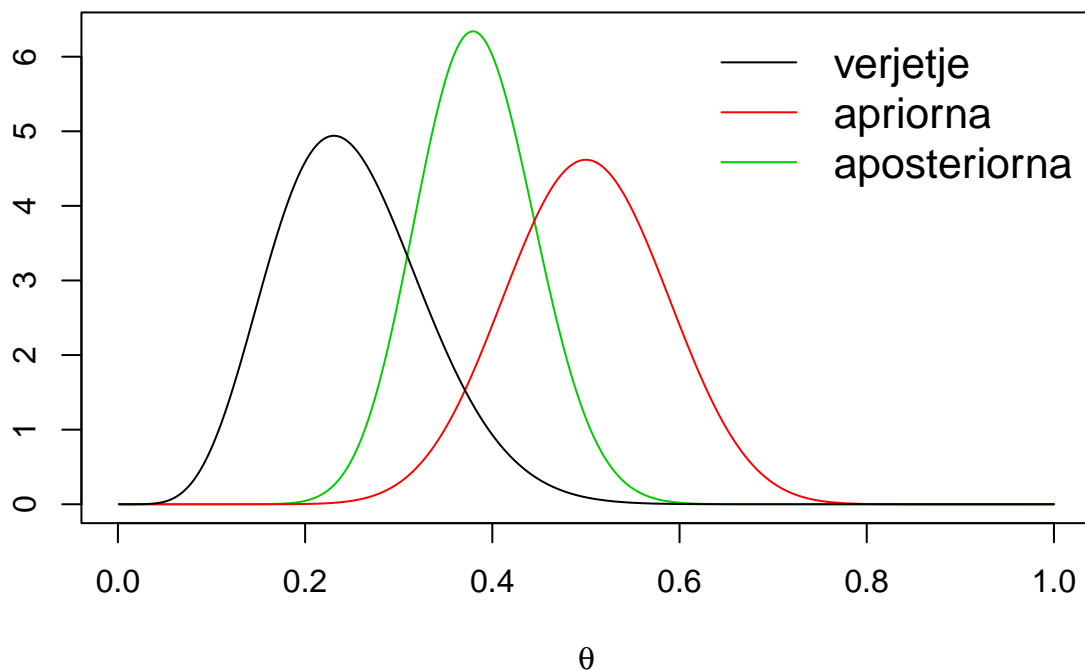


```
narisi_vse(0.4, 0.2)
```



Z zamenjavo parametrov kontroliramo, kje želimo imeti maksimum aposteriorne porazdelitve - v smislu, da ga lahko premikamo levo in desno.

```
#povečamo parametra
narisi_vse(17,17)
```



Ugotovimo sledeče: za manjša parametra α in β se graf aposteriorne porazdelitve bolj ujema z grafom verjetja, simetrično pa z večanjem parametrov dosežemo, da je graf aposteriorne bližje apriorni porazdelitvi.

2. Naloga

Glede na to, da imamo 4 možne odgovore, je naše apriorno prepričanje, da je θ približno 0.25. Za številsko oceno našega parametra ponavadi vzamemo pričakovano, zato poiščimo α in β iz apriorne porazdelitve taka, da bo

$$E(\pi) = 0.25 = \frac{\alpha}{(\alpha + \beta)}$$

. Ta enačba ima seveda za $\alpha, \beta > 0$ nešteto rešitev (α, β) , mi pa si oglejmo dva primera - v prvem smo močno prepričani, da je $\theta = 0.25$, v drugem pa ne zares.

Verjamemo $\theta = 0.25$

Ker močno verjamemo, da je $E(\pi) = 0.25$, želimo, da je "razpršenost" podatkov čim manjša. V ta namen poskušamo minimizirati varianco beta porazdelitve, $\text{var}(\pi) = \frac{(\alpha\beta)}{((\alpha+\beta)^2(\alpha+\beta+1))}$. Opazimo, da števec narašča kvadratno, medtem pa imenoalec kubično v $\phi = \alpha + \beta$. Zato izberimo večje parametre, npr:

```
alpha1 <- 10
beta1 <- 30
varianca_beta(alpha1, beta1)
```

```
## [1] 0.004573171
```

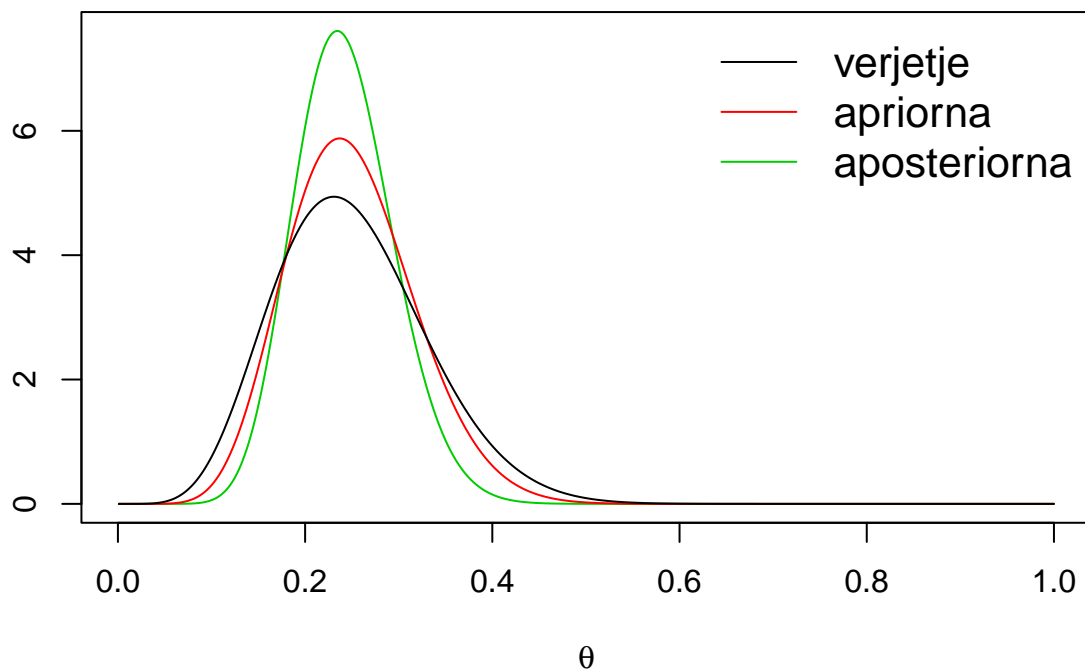
Kot omenjano, za številsko oceno parametra θ vzamemo pričakovano vrednost porazdelitve. Poglejmo torej pričakovano vrednost aposteriorne porazdelitve:

```
alpha1_apost = alpha1 + k
beta1_apost = n - k + beta1
alpha1_apost / (alpha1_apost + beta1_apost)
```

```
## [1] 0.2424242
```

Za predstavbo še narišemo in pogledamo kredibilnostne intervale:

```
narisi_vse(alpha1, beta1)
```



Kredibilnostni interval:

```
qbeta(c(0.025, 0.975), alpha1_apost, beta1_apost)
```

```
## [1] 0.1477231 0.3518967
```

Če smo močno prepričani v našo začetno oceno za θ , potem se to ne spremeni niti v aposteriorni porazdelitvi. Vsota $\phi = \alpha + \beta$ pa nam pove tudi, koliko je gostota zgoščena okrog pričakovane vrednosti in bolj kot je zgoščena, manjši bi bil tudi interval zaupanja.

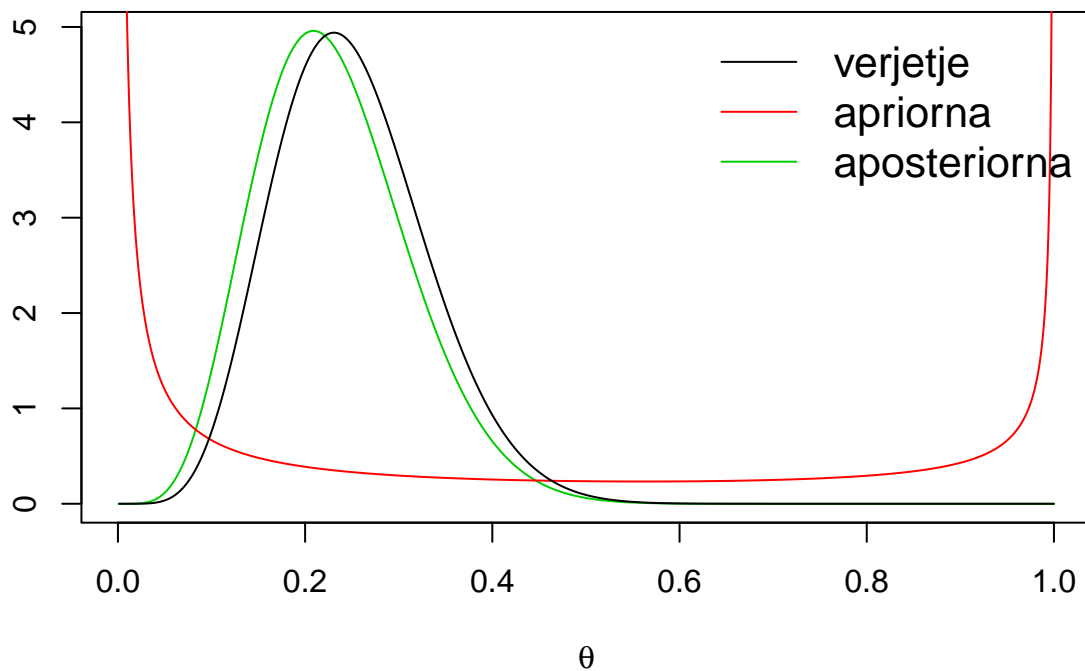
Nismo prepričani v $\theta = 0.25$

V tem primeru pa izberimo majhne parametre, narišimo grafe, ocenimo θ in pogledjmo kredibilnostni interval:

```
alpha2 = 0.1
beta2 = 0.3
varianca_beta(alpha2, beta2)
```

```
## [1] 0.1339286
```

```
narisi_vse(alpha2, beta2)
```



```
alpha2_apost = alpha2 + k
beta2_apost = n - k + beta2

#pričakovana vrednost apost
alpha2_apost / (alpha2_apost + beta2_apost)

## [1] 0.2310606

#kredibilnostni interval
qbeta(c(0.025, 0.975), alpha2_apost, beta2_apost)

## [1] 0.09458543 0.40597000
```

Kot smo pričakovali je sedaj aposteriorna ocena θ dlje od 0.25 kot v prejšnjem primeru. Prav tako je interval zaupanja občutni večji. To tudi ne preseneča, saj je tudi začetna varianca večja in podatki v tem primeru ne težijo k pričakovani vrednosti.