1. sklop: Binomski model

1 Primer

Izberite pravilni odgovor na spodnje vprasanje.

Vprasanje: Qskd senciljm dowdlq a?

(a) 25

(b) 625

(c) 1

(d) Nic od nastetega.

Zanima nas verjetnost, da odgovorimo pravilno.

2 Verjetnostni model za nas primer

Vzorec $X_1, X_2, \ldots, X_n$, kjer je:

- n stevilo studentov na vajah,
- X_i predstavlja pravilnost odgovora i-tega studenta, tj. $X_i = 1$, ce i-ti student odgovori pravilno, in $X_i = 0$, ce le-ta odgovori napacno.

Preucujemo $X_1 + X_2 + \ldots + X_n$, tj. stevilo vseh pravilnih odgovorov, ki ga oznacimo z X (druga standardna oznaka je Y v smislu izida, anglesko *outcome*).

- $X \mid \theta \sim \mathrm{Bin}(n, \theta)$
- $P(X = k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$; $k = 0, 1, ..., n$
- θ je verjetnost pravilnega odgovora – **parameter, ki nas zanima**
- $\mathrm{E}(X) = n\theta$, $\mathrm{Var}(X) = n\theta(1-\theta)$

Nas primer:

```
n <- 26
```

Nasi podatki (oznacimo s k realizacijo X na nasem vzorcu):

```
k <- 6
```

2.1 Kako bi ocenili nas parameter s "klasicno" frekventisticno statistiko? Katere metode bi lahko uporabili?

Po metodi najvecjega verjetja dobimo cenilko $\hat{\theta} = \frac{k}{n}$. Dobimo enako cenilko tudi z metodo momentov.

2.2 Bayesov model

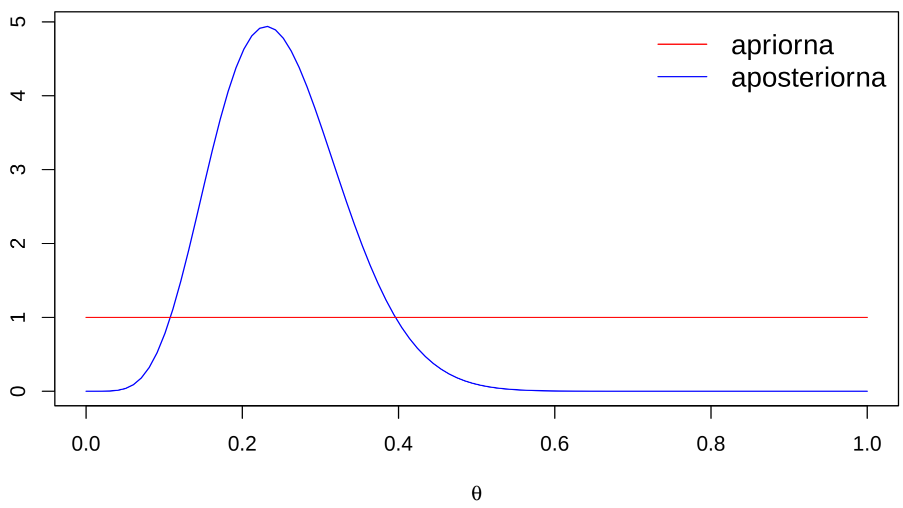
2.2.1 Opazenih je bilo 6 pravilnih izmed 26 odgovorov

2.2.2 Privzemite neinformativno apriorno Beta porazdelitev (Beta(1,1)). Izracunajte aposteriorno porazdelitev. Pomagajte si s funkcijami `dbeta`, `pbeta`

```
a <- 1
b <- 1

theta <- seq(0,1,length = 100)

apriorna_porazdelitev <- dbeta(theta, a, b)

a_apost <- a + k
b_apost <- b + n -k

aposteriorna_porazdelitev <- dbeta(theta, a_apost, b_apost)
```

2.2.3 Narisite gostoto apriorne in aposteriorne porazdelitve na istem grafu (`plot`, `lines`)

```
plot(theta, aposteriorna_porazdelitev, type='l', col='blue', ylab='', xlab=expression(th
lines(theta, apriorna_porazdelitev, col='red')
legend("topright", legend = c("apriorna","aposteriorna"),
col = c("red","blue"), lty = 1, bty = "n", cex = 1.3)
```



2.2.4 Izracunajte $P(\theta \leq 0.4)$ in $P(\theta \leq 0.4|X)$

```
pbeta(0.4, a, b)
```

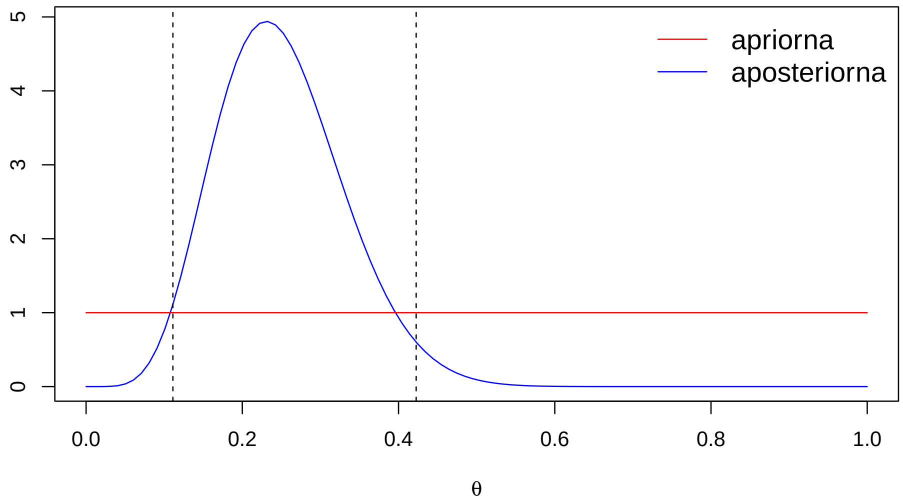
```
## [1] 0.4
```

```
pbeta(0.4, a_apost, b_apost)
```

```
## [1] 0.9579073
```

2.2.5 Izracunajte 95% centralni kredibilnostni interval. Pomagajte si s funkcijo `qbeta`. Narisite meje kredibilnostnega intervala na grafu iz tocke 3 (`abline(v= )`).

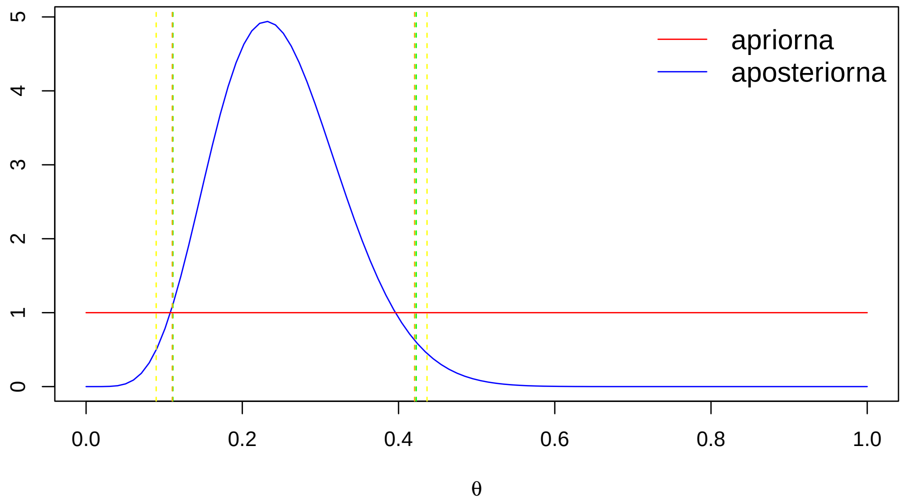
```
I1 <- qbeta(c(0.025,0.975), a_apost, b_apost)

plot(theta, aposteriorna_porazdelitev, type='l', col='blue', ylab='', xlab=expression(th
lines(theta, apriorna_porazdelitev, col='red')
legend("topright", legend = c("apriorna","aposteriorna"),
col = c("red","blue"), lty = 1, bty = "n", cex = 1.3)
abline(v=I1, lty='dashed')
```



2.2.6 Izracunajte se 95% interval zaupanja z aproksimacijo normalne porazdelitve in na podlagi metode Clopper-Pearson. Primerjajte rezultate

Opomba: pomagajte si z prop.test(k, n, correct=F)$conf in binom.test(k, n)$conf

```
I2 <- prop.test(k, n, correct=F)$conf
I3 <- binom.test(k, n)$conf

plot(theta, aposteriorna_porazdelitev, type='l', col='blue', ylab='', xlab=expression(th
lines(theta, apriorna_porazdelitev, col='red')
legend("topright", legend = c("apriorna","aposteriorna"),
col = c("red","blue"), lty = 1, bty = "n", cex = 1.3)
abline(v=I1, lty='dashed', col='green')
abline(v=I2, lty='dashed', col='orange')
abline(v=I3, lty='dashed', col='yellow')
```



2.2.7 Na predavanjih ste definirali pricakovano vrednost aposteriorne porazdelitve. Zapisite formulo in izracunajte oceno. Zapisite formulo v primeru neinformativne apriorne porazdelitve (Beta(1,1)). Ali je ocena enaka kakor pri frekventisticnem pristopu?

```
# Frekventisticen pristop:
k/n
```

```
## [1] 0.2307692
```

```
# Bayesov pristop:
a_apost / (a_apost + b_apost)
```

```
## [1] 0.25
```

2.2.8 DODATNA NALOGA: napovedovanje (ang. *prediction*)

Izpit je sestavljen iz desetih vprasanj (taksnih iz zacetka navodil tega sklopa).

1. Denimo, da bi pred zacetkom prvih vaj dali izpit v resevanje nekemu studentu. Kaj lahko povemo o porazdelitvi stevila njegovih pravilnih odgovorov?
2. Na prvih vajah smo pridobili vzorec, s katerim smo preizkusili, kako na vprasanje odgovarjamo, ce ne znamo cisto nic. Vzorec ste bili studentje, prisotni na prvih vajah. Izpit damo v resevanje studentu, **ki ni bil prisoten na prvih vajah** in se tudi ni ucil. Kaj lahko povemo o porazdelitvi stevila njegovih pravilnih odgovorov?

Odgovor na 1. vprasanje je **apriorna napovedna porazdelitev** (angl. *prior predictive distribution*).
Ta nas tipicno ne zanima.

Odgovor na 2. vprasanje je **aposteriorna napovedna porazdelitev** (angl. *posterior predictive distribution*).

Splosna formula za apriorno napovedno porazdelitev:

$$f(x_{\text{nov}}) = \int_\Theta f(x_{\text{nov}}, \theta)\, d\theta = \int_\Theta f(x_{\text{nov}} \mid \theta)\pi(\theta)\, d\theta.$$

Splosna formula za aposteriorno napovedno porazdelitev:

$$f(x_{\text{nov}} \mid x) = \int_\Theta f(x_{\text{nov}}, \theta \mid x)\, d\theta = \int_\Theta f(x_{\text{nov}} \mid \theta, x)\pi(\theta \mid x)\, d\theta = \int_\Theta f(x_{\text{nov}} \mid \theta)\pi(\theta \mid x)\, d\theta.$$

V nasem modelu (binomski model z apriorno beta porazdelitvijo) je:

- $\pi(\theta) \sim \text{Beta}(\alpha, \beta)$; izbrali smo $\alpha = 1, \beta = 1$
- $\pi(\theta \mid x) \sim \text{Beta}(\alpha_{\text{apost}}, \beta_{\text{apost}}) = \text{Beta}(k + \alpha, n - k + \beta)$; za nas vzorec velikosti $n = 26$ smo dobili $k = 6$
- za $x_{\text{nov}} \equiv K \in \{0, 1, \ldots, N\}$ je $f(x_{\text{nov}} \mid \theta) = \binom{N}{K}\theta^K(1-\theta)^{N-K}$; dolocili smo $N = 10$, zanimajo nas vsi mozni K

Izkaze se, da je iskana apriorna ali aposteriorna napovedna porazdelitev iz druzine t.i. **beta-binomske porazdelitve** (BetaBin). To je diskretna porazdelitev Y s parametri $N \in \mathbb{N}$ in $\tilde{\alpha}, \tilde{\beta} > 0$, ki lahko zavzame vrednosti $K \in \{0, 1, \ldots, N\}$ in je

$$P(Y = K) = \binom{N}{K} \frac{B(K + \tilde{\alpha}, N - K + \tilde{\beta})}{B(\tilde{\alpha}, \tilde{\beta})}.$$

Apriorna napovedna porazdelitev v binomskem modelu: $\text{BetaBin}(N, \alpha, \beta)$.

Aposteriorna napovedna porazdelitev v binomskem modelu: $\text{BetaBin}(N, \alpha_{\text{apost}}, \beta_{\text{apost}})$ oziroma $\text{BetaBin}(N, k + \alpha, n - k + \beta)$.

```r
# Beta-binomska porazdelitev v R (je vkljucena tudi v nekaterih paketih,
# ponekod drugace parametrizirana):
dbetabinom <- function(K, N, a, b){
  choose(N, K) * beta(K+a, N-K+b) / beta(a, b)
}
```

Narisite apriorno in aposteriorno napovedno porazdelitev.

Ali je to racunanje res potrebno?

- Nasa ocena parametra po upostevanju podatkov nasega vzorca je $\hat{\theta} = \alpha_{\text{apost}}/(\alpha_{\text{apost}} + \beta_{\text{apost}})$.
- Stevilo pravilnih odgovorov je porazdeljeno $\text{Bin}(10, \theta)$.
- Ali je preprosto aposteriorna porazdelitev kar $\text{Bin}(10, \hat{\theta})$? Primerjajte obe porazdelitvi