

# 1. sklop: Binomski model

## 1 Primer

Izberite pravilni odgovor na spodnje vprašanje.

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

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

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

## 2.2 Bayesov model

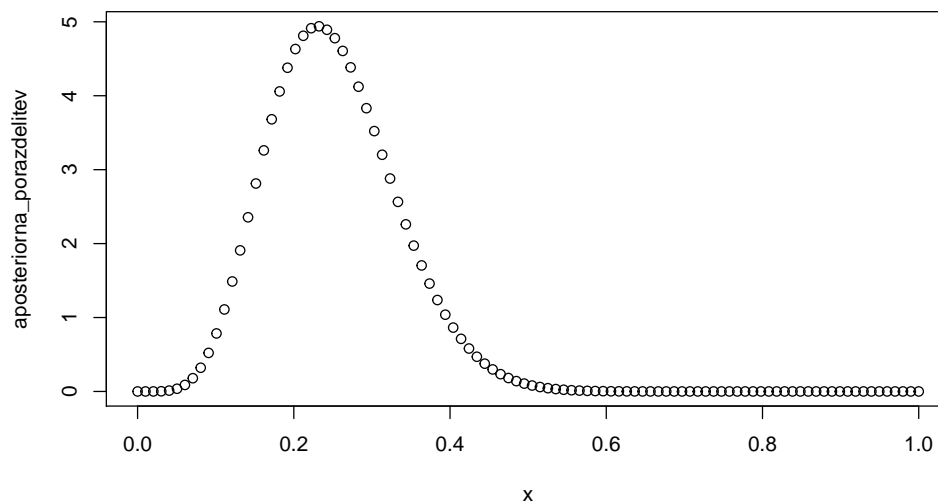
2.2.1 Opaznih je bilo 6 pravih izmed 26 odgovorov

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

```
x <- seq(0,1,length = 100)
aposteriorna_porazdelitev <- dbeta(x,1+k,1+n-k)
```

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

```
plot(x,aposteriorna_porazdelitev)
```



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

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

**2.2.6** Izračunajte 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`

**2.2.7** Na predavanjih ste definirali pričakovano vrednost aposteriorne porazdelitve. Zapišite formulo in izračunajte oceno. Zapišite formulo v primeru neinformativne apriorne porazdelitve ( $\text{Beta}(1,1)$ ). Ali je ocena enaka kakor pri frekventističnem pristopu?

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

Izpit je sestavljen iz desetih vprašanj (taksnih iz začetka navodil tega sklopa).

1. Denimo, da bi pred začetkom prvih vaj dali izpit v reševanje nekemu studentu. Kaj lahko povemo o porazdelitvi števila njegovih pravih odgovorov?
2. Na prvih vajah smo pridobili vzorec, s katerim smo preizkusili, kako na vprašanje odgovarjamo, če ne znamo čisto nič. Vzorec ste bili studentje, prisotni na prvih vajah. Izpit damo v reševanje studentu, **ki ni bil prisoten na prvih vajah** in se tudi ni učil. Kaj lahko povemo o porazdelitvi števila njegovih pravih odgovorov?

Odgovor na 1. vprašanje je **apriorna napovedna porazdelitev** (angl. *prior predictive distribution*).

Ta nas tipično ne zanima.

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

**Splosna formula za apriorno napovedno porazdelitev:**

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

**Splosna formula za aposteriorno napovedno porazdelitev:**

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

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

- $\pi(\theta) \sim \text{Beta}(\alpha, \beta)$ ; izbrali smo  $\alpha = 1, \beta = 1$
- $\pi(\theta | x) \sim \text{Beta}(\alpha_{\text{apost}}, \beta_{\text{apost}}) = \text{Beta}(k + \alpha, n - k + \beta)$ ; za nas vzorec velikosti  $n = 26$  smo dobili  $k = 6$

- za  $x_{\text{nov}} \equiv K \in \{0, 1, \dots, N\}$  je  $f(x_{\text{nov}} | \theta) = \binom{N}{K} \theta^K (1 - \theta)^{N-K}$ ; določili smo  $N = 10$ , zanimajo nas vsi možni  $K$

Izkaze se, da je iskana apriorna ali aposteriorna napovedna porazdelitev iz družine 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  $K \in \{0, 1, \dots, N\}$  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  $\text{BetaBin}(N, k + \alpha, n - k + \beta)$ .

```
# Beta-binomska porazdelitev v R (je vključena tudi v nekaterih paketih,
# ponekod drugače 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 računanje res potrebno?

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