1. sklop: Binomski model

Nina Ruzic Gorenjec

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: $X_i \sim \mathrm{Bern}(\theta) = \mathrm{Bin}(n=1, \theta)$

- 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*).

- $X \mid \theta \sim \mathrm{Bin}(n, \theta)$
- $P(X = k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$, $k = 0, 1, 2, \ldots, n$
- θ (na predavanjih ϑ) je verjetnost pravilnega odgovora – **parameter, ki nas zanima**
- $\mathrm{E}(X) = n\theta$

→ študentje neodvisni

Nas primer: → P(pravilni odg.) enaka za vsakega študenta

```r
n <- 26
```

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

```r
k <- 6
```

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

→ metoda največjega verjetja : $\hat{\theta} = \frac{k}{n}$

→ metoda momentov : dobimo enako

2.2 Bayesova formula

Na ravni dogodkov:

je proporcionalno

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \propto P(B \mid A)\, P(A).$$

Na ravni gostot z enotno oznako p (lahko bi pisali tudi f):

$$p(\theta \mid \text{podatki}) = \frac{p(\text{podatki} \mid \theta)\, p(\theta)}{p(\text{podatki})} \propto p(\text{podatki} \mid \theta)\, p(\theta).$$

V “standardnih” oznakah:

$\int = 1$

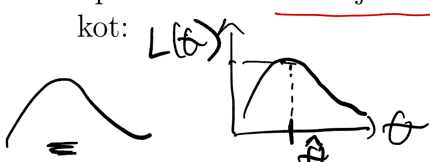
gostota

$\Rightarrow \int \pi(\theta|x)\, d\theta = 1$

$$\pi(\theta \mid x) = \frac{\boxed{f(x \mid \theta)}\, \pi(\theta)}{f(x)} \propto \boxed{f(x \mid \theta)\, \pi(\theta).}$$

Podatki x so znani. Zanima nas θ. **Poglejmo zato na zgornje kot na funkcijo θ.**

Spomnimo se funkcije verjetja (anglesko *likelihood*) $L(\theta \mid x) = L(\theta\,; x)$ in zapisimo zgornje kot:



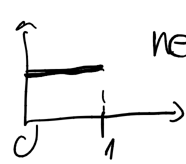
$$\boxed{\pi(\theta \mid x) \propto \boxed{L(\theta \mid x)}\, \pi(\theta).}$$

$= f(x_1,\dots,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

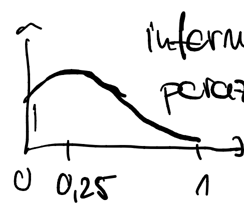
Trije gradniki Bayesove formule, ki jih bomo predstavili graficno kot funkcije θ:

- **Apriorna porazdelitev** $\pi(\theta)$ (njen integral je 1) – *nase vnaprejsnje (apriorno) vedenje/prepricanje o θ, preden zberemo podatke!*
- **Verjetje** $L(\theta \mid x)$ (potrebno mnoziti s konstanto, tako da bo integral enak 1) – *verjetnost podatkov pri razlicnih moznih vrednostih parametra θ.*
- **Aposteriorna porazdelitev** $\pi(\theta \mid x)$ – *preko Bayesove formule posodobimo nase apriorno vedenje o parametru ($\pi(\theta)$) s tem, kar nam povedo podatki ($L(\theta \mid x)$).*

$\pi(\theta) \sim \text{Unif}\,[0,1]$



neinformativna porazdelitev



informativna porazdelitev

$$L(\theta|x) = f(x_1,\dots,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}\cdot(1-\theta)^{\sum_{i=1}^n (1-x_i)} =$$

Bern(θ) $\quad x_i = \begin{cases} 1; & \text{prav} \\ 0; & \text{narobe} \end{cases}$

$$= \theta^{k}\cdot(1-\theta)^{\sum_{i=1}^n 1 - \sum_{i=1}^n x_i} = \theta^k(1-\theta)^{n-k}$$

2.3 Verjetje

Stevilo pravilnih odgovorov med n studenti je enako k:

$$L(\theta \mid x) = \theta^k\,(1-\theta)^{n-k}.$$

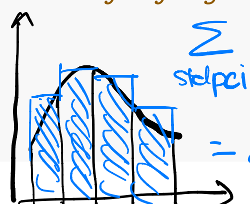
V R-u:

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

#Z mnozenjem s konst dosezemo, da je integral verjetja glede na theta enak 1.
konst <- function(k, n){
  theta <- seq(0.001, 1, 0.001)
  1 / (0.001 * sum(verjetje(theta, k, n)))
}
```

$\binom{n}{k} = \binom{26}{6} =$ konst.

$c =$ konst, $\quad$ 0.001 $= d$, $\quad \theta_i$: theta, $\quad g(\theta_i)$ = verjetje(theta, k, n)



$\sum$ višina · širina $=$
$= \sum g(\theta_i)\cdot d =$
$= d\cdot\sum g(\theta_i)$ $\quad (*)$
$\doteq \int_\theta g(\theta)\,d\theta$

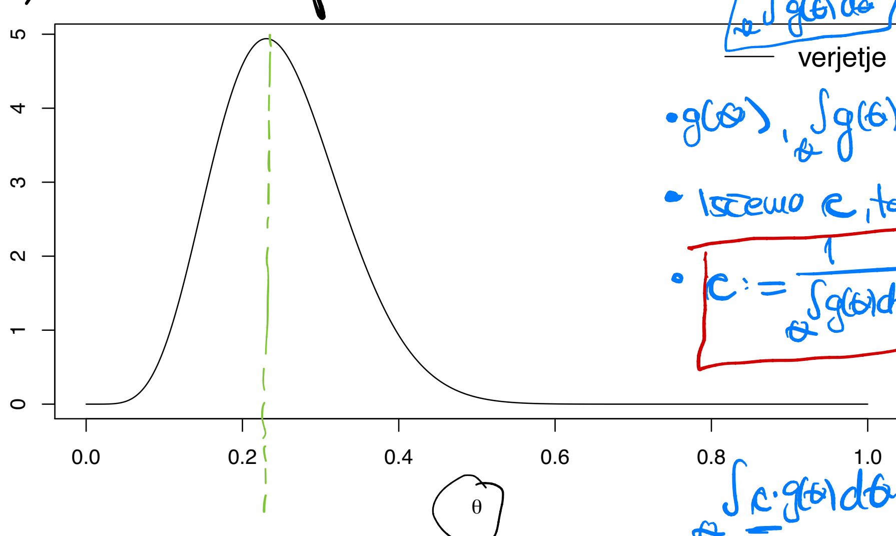
Narisemo za nas vzorec:

```
theta <- seq(0, 1, 0.001)
konst.verjetje <- konst(k, n) * verjetje(theta, k, n)
plot(theta, konst.verjetje, type = "l",
     xlab = expression(theta), ylab ="")
legend("topright", legend = c("verjetje"), col = c("black"),
       lty = 1, bty = "n", cex = 1.3)
```

Iskali konst. tako da bo $\int$ od konst g enaz 1.

$\theta^6(1-\theta)^{20}$ · normalizacijska konstanta

$\boxed{\frac{1}{\int_\theta g(\theta)\,d\theta}}\cdot g(\theta)$ iskana konstanta



- $g(\theta)$, $\int_\theta g(\theta)\,d\theta \neq 1$
- Iščemo c, tako da $\int_\theta \boxed{c\cdot g(\theta)}\,d\theta = 1$.
- $\boxed{c := \frac{1}{\int_\theta g(\theta)\,d\theta}}$, ker:

$$\int_\theta c\cdot g(\theta)\,d\theta = c\cdot\int_\theta g(\theta)\,d\theta = \frac{1}{\int_\theta g(\theta)\,d\theta}\cdot\int_\theta g(\theta)\,d\theta = 1 \checkmark$$

- $c \doteq \frac{1}{d\cdot\sum_i g(\theta_i)}$

ocena za θ po metodi največjega verjetja $= 6/26 = 0{,}23$

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

2.4 Apriorna porazdelitev

konjugirana porazdelitev

Za apriorno porazdelitev si izberemo beta porazdelitev, ki je v primeru binomske porazdelitve podatkov *conjugate prior* (pomeni, da apriorna in aposteriorna porazdelitev pripadata enaki druzini porazdelitev), zato se lahko uporablja tudi izraz **beta-binomski model**.

Za apriorno porazdelitev imamo torej gostoto beta porazdelitve pri parametrih $\alpha, \beta > 0$:

$$\pi(\theta) = \pi(\theta \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1},$$

kjer je funkcija beta $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ in je funkcija gama $\Gamma(a) = (a-1)!$ za pozitivna cela stevila a. Spomnimo se:

- $\mathrm{E}(\mathrm{Beta}(\alpha, \beta)) = \frac{\alpha}{\alpha+\beta}$,
- $\mathrm{var}(\mathrm{Beta}(\alpha, \beta)) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

$\theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^0 \cdot (1-\theta)^0 = 1$

Najprej poskusimo $\alpha = \beta = 1$, s cimer dobimo enakomerno zvezno porazdelitev U[0,1] - **neinformativna apriorna porazdelitev**.

Narisemo v R-u:

$$// \frac{1}{\int_0^1 \theta^6(1-\theta)^{20}\, d\theta} \cdot \theta^6(1-\theta)^{20} //$$

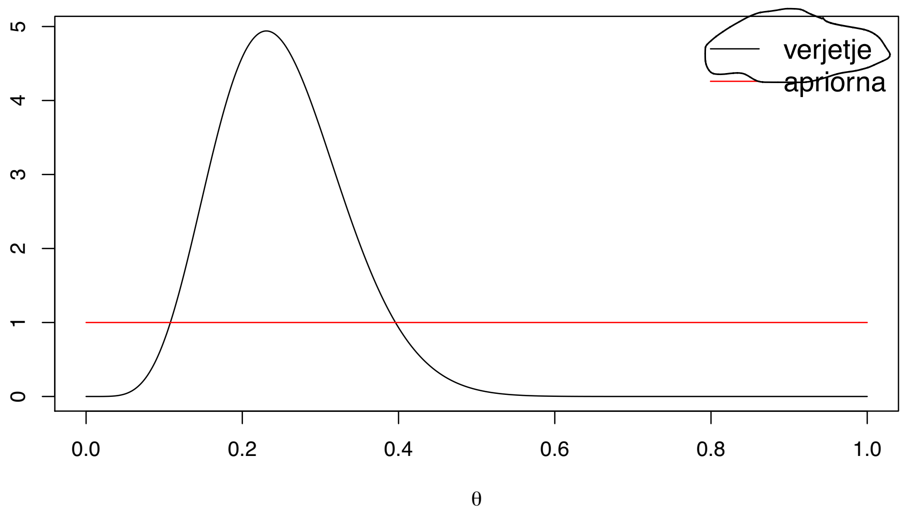
```
alpha <- 1
beta <- 1

theta <- seq(0, 1, 0.001)
apriorna <- dbeta(theta, alpha, beta)

konst.verjetje <- konst(k, n) * verjetje(theta, k, n)

y.max <- max(c(konst.verjetje, apriorna))
plot(theta, konst.verjetje, ylim = c(0, y.max), type = "l",
     xlab = expression(theta), ylab = "")
lines(theta, apriorna, col = "red")
legend("topright", legend = c("verjetje", "apriorna"), col = c("black", "red"),
       lty = 1, bty = "n", cex = 1.3)
```

← ponavljeno



2.5 Aposteriorna porazdelitev

Ker smo uporabili *conjugate prior*, bo aposteriorna porazdelitev tudi iz druzine beta porazdelitev, njena parametra sta enaka:

- $\alpha_{\text{apost}} = k + \alpha,$
- $\beta_{\text{apost}} = n - k + \beta.$

$L(\theta|x) \cdot \pi(\theta) \propto \theta^k (1-\theta)^{n-k} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1} = \theta^{\boxed{k+\alpha}-1} \cdot (1-\theta)^{\boxed{n-k+\beta}-1}$

$\sim \text{Beta}(\alpha_{apost}, \beta_{apost})$

Narisemo v R-u:

```
alpha.apost <- k + alpha
beta.apost <- n - k + beta

theta <- seq(0.001, 1, 0.001)
aposteriorna <- dbeta(theta, alpha.apost, beta.apost)

konst.verjetje <- konst(k, n) * verjetje(theta, k, n)
apriorna <- dbeta(theta, alpha, beta)

y.max <- max(c(konst.verjetje, apriorna, aposteriorna))
plot(theta, konst.verjetje, ylim=c(0, y.max), type = "l",
     xlab = expression(theta), ylab = "")
lines(theta, apriorna, col = "red")
lines(theta, aposteriorna, col = "green3")
legend("topright", legend = c("verjetje", "apriorna","aposteriorna"),
       col = c("black","red","green3"), lty = 1, bty = "n", cex = 1.3)
```

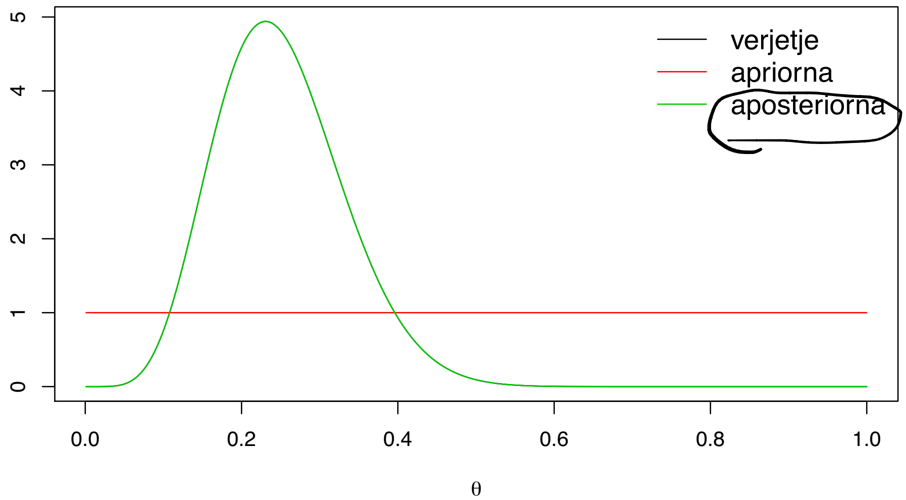
ponovili

podobno

Frekventistična: $\hat{\theta} \rightarrow$ napaka ocene (SE) $\rightarrow$ Interval zaupanja

Bayes: dobimo porazdelitev parametra tj. aposteriorna por.

$\rightarrow \hat{\theta}$



2.6 Ocena parametra θ

Druga možnost: mediana

Ena moznost je pricakovana vrednost aposteriorne porazdelitve:

$$\hat{\theta} = \frac{\alpha_{\text{apost}}}{\alpha_{\text{apost}} + \beta_{\text{apost}}} = \frac{k + \alpha}{(k + \alpha) + (n - k + \beta)} = \frac{k + \alpha}{n + \alpha + \beta}.$$

```
alpha.apost / (alpha.apost + beta.apost)
```

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

Ali dobimo enako kakor pri frekventisticnemu pristopu?

```
k/n
```

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

V primeru neinformativne porazdelitve $\alpha = \beta = 1$, dobimo

$$\hat{\theta} = \frac{k + 1}{n + 2}.$$

uteži (vsota je 1)

Na predavanjih ste $\hat{\theta}$, ocenjen preko pricakovane vrednosti aposteriorne porazdelitve, zapisali kot

$$\hat{\theta} = \frac{\phi}{\phi + n} \mu + \frac{n}{\phi + n} \cdot \frac{k}{n},$$

kjer je $\mu = \mathrm{E}(\mathrm{Beta}(\alpha, \beta)) = \frac{\alpha}{\alpha+\beta}$ in $\phi = \alpha + \beta$.

Ideja: $\hat{\theta}$ je utezeno povprecje med E(apriorna) in E(X), kjer preko ϕ kontroliramo, kako mocno verjamemo apriorni pricakovani vrednosti.

2.7 Interval (obmocje) zaupanja v Bayesovi statistiki

Pri danih podatkih $X = x$ ima interval $[L_B(x), U_B(x)]$ 95% **Bayesovo pokritje** za θ, ce velja

6/26

$$P(L_B(x) < \theta < U_B(x) \mid X = x) = 0,95.$$

kredibilnostni interval, credible interval

sluč. sprem.

Preden zberemo podatke ima interval $[L(X), U(X)]$ 95% **frekventisticno pokritje** za θ, ce velja

$$P(L(X) < \theta < U(X) \mid \theta) = 0,95.$$

$\theta \to$ poznate porazdelitev X

$\to L(X), U(X)$

Ko poznamo podatke $X = x$, jih vstavimo v $L(X)$ in $U(X)$ ter s tem dobimo $L(x)$ in $U(x)$.

$$A = \text{Koliko je } P(L(x) < \theta < U(x) \mid \theta)\ ?$$

Ce je $\theta = 0,2$, $A = 1$

$\theta = 0,5$, $A = 0$

konst = 0,1; 0,3

(ne pa verjetnost)

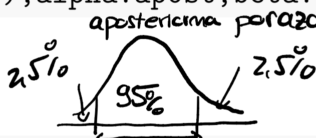
Odg.: 0 ali 1. [0,1 ; 0,3] s 95% zaupanjem, je θ med 0,1 in 0,3.

Interval zaupanja v Bayesovi statistiki (angl. pogosto *credible interval*, lahko tudi *confidence interval*) je torej katerikoli interval, v katerem "je vsebovanih 95% gostote aposteriorne porazdelitve". Seveda pa zelimo, da je "centralen glede na porazdelitev".

Najbolj preprosta varianta preko kvantilov porazdelitve (smiselna, ce je porazdelitev priblizno simetricna):

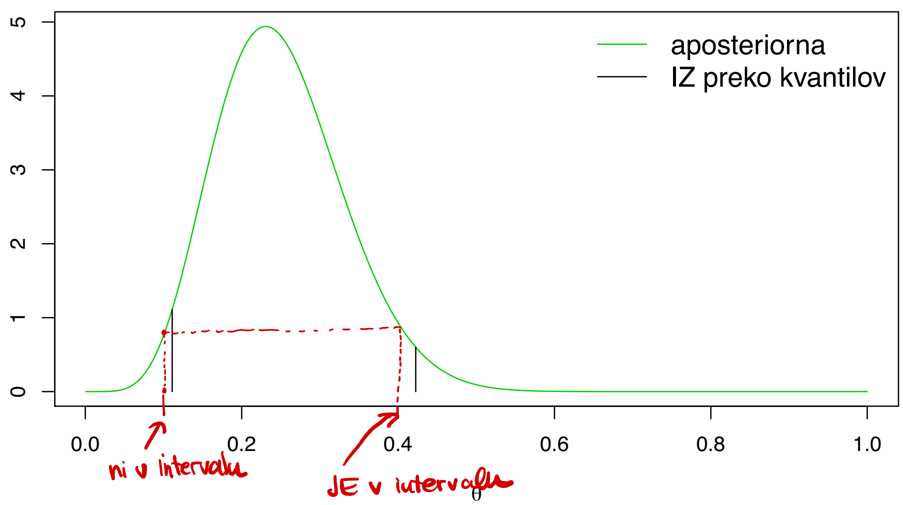
```r
(iz <- qbeta(c(0.025,0.975),alpha.apost,beta.apost))
```


```
## [1] 0.1111446 0.4225831
```



ni dobro, ker ni centralen

```r
plot(theta, aposteriorna, type = "l", col="green3",
     xlab = expression(theta), ylab = "")
segments(x0 = iz[1], y0 = 0,
         x1 = iz[1], y1 = dbeta(iz[1], alpha.apost, beta.apost))
segments(x0 = iz[2], y0 = 0,
         x1 = iz[2], y1 = dbeta(iz[2], alpha.apost, beta.apost))
legend("topright", legend = c("aposteriorna","IZ preko kvantilov"),
       col = c("green3","black"), lty = 1, bty = "n", cex = 1.3)
```



Highest posterior density (HPD) region:

apost.

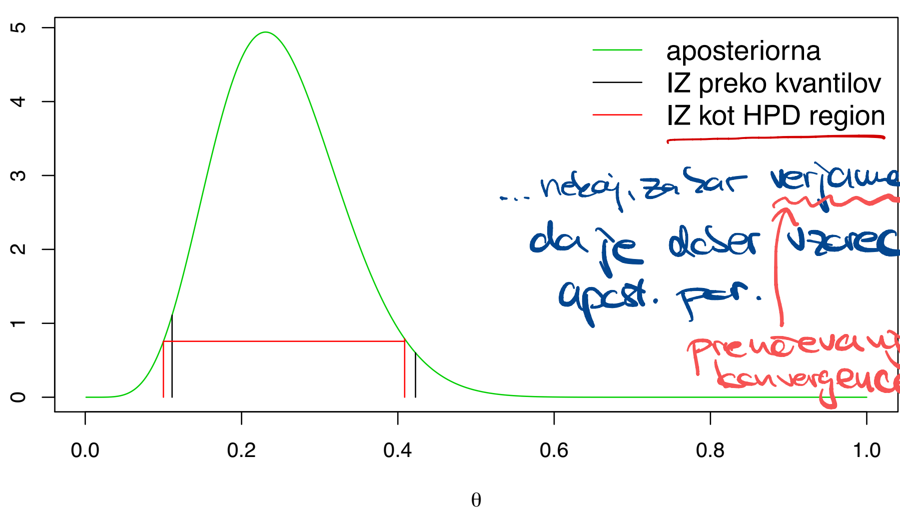
zelo velik vzorec iz porazdelitve

```
#install.packages("HDInterval")
library(HDInterval)

aposteriorna.sample <- rbeta(100000, alpha.apost, beta.apost)
(iz.hdi <- hdi(aposteriorna.sample, credMass = 0.95))
```

```
##     lower     upper
## 0.1000250 0.4086978
## attr(,"credMass")
## [1] 0.95
```

MCMC metode
→ output: vzorec določene velikosti iz apost. porazdelitve ca ....



...nekaj, za kar verjamete, da je dober vzorec iz apost. por.

preverjanje konvergence

Kaksen interval zaupanja dobimo s frekventisticnim pristopom? Katere intervale zaupanja imamo na voljo?

$\hat{p} = \frac{k}{n}$, $\hat{p} \pm 1{,}96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ približen IZ

(namesto ...=95% imamo

IZ1

```
prop.test(k, n, correct=F)$conf #IZ z aproksimacijo normalne porazdelitve
```

```
## [1] 0.1103385 0.4205155
## attr(,"conf.level")
## [1] 0.95
```

$\doteq 95\%$) $<$

IZ2

```
binom.test(k, n)$conf #Clopper-Pearsonov IZ
```

```
## [1] 0.08974011 0.43647510
## attr(,"conf.level")
## [1] 0.95
```

eksakten (pokritje kvečjemu večje od 95%)

```
iz #Bayesov IZ, metoda s kvantili
```

```
## [1] 0.1111446 0.4225831
```

vedno večja: $IZ1 \subset IZ2$

OTS

2.8 Testiranje hipotez

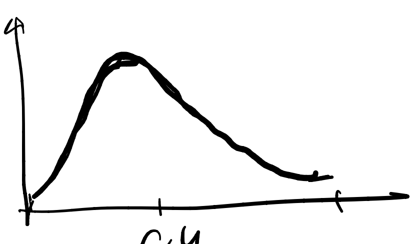
Namen stirih moznih odgovorov je bil, da je verjetnost pravilnega odgovora brez kakrsnegakoli ucenja dovolj majhna. Ali lahko sklepamo, da je verjetnost pravilnega odgovora manjsa od 0,4? Ali je $\theta < 0{,}4$?

Kako testiramo to hipotezo s Bayesovim pristopom? Kaj mora biti nicelna in kaj alternativna hipoteza?

Kako testiramo s frekventisticnim pristopom?

apriorna

π že določili → določili že aposteriorno



I. način: 95% CI; Ali je CI pod 0,4? Potem: da, lahko sklepamo, da $\theta < 0.4$.

II. način: $P(\theta < 0{,}4)$ = verjetnost naše domneve.

$P(\theta \leq 0{,}4)$

MCMC → sample = nek vektor

- mean(sample < 0,4)

|||

- sum(sample < 0,4)/length(sample)

```r
#Beayesonski pristop
pbeta(0.4, alpha.apost, beta.apost)
```

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

0,96

```r
#Test z aproksimacijo normalne porazdelitve
prop.test(k, n, p = 0.4, alternative = "less", correct = FALSE)$p.value
```

```
## [1] 0.03908454
```

```r
#Binomski eksaktni test
binom.test(k, n, p = 0.4, alternative = "less")$p.value
```

→ dualen Clopper-Pearsonovemu IZ

```
## [1] 0.05588404
```

⟹ HA velja ni $\theta < 0.4$

Frekventistična:

$H_0: \theta \geq 0{,}4$

$H_A: \theta < 0{,}4$ ← v alternativno, kar „želimo dokazati"

→ test

→ vrednost p

Če p majhna, ni verjetno, da H_0 velja pri danih podatkih.

3 Primerjava dveh neodvisnih delezev

Nase vprasanje iz zacetka navodil tega sklopa zastavimo se skupini studentov, ki se je ucila.

Od 30 studentov, jih je 21 odgovorilo pravilno.

Verjetnost pravilnega odgovora za studenta, ki se uci, naj bo θ_{uci}. Ocenimo jo z uporabo neinformativne apriorne porazdelitve (tako kakor prej, $\alpha = \beta = 1$). Dobimo aposteriorno porazdelitev

$$\theta_{\text{uci}} \sim \text{Beta}(21 + 1, 30 - 21 + 1) = \text{Beta}(22, 10).$$

Ocenimo $\hat{\theta}_{\text{uci}} = 22/(22 + 10) = 0,6875$.

Kaj smo predpostavili, da smo lahko uporabili ta model?

Prej smo ocenili verjetnost pravilnega odgovora za studenta, ki nakljucno izbere pravilni odgovor, kot

$$\theta_{\text{blef}} \sim \text{Beta}(6 + 1, 26 - 6 + 1) = \text{Beta}(7, 21).$$

Ocenimo $\hat{\theta}_{\text{blef}} = 7/(7 + 21) = 0,25$.

Zelimo primerjati ti dve verjetnosti.

Na kaksne nacine ju lahko primerjamo? Katere mere povezanosti lahko izracunamo?

- Razlika delezev (angl. *risk difference*): $\theta_{\text{uci}} - \theta_{\text{blef}}$.
- Relativno tveganje (angl. *risk ratio*): $\theta_{\text{uci}}/\theta_{\text{blef}}$.
- Razmerje obetov (angl. *odds ratio*):

$$\frac{\theta_{\text{uci}}/(1-\theta_{\text{uci}})}{\theta_{\text{blef}}/(1-\theta_{\text{blef}})}.$$

Kako ocenimo vsako izmed teh mer?

Ali znate s orodji frekventisticne statistike testirati, da je razlika delezev vecja od nic?

Kako bi testirali, da je relativno tveganje vecje od ena (studentje, ki se ucijo, imajo torej vecjo verjetnost pravilnega odgovora)? Kako bi izracunali interval zaupanja za relativno tveganje?

Frekventisticno?

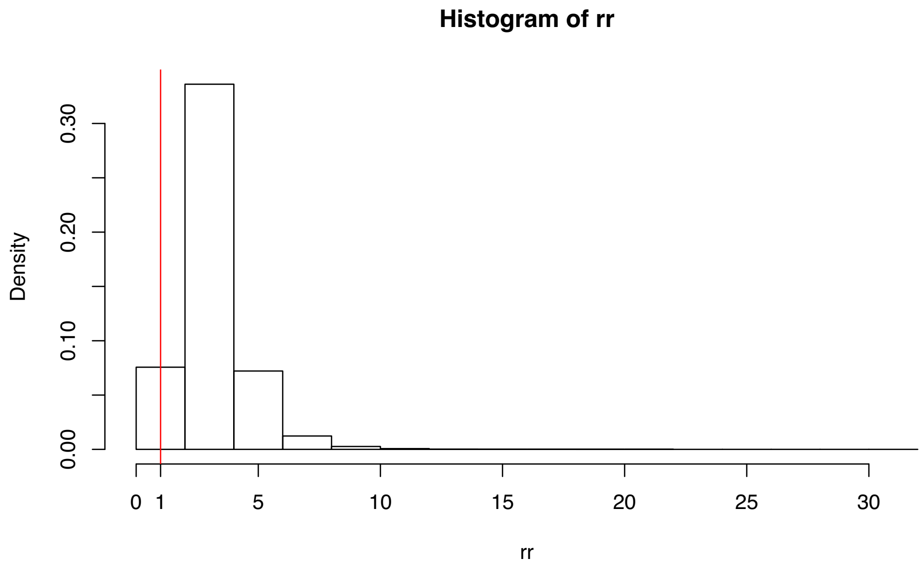
Bayesovsko?

Moramo pri teh dveh pristopih kaj izpeljati?

```r
uci <- rbeta(1000000, 22, 10)
blef <- rbeta(1000000, 7, 21)

rr <- uci/blef
```

```r
hist(rr, prob = TRUE)
axis(1, at = 1)
abline(v = 1, col = "red")
```




```r
## Ocena:
mean(rr)
```

```
## [1] 3.092614
```

```r
## Interval zaupanja:
quantile(rr, probs = c(0.025, 0.975)) # preko kvantilov
```

```
##     2.5%    97.5%
## 1.523910 6.323553
```

```r
hdi(aposteriorna.sample, credMass = 0.95) #preko HPD region
```

```
##     lower     upper
## 0.1000250 0.4086978
## attr(,"credMass")
## [1] 0.95
```

```r
## Verjetnost hipoteze, da ucenje pomaga:
sum(rr>1)/length(rr)
```

```
## [1] 0.999793
```

4 Napovedovanje (angl. *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: $\text{BetaBin}(N, \alpha, \beta)$.

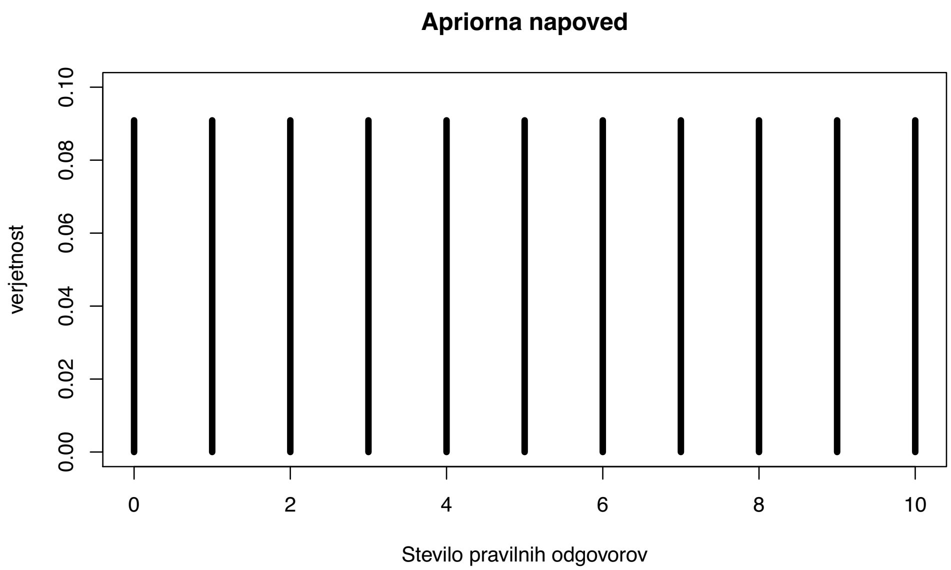
Aposteriorna napovedna porazdelitev v binomskem modelu: $\text{BetaBin}(N, \alpha_{\text{apost}}, \beta_{\text{apost}})$ oziroma $\text{BetaBin}(N, k + \alpha, n - k + \beta)$.

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)
}
```

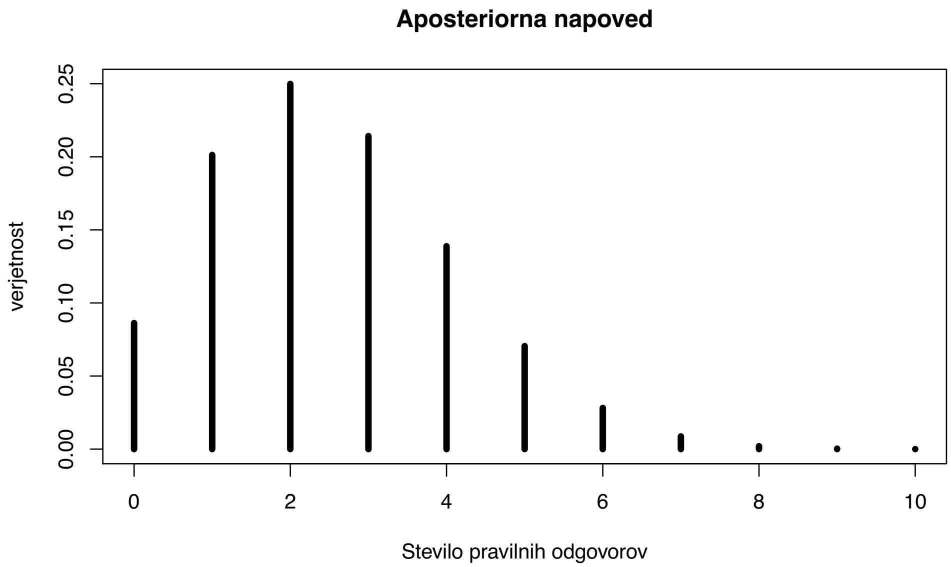
Narisemo apriorno napovedno porazdelitev.

```
plot(0:10, dbetabinom(0:10, N = 10, a = alpha, b = beta), type = "h",
     xlab = "Stevilo pravilnih odgovorov", ylab = "verjetnost",
     main = "Apriorna napoved", ylim = c(0, 0.1), lwd = 5)
```



Narisemo aposteriorno napovedno porazdelitev.

```
plot(0:10, dbetabinom(0:10, N = 10, a = alpha.apost, b = beta.apost), type = "h",
     xlab = "Stevilo pravilnih odgovorov", ylab = "verjetnost",
     main = "Aposteriorna napoved", lwd = 5)
```



Ali je bilo vse to racunanje res potrebno?

- Nasa ocena parametra po upostevanju podatkov nasega vzorca je $\hat{\theta} = \alpha_{\text{apost}}/(\alpha_{\text{apost}} + \beta_{\text{apost}}) = 0.25$.
- Stevilo pravilnih odgovorov je porazdeljeno $\text{Bin}(10, \theta)$.
- Ali je preprosto aposteriorna porazdelitev kar $\text{Bin}(10, \hat{\theta})$?

5 Jeffreyeva apriorna porazdelitev

To je **neinformativna** apriorna porazdelitev, ki je proporcionalna $\sqrt{\det \mathcal{I}(\theta)}$. Zanjo je znacilno, da je invariantna glede na razlicne reparametrizacije prostora parametrov.

Pri binomskem modelu je Jeffreyeva apriorna porazdelitev Beta($\alpha = 0.5$, $\beta = 0.5$).

Pri nasih podatkih dobimo: