2. sklop: Poissonov model

1 Podatki

Imamo zgodovinske podatke o stevilu rojstev cetverckov na leto v Prusiji za obdobje 69 let (Ladislaus von Bortkiewicz), za katere je znano, da se dobro prilegajo Poissonovi porazdelitvi.

```
(podatki <- data.frame(stevilo.cetverckov = 0:6,
                       stevilo.let = c(14,24,17,9,2,2,1)))
```

```
##   stevilo.cetverckov stevilo.let
## 1                  0          14
## 2                  1          24
## 3                  2          17
## 4                  3           9
## 5                  4           2
## 6                  5           2
## 7                  6           1
```

2 Verjetnostni model za nas primer

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

- $n = 69$ stevilo let,
- X_i predstavlja stevilo rojstev cetverckov v i-tem letu,
- $X_i \mid \lambda \sim \text{Poiss}(\lambda)$,
- $P(X_i = k \mid \lambda) = \frac{1}{k!}\lambda^k e^{-\lambda}$ za $k \in \{0, 1, 2, \ldots\}$,
- $\text{E}(X_i) = \lambda$ (**parameter, ki nas zanima**),
- $\text{Var}(X_i) = \lambda$.

Za vsako leto imamo sledece stevilo rojenih cetverckov:

```
(x <- rep(podatki$stevilo.cetverckov, podatki$stevilo.let))
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 5 5 6
```

Vse in se vec o Poissonovi porazdelitvi najdete tu (pripravil doc. dr. Gaj Vidmar):

- clanek: http://ims.mf.uni-lj.si/archive/17(2)/31.pdf

3 Naloge

3.1 Kako bi ocenili parameter λ s "klasicno" frekventisticno statistiko? Katere metode bi lahko uporabili?

3.2 Ocenjevanje v Bayesovi statistiki

Bayesova formula:

$$f(\vartheta|\vec{x}) = \frac{f(\vec{x}|\vartheta)f(\vartheta)}{f(\vec{x})} \propto f(\vec{x}|\vartheta)f(\vartheta).$$

Opazeni podatki:

- n: stevilo let
- $k = \sum_{i=1}^{n} X_i$: stevilo cetverckov

```
(n <- length(x))
```

```
## [1] 69
```

```
(k <- sum(x))
```

```
## [1] 109
```

3.2.1 Izracunajte parametre apriorne porazdelitve

Za apriorno porazdelitev izberite Gamo porazdelitev, ki je v primeru Poissonove porazdelitve podatkov konjugirana porazdelitev (ang. *conjugate prior*; pomeni, da apriorna in aposteriorna porazdelitev pripadata enaki druzini porazdelitev), zato se lahko uporablja tudi izraz **Gama-Poissonov model**.

Gostota Gama porazdelitve pri parametrih $\alpha, \beta > 0$:

$$f(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda},$$

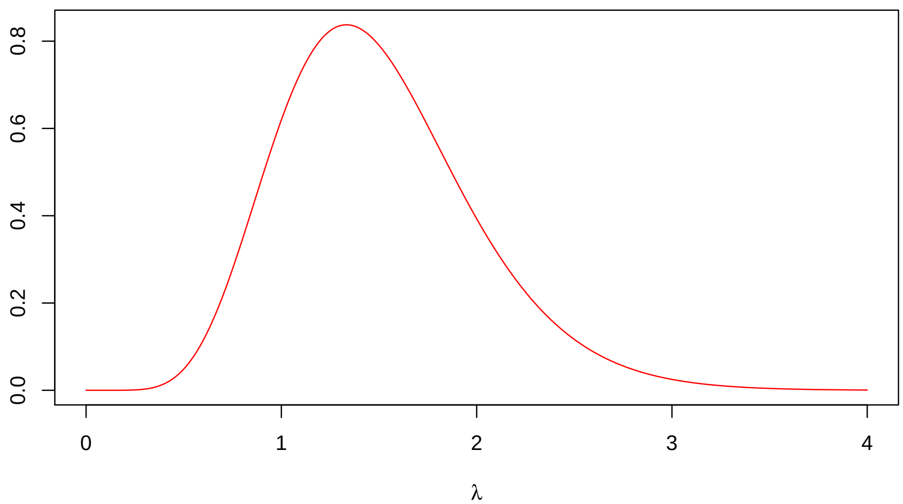
- $\mathrm{E}(\mathrm{Gama}(\alpha, \beta)) = \frac{\alpha}{\beta}$,
- $\mathrm{Var}(\mathrm{Gama}(\alpha, \beta)) = \frac{\alpha}{\beta^2}$.

Izracunajte α in β apriorne porazdelitve, pri cemer upostevajte, da je povprecje apriorne porazdelitve enako 1.5 (torej toliksno je povprecno stevilo rojenih cetverckov na leto), medtem ko standardni odklon znasa 0.5.

```
E <- 1.5
SD <- 0.5

beta <- E/SD/SD
alpha <- E * beta
```

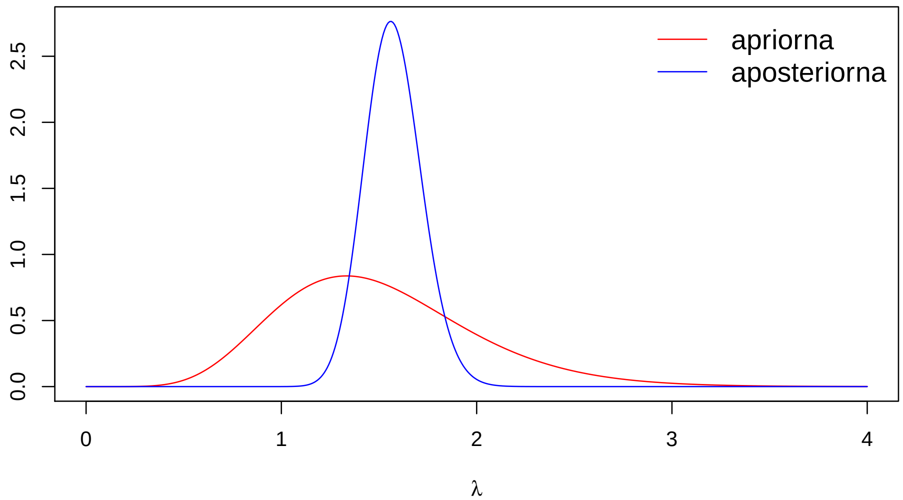
3.2.2 Narisite gostoto apriorne porazdelitve (funkcija `dgamma`)

```
lambda <- seq(0, 4, length=500)
apriorna_porazdelitev <- dgamma(lambda, alpha, beta)

plot(lambda, apriorna_porazdelitev, type='l', col='red', ylab='',
     xlab=expression(lambda))
```



3.2.3 Vemo, da je aposteriorna porazdelitev porazdeljena Gama. Izracunajte parametre aposteriorne porazdelitve in narisite njeno gostoto (na istem grafu kot za apriorno porazdelitev)

```
alpha_apost <- alpha + k
beta_apost <- beta + n
aposteriorna_porazdelitev <- dgamma(lambda, alpha_apost, beta_apost)

plot(lambda, apriorna_porazdelitev, type='l', col='red', ylab='',
     xlab=expression(lambda),
     ylim=c(0, max(c(max(apriorna_porazdelitev),
                     max(aposteriorna_porazdelitev)))))
lines(lambda, aposteriorna_porazdelitev, col='blue')
legend("topright", legend = c("apriorna","aposteriorna"),
       col = c("red","blue"), lty = 1, bty = "n", cex = 1.3)
```



3.2.4 Izracunajte $P(1.1 \leq \lambda \leq 1.9)$ in $P(1.1 \leq \lambda \leq 1.9|\vec{x})$

```
sum(pgamma(c(1.1, 1.9), alpha, beta) * c(-1, 1))
```

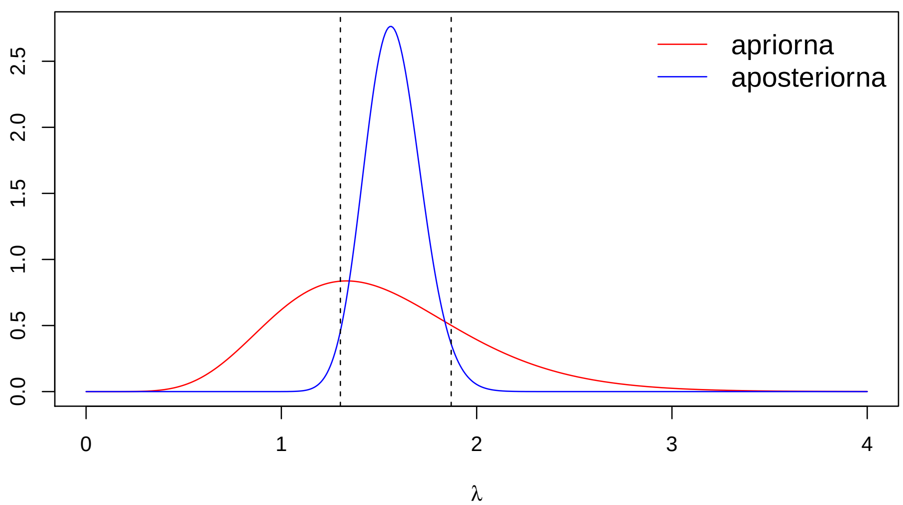
```
## [1] 0.5811641
```

```
sum(pgamma(c(1.1, 1.9), alpha_apost, beta_apost) * c(-1, 1))
```

```
## [1] 0.9839279
```

3.2.5 Izracunajte 95% centralni kredibilnostni interval za λ. Pomagajte si s funkcijo `qgamma`. Narisite meje kredibilnostnega intervala na grafu iz tocke 3.2.3 (`abline(v= )`).

```
I <- qgamma(c(0.025,0.975), alpha_apost, beta_apost)

plot(lambda, apriorna_porazdelitev, type='l', col='red', ylab='',
     xlab=expression(lambda),
     ylim=c(0, max(c(max(apriorna_porazdelitev),
                     max(aposteriorna_porazdelitev)))))
lines(lambda, aposteriorna_porazdelitev, col='blue')
legend("topright", legend = c("apriorna","aposteriorna"),
       col = c("red","blue"), lty = 1, bty = "n", cex = 1.3)
abline(v=I, lty='dashed')
```



3.2.6 Srediscna cena parametra λ

Na predavanjih ste zapisali, da je pricakovana vrednost aposteriorne porazdelitve (ki je hkrati ocena za λ, oznacimo jo z $\hat{\lambda}$) enaka

$$\hat{\lambda} = \frac{\beta}{\beta + n} \cdot \frac{\alpha}{\beta} + \frac{n}{\beta + n} \cdot \frac{k}{n}.$$

Primerjajte $\hat{\lambda}$, apriorno oceno za λ (to je $\frac{\alpha}{\beta}$) in frekventisticno oceno $\frac{k}{n}$ na danem vzorcu.

Kateri vrednosti je ocena $\hat{\lambda}$ blizje? Zakaj?

```
lambda_apriorna <- alpha/beta
lambda_aposteriorna <- beta/(beta+n)*alpha/beta + n/(beta+n)*k/n
lambda_frekventisticna <- k/n
```

3.2.7 Kolikšna je verjetnost, da se v povprecju na leto rodijo eni do dvoji cetvorcki?

3.2.8 DODATNA NALOGA: Napovedovanje

Zanima nas, kaj lahko povemo o stevilu cetvorckov v prihajajocem letu ob upostevanju podatkov zadnjih 69 let, tj. zanima nas **aposteriorna napovedna porazdelitev**.

(Ce bi nas zanimalo stevilo cetvockov v prihajajocem letu brez upostevanja podatkov 69 let, potem bi nas zanimala **apriorna napovedna porazdelitev**).

V Poissonovem modelu z apriorno Gama porazdelitvijo lahko hitro izpeljemo apriorno/aposteriorno napovedno porazdelitev:

$$P(Y = K) = \frac{\Gamma(K + \tilde{\alpha})}{\Gamma(\tilde{\alpha})\, K!} \tilde{\beta}^{\tilde{\alpha}} / (\tilde{\beta} + 1)^{K + \tilde{\alpha}} \quad \text{za } K \in \{0, 1, 2, \ldots\}.$$

To je ravno negativna binomska porazdelitev s parametroma $r = \tilde{\alpha}$ in $p = 1/(1+\tilde{\beta})$, zasledimo pa lahko tudi poimenovanje **Gama-Poissonova porazdelitev**.

Za $\tilde{\alpha}, \tilde{\beta}$ vstavimo primerna parametra Gama apriorne oz. aposteriorne porazdelitve.

- Izracunajte apriorno in aposteriorno napovedno porazdelitev.
- Poglejte, kaksna je razlika med pravilno izracunano aposteriorno napovedno porazdelitvijo in tisto, ki jo dobimo, ce v Poissonovo porazdelitev vstavimo naso oceno parametra $\hat{\lambda} = \alpha_{\text{apost}}/\beta_{\text{apost}}$.