

Бајесова метода

7.час

19. април 2016.

Први корак у Бајесовој методи је да се одреди априорна оцена $\pi(\theta)$ за параметар θ .

Апостериорна оцена параметра θ ако знамо $y_1, \dots, y_n$

$$\pi(\theta|y_1, \dots, y_n) = \frac{f(y_1, \dots, y_n|\theta)\pi(\theta)}{f(y_1, \dots, y_n)} = \frac{[\prod_i f(y_i|\theta)] \pi(\theta)}{\int_{\Omega} [\prod_i f(y_i|\theta)] \pi(\theta) d\theta}$$

Пример - Биномни модел

Вршено је истраживање чији је циљ био да одреди да ли су жене старости од 65 година срећне. Нека је $Y_i = 1$ ако испитаник је генерално задовољан и нека $Y_i = 0$ у супротном. На располагању имамо узорак обима $n = 129$.

Из претходног закључујемо да условну вероватноћу за $(y_1, y_2, \dots, y_{129})$, ако је θ познато, можемо изразити као:

$$p(y_1, y_2, \dots, y_{129} | \theta) = \theta^{\sum_{i=1}^{129} y_i} (1 - \theta)^{129 - \sum_{i=1}^{129} y_i}.$$

Даље је потребно одредити априорну расподелу за параметар θ .

Пример - Биномни модел

Параметар θ је неки непознат број између 0 и 1. Претпоставимо да нам је од раније познато да су сви подинтервали интервала $[0, 1]$ једнаке дужине и сви имају исту вероватноћу, односно

$$P\{a \leq \theta \leq b\} = P\{a + c \leq \theta \leq b + c\}, \quad 0 \leq a < b < b + c \leq 1.$$

Из претходног услова можемо закључити да параметар θ мора имати униформну расподелу:

$$p(\theta) = 1, \quad \theta \in [0, 1].$$

Пример - Биномни модел

За априору расподелу параметра θ Бајесово правило даје

$$p(\theta|y_1, y_2, \dots, y_{129}) = \frac{p(y_1, y_2, \dots, y_{129}|\theta)p(\theta)}{p(y_1, y_2, \dots, y_{129})}$$

$$p(\theta|y_1, y_2, \dots, y_{129}) = p(y_1, y_2, \dots, y_{129}|\theta) \cdot \frac{1}{p(y_1, y_2, \dots, y_{129})}$$

Из последње једнакости можемо закључити да су $p(\theta|y_1, y_2, \dots, y_{129})$ и $p(y_1, y_2, \dots, y_{129}|\theta)$ пропорционалне као функције зато што је расподела $p(\theta|y_1, y_2, \dots, y_{129})$ једнака $p(y_1, y_2, \dots, y_{129}|\theta)$ подељено са нечим што не зависи од θ . То значи да ове две функције имају исти облик, али не и исте коефицијенте.

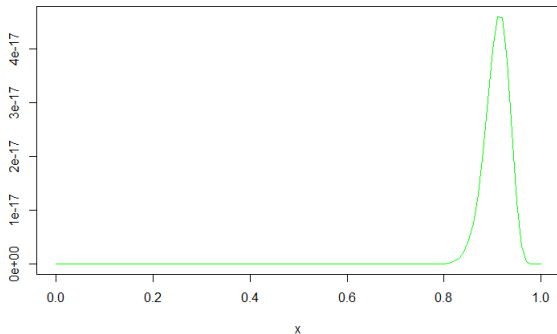
Пример - Биномни модел

На основу подаката које имамо на располагању

- 129 жена је испитано
- 118 жена је генерално срећно (91%)
- 11 жена није срећно (9%)

можемо да израчунамо вероватноћу $p(y_1, y_2, \dots, y_{129} | \theta) = \theta^{118}(1 - \theta)^{11}$.

```
> f <- function(x) x^118*(1-x)^11  
> curve(f, col = 'green')
```



Пример - Биномни модел

Примењујући Бајесово правило добијамо следећу једнакост:

$$p(\theta|y_1, y_2, \dots, y_{129}) = \theta^{118}(1 - \theta)^{11} \cdot \frac{1}{p(y_1, y_2, \dots, y_{129})}.$$

Даље је потребно да се одреди коефицијент условне расподеле $p(\theta|y_1, y_2, \dots, y_{129})$ користећи особину густине расподеле

$$\int_0^1 p(\theta|y_1, y_2, \dots, y_{129})d\theta = 1.$$

Пример - Биномни модел

$$1 = \int_0^1 p(\theta|y_1, y_2, \dots, y_{129})d\theta$$

$$1 = \int_0^1 \theta^{118}(1 - \theta)^{11}/p(y_1, y_2, \dots, y_{129})d\theta$$

$$1 = \frac{1}{p(y_1, y_2, \dots, y_{129})} \int_0^1 \theta^{118}(1 - \theta)^{11}d\theta$$

$$1 = \frac{1}{p(y_1, y_2, \dots, y_{129})} \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

$$p(y_1, y_2, \dots, y_{129}) = \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

Пример - Биномни модел

Коначно добијамо да је:

$$p(\theta|y_1, y_2, \dots, y_{129}) = \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)}\theta^{118}(1 - \theta)^{11},$$

што се може написати као

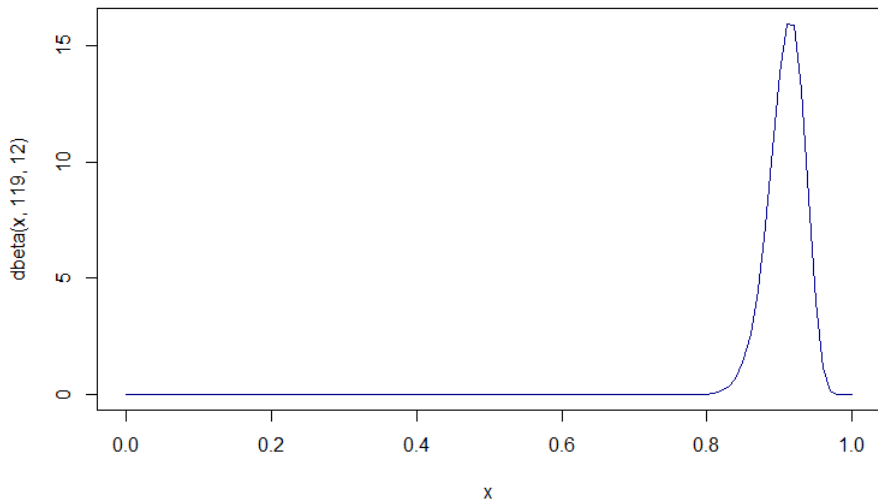
$$p(\theta|y_1, y_2, \dots, y_{129}) = \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)}\theta^{119-1}(1 - \theta)^{12-1}.$$

Дакле, апостериорна расподела за θ је бета расподела са параметрима $a = 119$ и $b = 12$.

$$\theta|y_1, y_2, \dots, y_{129} \sim \text{beta}(119, 12)$$

Пример - Биномни модел

```
curve(dbeta(x, 119, 12), col = 'darkblue')
```



Познато је да ако претпоставимо да случајне величине $Y_1, \dots, Y_n | \theta$ имају Бернулијове расподелу са параметром θ , тада $Y = \sum_{i=1}^n Y_i$ има биномну расподелу $\mathcal{B}(n, \theta)$.

$$P\{Y = y | \theta\} = \binom{n}{y} \theta^y (1 - \theta)^{(n-y)}$$

Сматрајући да је $Y = y$ можемо да израчунамо апостериорну оцену параметра θ .

$$p(\theta | y) = \frac{\binom{n}{y} \theta^y (1 - \theta)^{n-y} p(\theta)}{p(y)} = c(y) \theta^y (1 - \theta)^{n-y} p(\theta),$$

где је $c(y)$ функција од y , али не и од θ .

Користећи особину вероватноће

$$1 = \int_0^1 c(y)\theta^y(1-\theta)^{n-y}d\theta,$$

добијамо да је

$$c(y) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}.$$

Коначно добијамо да је

$$p(\theta|y) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y},$$

односно $\theta|y \sim \text{beta}(y+1, n-y+1)$.

У претходном примеру смо претпоставили да је априорна расподела параметра θ униформна расподела на интервалу $[0, 1]$. Ова расподела се може посматрати као бета расподела са параметрима $a = 1$ и $b = 1$.

$$p(\theta) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\theta^{1-1}(1-\theta)^{1-1} = 1$$

Дакле, ако $Y \sim \mathcal{B}(n, \theta)$ и $\theta \sim \text{beta}(1, 1)$, тада

$$\theta|Y = y \sim \text{beta}(1 + y, 1 + n - y).$$

Важи и у општем случају, ако $Y \sim \mathcal{B}(n, \theta)$ и $\theta \sim \text{beta}(a, b)$, тада

$$\theta|Y = y \sim \text{beta}(a + y, b + n - y).$$

Пример - процена вероватноће неког догађаја

Претпоставимо да нас занима распрострањеност заразне болести у неком малом граду. На располагању имамо случајан узорак који се састоји од 20 особа из тог града за које је испитано да ли су инфицирани.

Нека је θ случајна величина која представља удео заражених особа и Y случајна величина која представља укупан број заражених особа, тада је

$$\theta \in [0, 1] \quad Y = \{0, 1, \dots, 20\}.$$

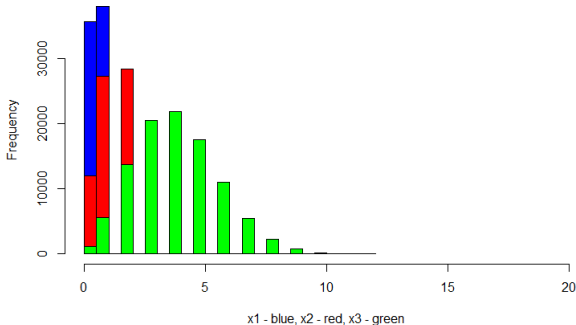
Ако је позната вредност θ , тада случајна величина има $Y|\theta$ има биномну расподелу:

$$Y|\theta \sim \mathcal{B}(20, \theta).$$

Пример - процена вероватноће неког догађаја

```
> hist(x2, col="red", add=TRUE)
> hist(x3, offset=0.66, col="green", add=TRUE)
> x1 <- rbinom(100000, 20, 0.05)
> x2 <- rbinom(100000, 20, 0.1)
> x3 <- rbinom(100000, 20, 0.2)
> hist(x1, col="blue", xlim=c(0,20),
+ main="Histogram of x1, x2 & x3", xlab="x1 - blue, x2 - red, x3 - green")
> hist(x2, col="red", add=TRUE)
> hist(x3, col="green", add=TRUE)
```

Histogram of x1, x2 & x3



Пример - процена вероватноће неког догађаја

Нека друга истрживања из разних делова земље показују да је стопа заражених особа од 0,05 до 0,20 са просеком око 0,1. Ове претходне информације треба да нам укажу коју расподелу да претпоставимо за θ . Међутим, постоји много расподела са овим особинама.

Претпоставимо да θ има бета расподелу:

$$\theta \sim \text{beta}(2, 20).$$

Тада важи и следеће:

$$E(\theta) = 0.09$$

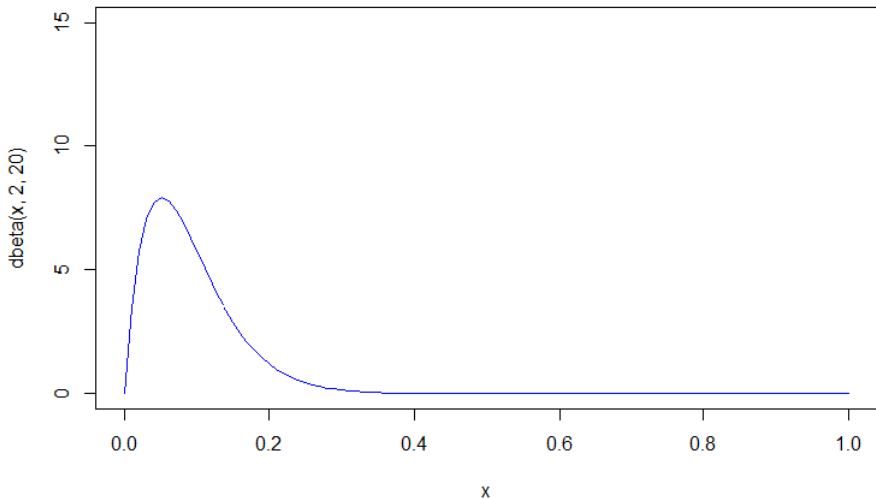
$$\text{mod}(\theta) = 0.05$$

$$P\{\theta < 0.10\} = 0.64$$

$$P\{0.05 < \theta < 0.20\} = 0.66$$

Пример - процена вероватноће неког догађаја

```
curve(dbeta(x, 2, 20), ylim = c(0,15), col = 'blue')
```



Пример - процена вероватноће неког догађаја

Ако $Y|\theta \sim \mathcal{B}(n, \theta)$ и $\theta \sim \text{beta}(a, b)$, тада је апостериорна расподела $\theta|Y \sim \text{beta}(a + y, b + n - y)$.

Приметимо да ако је $Y = 0$, тада је апостериорна оцена за θ бета расподела са параметрима 2 и 40.

$$\theta|\{Y = 0\} \sim \text{beta}(2, 40)$$

Тада важи и следеће:

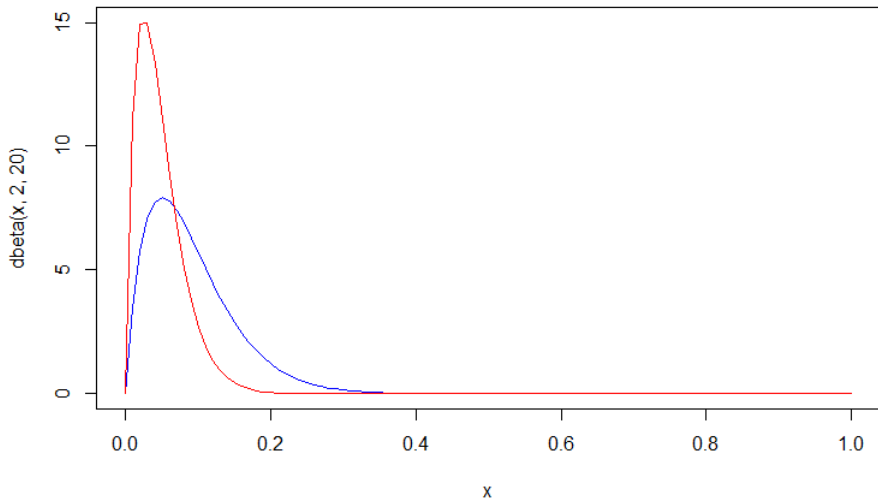
$$E(\theta|Y = 0) = 0.048$$

$$\text{mod}(\theta|Y = 0) = 0.025$$

$$P\{\theta < 0.10|Y = 0\} = 0.93$$

Пример - процена вероватноће неког догађаја

```
curve(dbeta(x, 2, 40), ylim = c(0,15), col = 'red', add = T)
```



Хвала на пажњи!

