Rešitve - cenilke, nepristranskost, doslednost

Nataša Kejžar

Naloga 1 - Hardy-Weinberg

X	AA	Aa	aa
p(X)	$(1-\theta)^2$	$2\theta(1-\theta)$	θ^2
frekvence	n_{AA}	n_{Aa}	n_{aa}

Velja, da je velikost vzorca (fiksna številka) $n = n_{AA} + n_{Aa} + n_{aa}$. Številke n_{ij} se glede na vzorec, spreminjajo, torej so spremenljivke, zato jih bomo pri izrazu za cenilko zapisali z velikimi črkami.

Intuitivna cenilka: $\hat{\theta} = \sqrt{\frac{N_{aa}}{n}}$

Pričakovana vrednost

$$\begin{aligned} E(\hat{\theta}) &= E(\sqrt{\frac{N_{aa}}{n}}) \\ &= \frac{1}{\sqrt{n}} E(\sqrt{N_{aa}}) \end{aligned}$$

Uporabimo Jensenovo neenakost za konkavne funkcije in dejstvo, da je N_{aa} porazdeljena po $Bin(n, \theta^2)$. Velja, če $Y \sim Bin(n, \pi)$, potem $E(Y) = n\pi$.

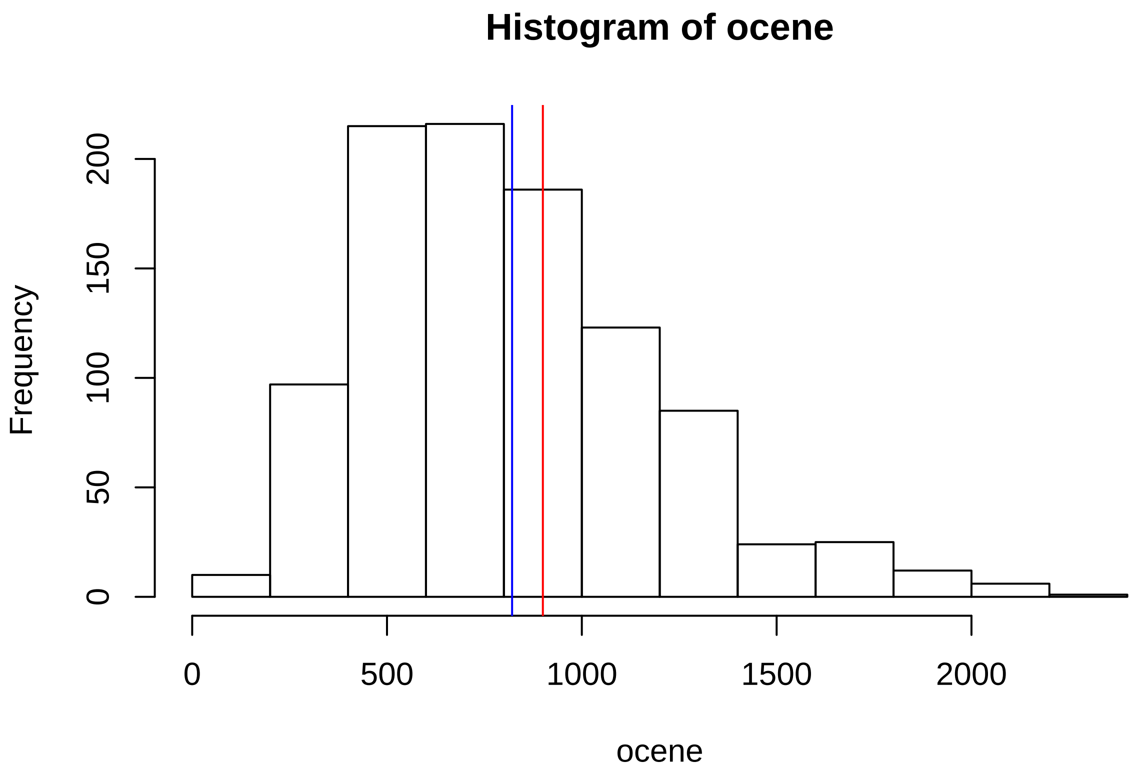
$$\begin{aligned} E(\hat{\theta}) &\leq \frac{1}{\sqrt{n}} \sqrt{E(N_{aa})} \\ &\leq \theta \end{aligned}$$

Cenilka torej podcenjuje θ, ni nepristranska.

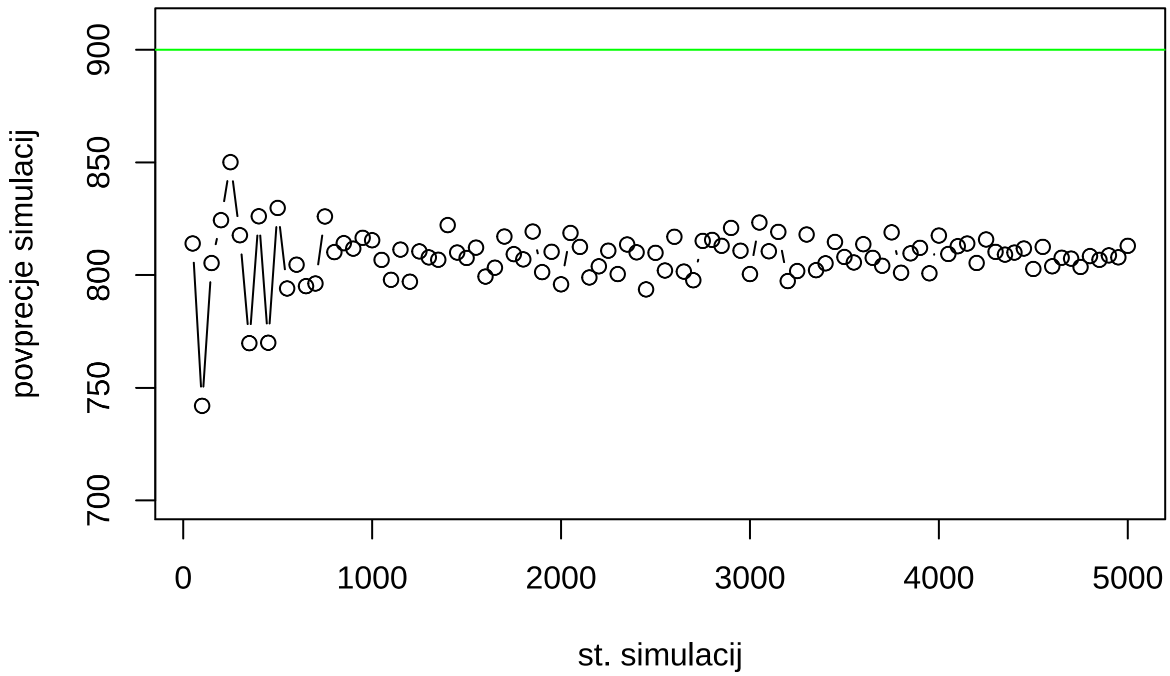
Naloga 2 - pristranska varianca

```r
# definiramo funkcijo za izracun cenilke
cenilka = function(x){sum((x-mean(x))^2)/length(x)}
# definiramo funkcijo za simulacijo n vzorcev in izracun n ocen
simN = function(n){
  ocene = NULL # inicializiramo vektor simuliranih ocen
  for(i in 1:n){
    vzorec = rnorm(n = 10,mean = 120, sd=30) # generiramo vzorec
    ocene = c(ocene,cenilka(vzorec))} # izracunamo cenilko
  return(ocene) #vrnemo vektor ocen
}
# izracun ocen, histogram + populacijsko vrednost + povprecje
```

```r
ocene = simN(1000)
hist(ocene); abline(v=30^2,col="red")
abline(v=mean(ocene),col="blue")
```




```r
# natancnost simulacije
# pozenemo za razlicno veliko vzorcev
# vec vzorcev, bolj natancna ocena pricakovane vrednosti
stVzorcev = seq(50,5000,by = 50)
povpr = NULL
for(i in stVzorcev){
  povpr = c(povpr,mean(simN(i)))
}
plot(stVzorcev,povpr,xlab="st. simulacij",
     ylab="povprecje simulacij",main="",type="b",ylim=c(700,910))
abline(h=30^2,col="green") # populacijska pricakovana vrednost
```



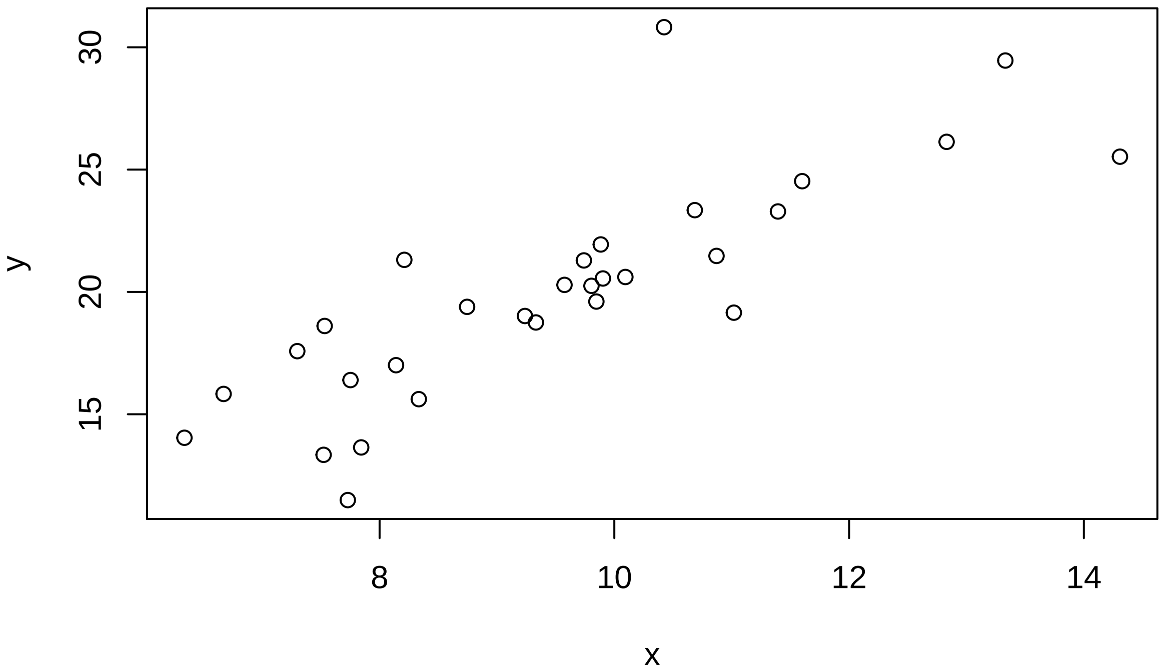
Naloga 3 - standardni odklon

Naloga 4, 5 - linearna regresija

```
# generiraj podatke
y= 2*(x=rnorm(30,mean=10,sd=2)) + rnorm(30,mean=0,sd=3)
lm(y~x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##       1.921        1.897
```

```
plot(x,y)
```




```
#funkcija za izracun cenilke
cenilka = function(x,y){
  return(sum((x-mean(x))*(y-mean(y)))/sum((x-mean(x))^2))
}
# preverimo, ali deluje prav
y= 2*(x=rnorm(30,mean=10,sd=2)) + rnorm(30,mean=0,sd=3)
cenilka(x,y)
```

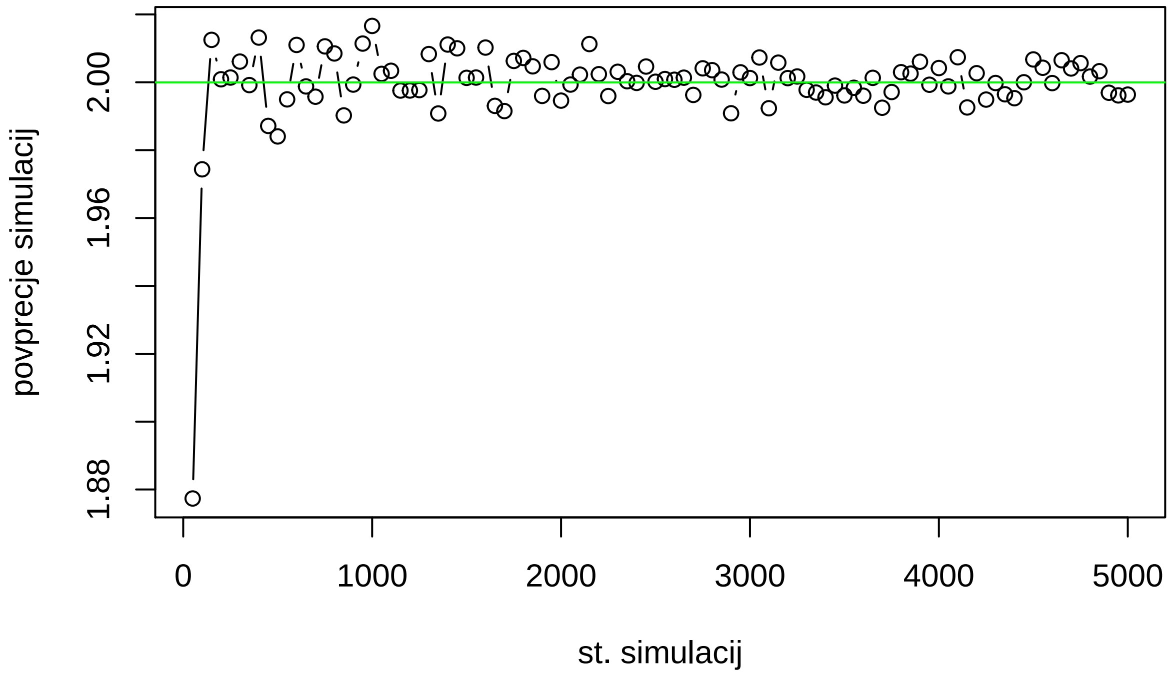
```
## [1] 1.97201
```

```
lm(y~x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     -0.4717       1.9720
```

```
# definiramo funkcijo za simulacijo N vzorcev in izracun N ocen
simN = function(N){
  ocene = NULL # inicializiramo vektor simuliranih ocen
  for(i in 1:N){
    vzorecx = rnorm(30,mean=10,sd=2) # generiramo vzorec x-ov
    vzorecy = 2*vzorecx + rnorm(30,mean=0,sd=3)
    ocene = c(ocene,cenilka(vzorecx,vzorecy))} # izracunamo cenilko
  return(ocene) #vrnemo vektor ocen
}
# nepristranskost
stVzorcev = seq(50,5000,by = 50)
povpr = NULL
for(i in stVzorcev){
  povpr = c(povpr,mean(simN(i)))}
```

```r
plot(stVzorcev,povpr,xlab="st. simulacij",
ylab="povprecje simulacij",main="",type="b")
abline(h=2,col="green") # populacijska vrednost
```

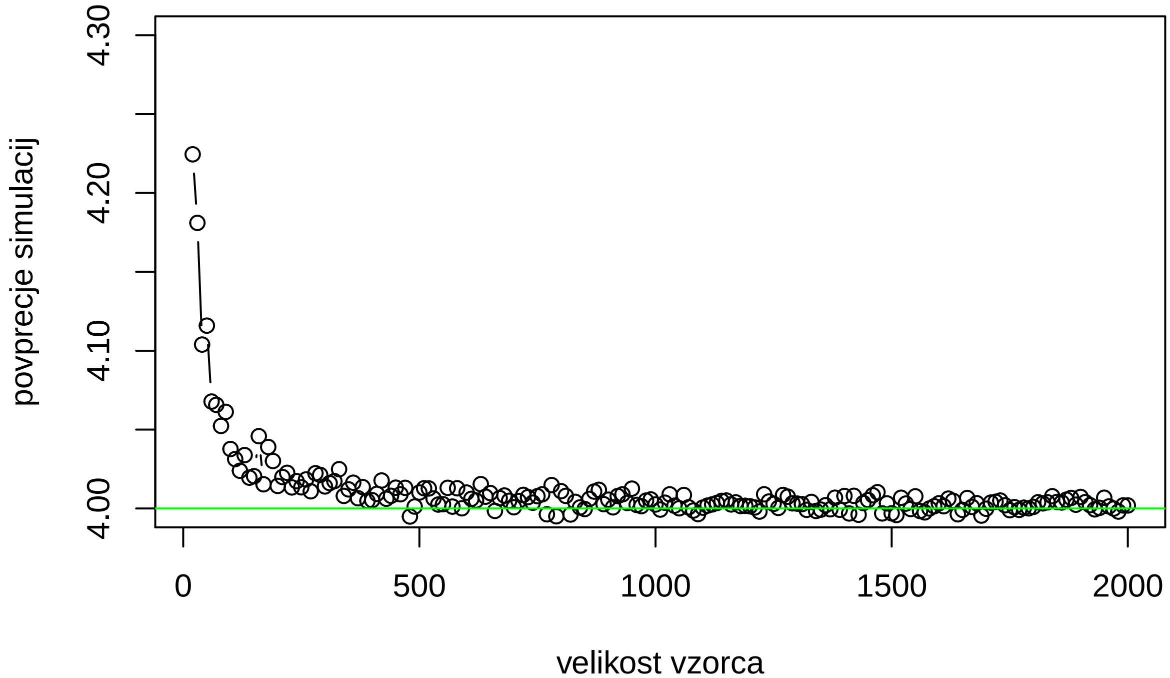


Naloga 6 - eksponentna porazdelitev

```r
#funkcija za izracun cenilke
cenilka = function(x){
  return(length(x)/sum(x))}

# definiramo funkcijo za simulacijo
simN = function(n,N){ # n - velikost vzorca se spreminja
  ocene = NULL
  for(i in 1:N){
    vzorec = rexp(n,rate=4)
    ocene = c(ocene,cenilka(vzorec))}
  return(ocene)
}

# doslednost
velikostVzorca = seq(20,2000,by = 10)
povpr = NULL
for(i in velikostVzorca){
  povpr = c(povpr,mean(simN(i,1000)))}
plot(velikostVzorca,povpr,xlab="velikost vzorca",
ylab="povprecje simulacij",main="",type="b",ylim=c(4,4.3))
abline(h=4,col="green") # populacijska vrednost
```

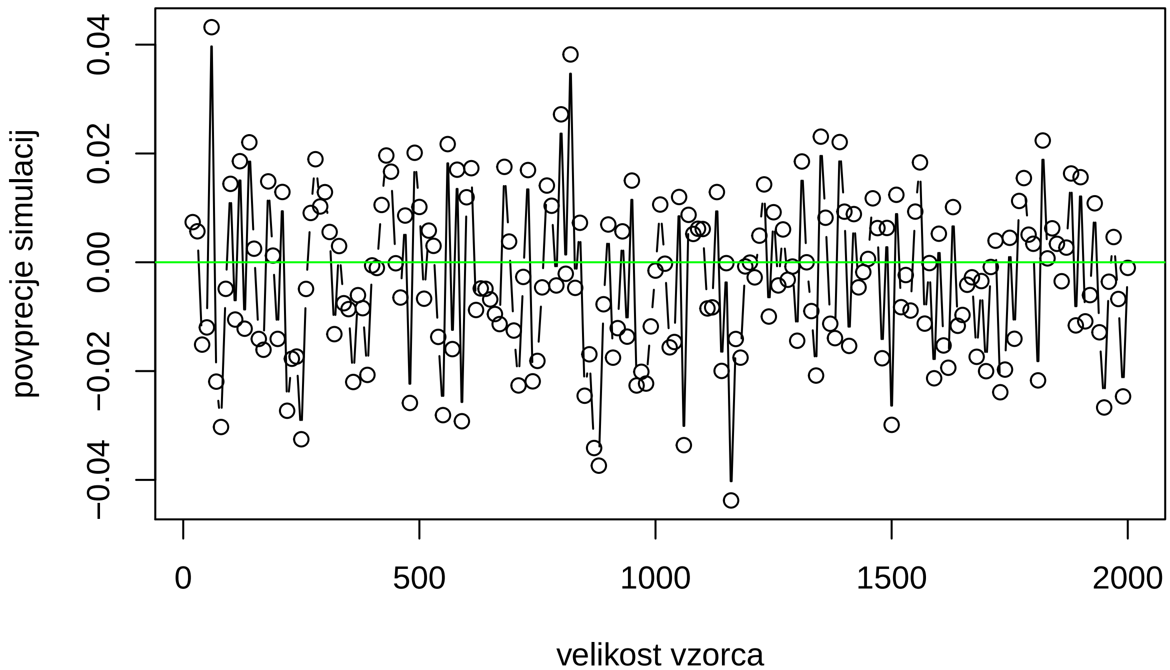


Naloga 7 - prvih 5 enot

```
cenilka = function(x){
  return(mean(x[1:5]))}

# definiramo funkcijo za simulacijo
simN = function(n,N){ # n - velikost vzorca se spreminja
  ocene = NULL
  for(i in 1:N){
    vzorec = rnorm(n) # izberemo npr. N(0,1)
    ocene = c(ocene,cenilka(vzorec))}
  return(ocene)
}

# doslednost
velikostVzorca = seq(20,2000,by = 10)
povpr = NULL
for(i in velikostVzorca){
  povpr = c(povpr,mean(simN(i,1000)))}
plot(velikostVzorca,povpr,xlab="velikost vzorca",
ylab="povprecje simulacij",main="",type="b")
abline(h=0,col="green") # populacijska vrednost
```




```
# variabilnost se ne zmanjsuje, ne konvergira
```

Naloga 8 - velik vzorec

```
#(a)
varianca = function(x){1/length(x) * sum((x-mean(x))^2)}
vzorec = rnorm(10000,mean=120,sd=30)
varianca(vzorec) # kolicina a
```

```
## [1] 899.311
```

```
#(b)
#dolocimo indekse za podovzorce
mejaZg = seq(10,10000,10)
mejaSp = seq(1,9991,10)
variance = NULL
for(i in 1:1000){
  mali = vzorec[mejaSp[i]:mejaZg[i]]
  variance = c(variance,varianca(mali))}
mean(variance) # kolicina b
```

```
## [1] 808.564
```

```
### drug način - ista naloga
vzorec = rnorm(10000,mean=120,sd=30)
varianca(vzorec)
```

```
## [1] 904.0728
```

```
vars = sapply(seq(1,10000,10),FUN=function(x)varianca(vzorec[x:(x+9)]))
mean(vars)
```

```
## [1] 813.4572
```

Naloga 9 - enakomerna porazdelitev

Gostota porazdelitve za X:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{a} \leq \text{x} \leq \text{b} \\ 0 & \text{sicer} \end{cases}$$

$$\begin{aligned} E(X) &= \int_a^b x f(x) dx = \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \frac{x^2}{2} |_a^b \\ &= \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{b^2}{2} \right) = \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

$$\begin{aligned} var(X) &= \int_a^b (x-\mu)^2 f(x) dx = \frac{1}{b-a} \left(\int_a^b x^2 dx + 2 \int_a^b \mu x dx + \mu^2 \int_a^b dx \right) \\ &= \frac{1}{b-a} \left(\frac{x^3}{3} |_a^b + (b+a) \frac{x^2}{2} |_a^b + \frac{(b+a)^2}{4} x |_a^b \right) \\ &= \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} - \frac{(b+a)(b^2-a^2)}{2} \right) = \frac{(b+a)^2 (b-a)}{4} \\ &= \frac{4(b-a)}{(b-a)} \frac{(b^2 + ab + a^2) - 6(b+a)^2 + 3(b+a)^2)}{12} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$