cvic10 – bodové odhady

Bodový odhad pro $U(0, \vartheta)$

```
theta = 4.7   # parametr, který ve skutečné aplikaci NEZNÁME
n = 10        # počet měření, tzv. rozsah náhodného výběru

t = runif(n,0,theta)  # realizace náhodného výběru ... tj. konkrétní n-tice čísel
t
```

```
##  [1] 4.4371492 1.3930071 0.5196741 3.6382322 3.1086725 0.9038945 0.3216326
##  [8] 4.5524948 3.3286946 1.5192835
```

Jeden možný bodový odhad: dvojnásobek průměru z čísel, co jsme dostali. A jeho kvadratická chyba.

```
theta2 = 2*mean(t)
theta2
```

```
## [1] 4.744547
```

```
(theta2-theta)^2
```

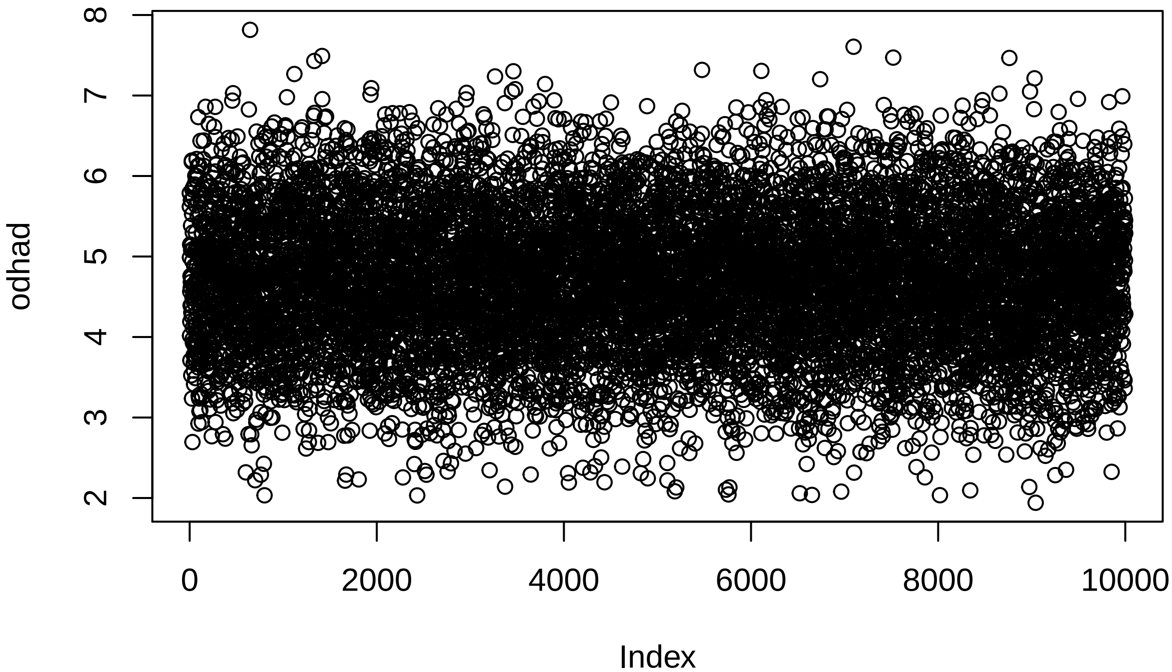
```
## [1] 0.001984439
```

Co znamenají vlastnosti bodového odhadu? A jak je ověřit? Např. nestrannost: jde nám o střední hodnotu výrazu theta2 = 2*mean(t), chceme, aby se rovnala theta. To můžeme spočíst "teoreticky na papíře" (a to přesně), tady zkusme samplovat. Budeme tedy opakovat celý pokus znovu a znovu a dělat průměr výrazu 2*mean(x). Pro přehlednost for-cyklem, abychom neodváděli pozornost k elegantním R-kovým konstrukcím.

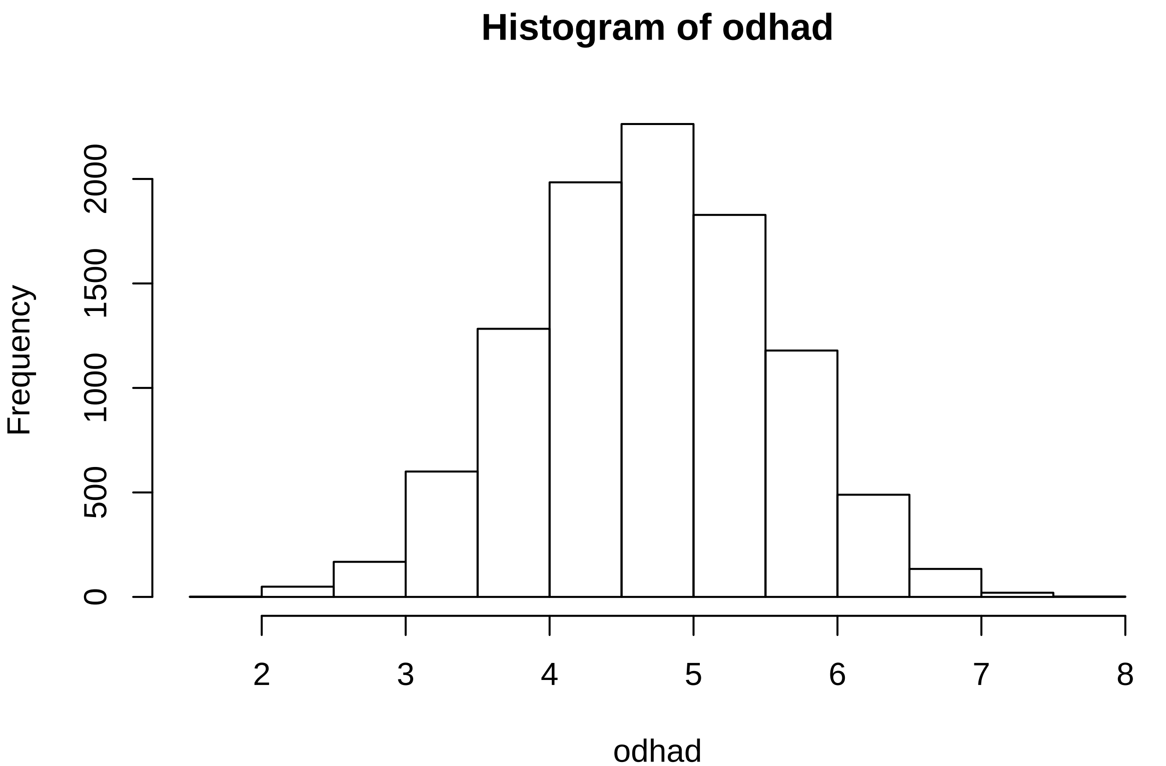
```
N = 10^4
odhad = rep(0,N)

for (cnt in 1:N) {
  odhad[cnt] = 2*mean(runif(n,0,theta))
}

plot(odhad)
```




```r
hist(odhad)
```




```r
mean(odhad)
```

```
## [1] 4.702214
```

```r
mean((odhad-theta)^2)
```

```
## [1] 0.7322249
```

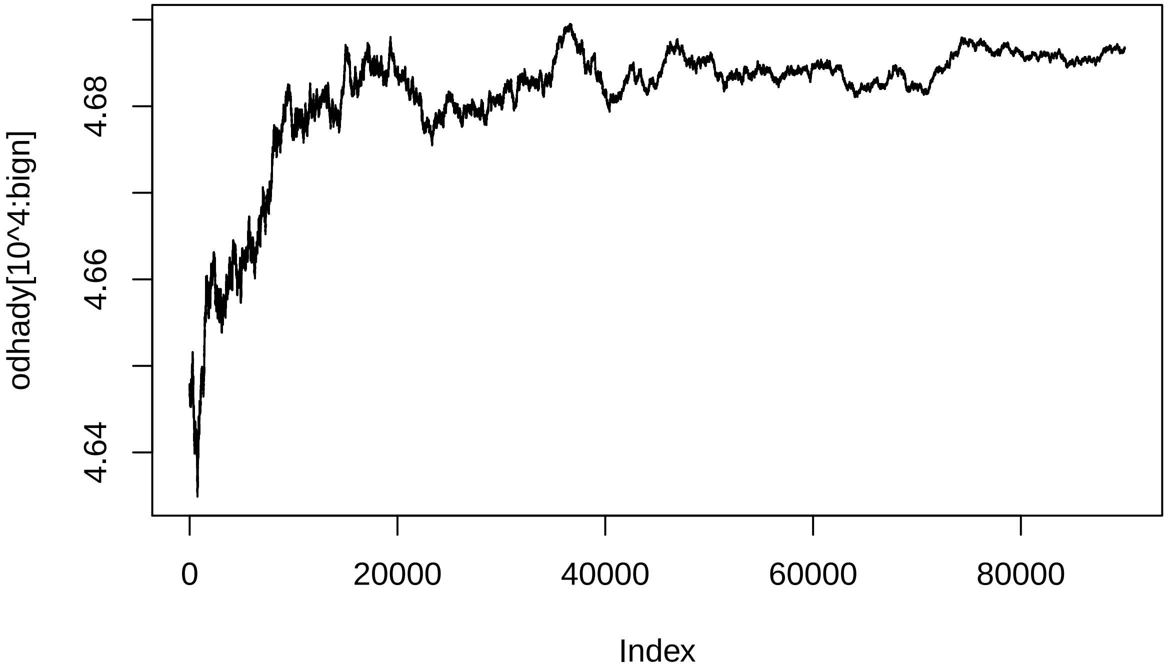
Konzistence znamená, že pro velká n dostaneme "skoro vždy" správnou hodnotu.

```
bign = 10^5
bigx = runif(bign,0,theta)

#hloupý způsob jak získat průměry ze všech prefixů: v kvadratickém čase
#odhady = 0*bigx
#for (n in 1:bign) {
#   odhady[n] = 2*mean(bigx[1:n])
#}

#chytřejší způsob v lineárním čase:
odhady = 2*cumsum(bigx)/seq(1:bign)

plot(odhady[10^4:bign], pch=2, type='l', cex=1)
lines(c(0,bign),c(4.7,4.7),col='red')
```



Teď totéž pro odhad získaný metodou max. věrohodnosti.

```
theta = 4.7   # parametr, který ve skutečné aplikaci NEZNÁME
n = 10        # počet měření, tzv. rozsah náhodného výběru

t = runif(n,0,theta)  # realizace náhodného výběru ... tj. konkrétní n-tice čísel
t
```

```
##  [1] 2.725071 3.328184 4.119362 1.040495 3.766770 1.886594 2.743624 3.852408
##  [9] 1.537381 1.920892
```

```
max(t)  #*(n+1)/n
```

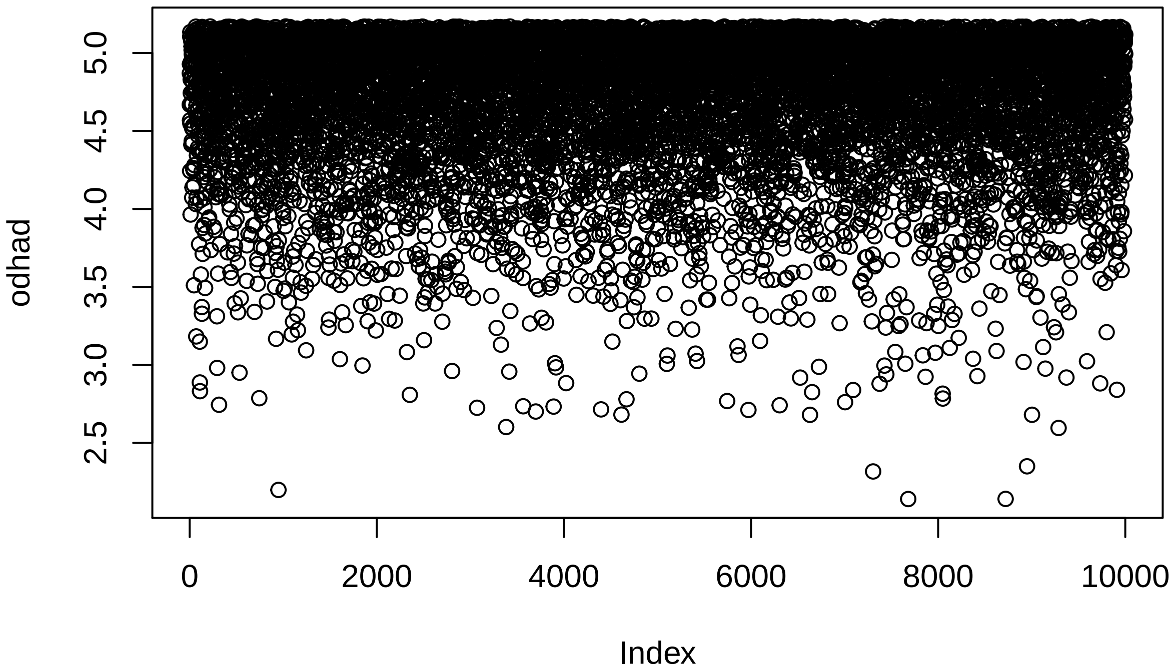
```
## [1] 4.119362
```

Nestrannost a MSE.

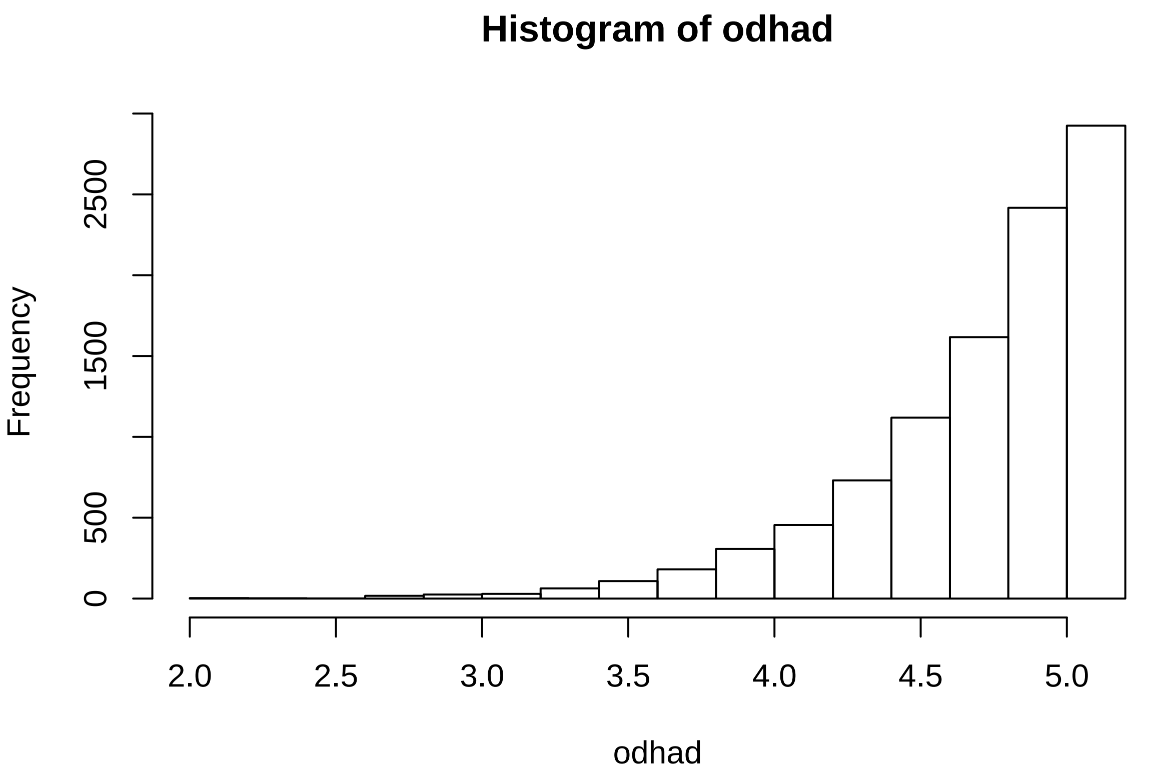
```
N = 10^4
odhad = rep(0,N)
```

```
for (cnt in 1:N) {
  odhad[cnt] = max(runif(n,0,theta))*(n+1)/n
}

plot(odhad)
```




```
hist(odhad)
```



```r
mean(odhad)
```

```
## [1] 4.710095
```

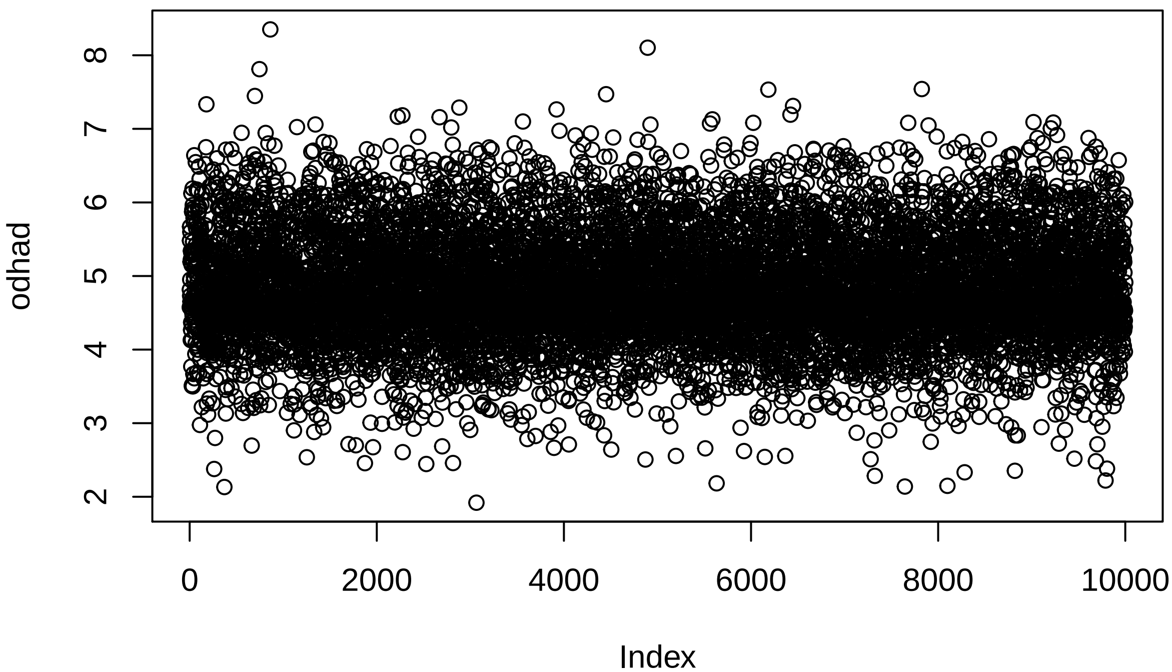
```r
mean((odhad-theta)^2)
```

```
## [1] 0.1797044
```

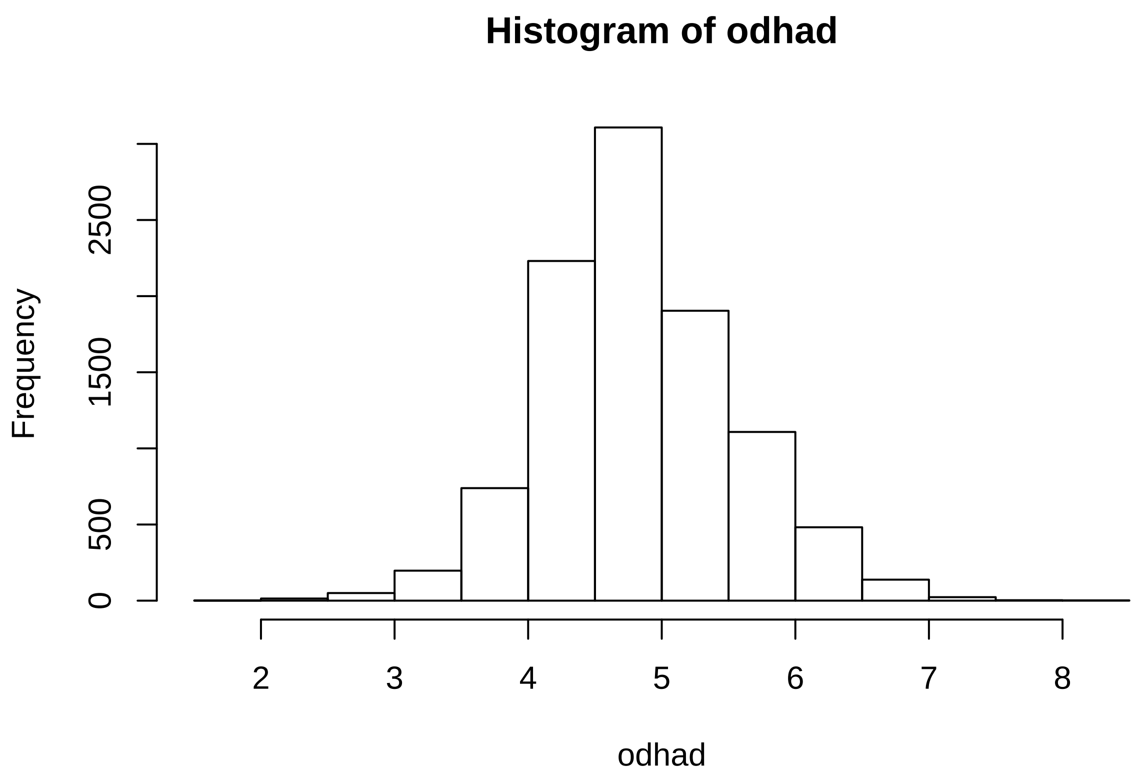
```r
N = 10^4
odhad = rep(0,N)

for (cnt in 1:N) {
  t = runif(n,0,theta)
  a = max(t) #*(n+1)/n
  b = 2*mean(t)
  odhad[cnt] = max(a,b)
}

plot(odhad)
```




```r
hist(odhad)
```




```
mean(odhad)
```

```
## [1] 4.828508
```

```
mean((odhad-theta)^2)
```

```
## [1] 0.5318131
```