

Výběrový rozptyl dvě verze

Podívejme se na dva vzorce pro výpočet výběrového rozptylu. To, co R počítá je vzorec daný součtem kvadratických odchylek od výběrového průměru – vydělené $n-1$. Tak se to aspoň píše v dokumentaci – ověříme:

```
data = rnorm(10)
var(data) # zabudovaná verze
```

```
## [1] 1.351882
```

```
m = mean(data)
sum((data-m)^2)/9 # vzorec explicitně napsaný
```

```
## [1] 1.351882
```

Ten druhý odhad samozřejmě dostaneme jako $\text{var}(\text{data}) * 9/10$. Teď ověříme, jaké mají odhady vlastnosti.

První vzorec

Napřed vyzkoušíme výběrový rozptyl “kde se dělí $n-1$ ”. Vidíme, že $\text{mean}(\text{bias})$ je skoro nula, takže se zdá, že se opravdu jedná o nestranný odhad. Dále ověříme, že MSE je rovno $\text{bias}^2 + \text{var}$.

```
estimates = replicate(10^6, var(rnorm(10)))
```

```
bias = mean(estimates) - 1
estvar = var(estimates)
mean(bias)
```

```
## [1] 0.0006247646
```

```
bias^2 + estvar
```

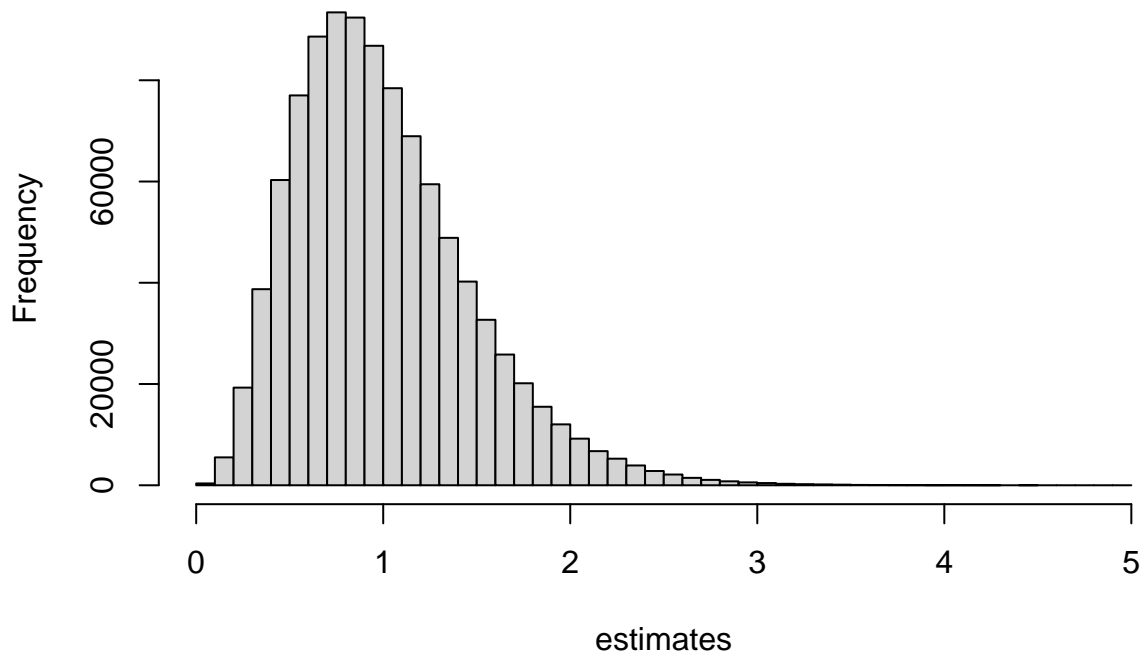
```
## [1] 0.2225657
```

```
mean((estimates-1)^2)
```

```
## [1] 0.2225654
```

```
hist(estimates, breaks = seq(0,5,.1))
```

Histogram of estimates



vzorec

Teď zkusíme totéž s druhým vzorcem pro výběrový rozptyl – tím, kde se dělí n . Opět ověříme, že $MSE = bias^2 + var$. A rovněž si všimneme, že zatímco teď má bias nenulovou střední hodnotu, MSE vyšlo menší – to proto, že rozptyl se snížil.

```
estimates2 = replicate(106, 9/10*var(rnorm(10)))
```

```
bias2 = mean(estimates2) - 1  
estvar2 = var(estimates2)  
mean(bias2)
```

```
## [1] -0.1000283
```

```
bias22 + estvar2
```

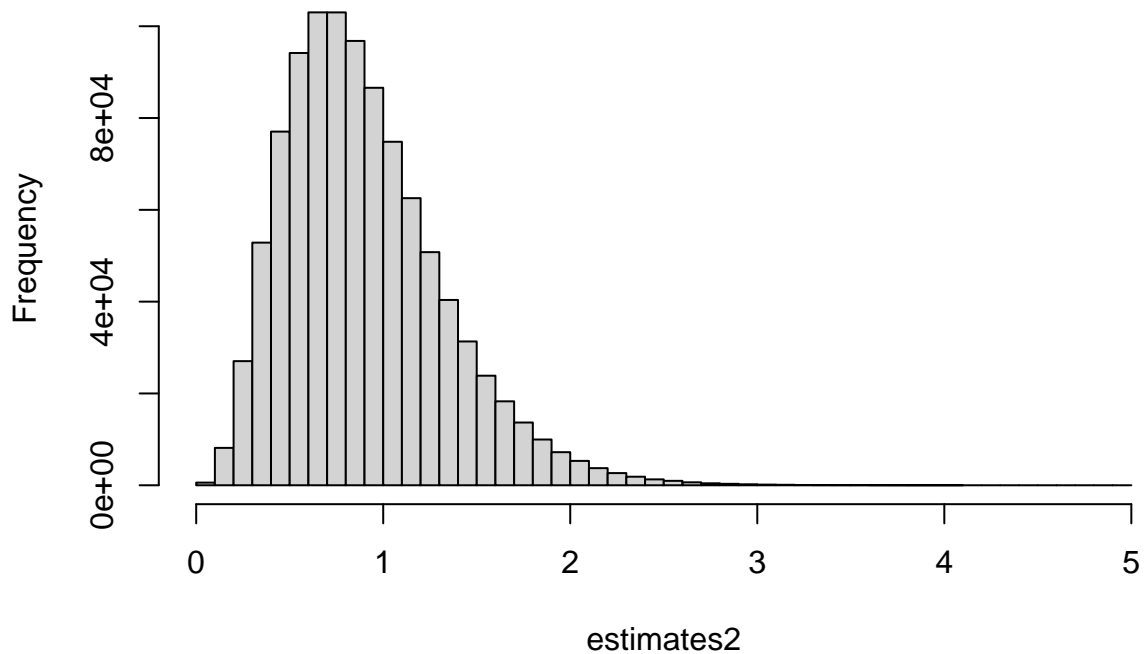
```
## [1] 0.1899712
```

```
mean((estimates2-1)2)
```

```
## [1] 0.1899711
```

```
# hist(estimates2)  
hist(estimates2, breaks = seq(0,5,.1))
```

Histogram of estimates2



pro exponenciální rozdělení

```
ex_estimates1 = replicate(10^4, var(rexp(10)))  
ex_estimates2 = replicate(10^4, 9/10*var(rexp(10)))  
mean(ex_estimates1)
```

```
## [1] 1.010793
```

```
mean(ex_estimates2)
```

```
## [1] 0.9042599
```

```
mean((ex_estimates1-1)^2)
```

```
## [1] 0.8407787
```

```
mean((ex_estimates2-1)^2)
```

```
## [1] 0.6889338
```

```
(var(rexp(10^5))-1)^2
```

```
## [1] 0.0002385209
```

Totěž pro uniformní rozdělení

```
un_estimates1 = replicate(10^5, var(runif(10)))  
un_estimates2 = replicate(10^5, 9/10*var(runif(10)))  
mean(un_estimates1)
```

```
## [1] 0.08328531
```

```
mean(un_estimates2)
```

```
## [1] 0.07487993
```

```
1/12
```

```
## [1] 0.08333333
```

```
mean((un_estimates1-1/12)^2)
```

```
## [1] 0.0007060001
```

```
mean((un_estimates2-1/12)^2)
```

```
## [1] 0.0006497206
```

```
(var(runif(10^6))-1/12)^2
```

```
## [1] 8.089293e-09
```