Properties of the exponential distribution

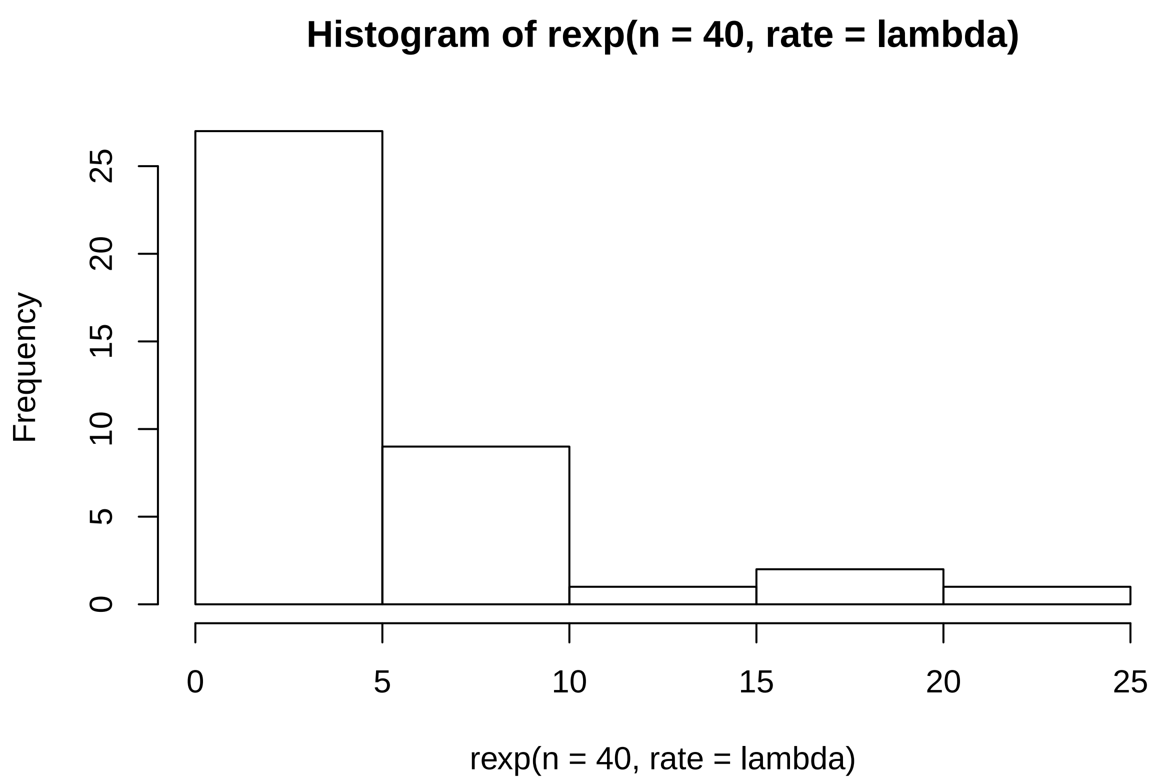
Overview

In this exercise 1000 exponential distributions with 40 values each are randomly generated. The rate parameter `lambda` is defined as 0.2. The mean value of each simulated distribution is calculated yielding a distribution of mean exponentials. Based on the Central Limit Theorem this distribution should be very similar to a normal distribution with parameters `mean = 1/lambda` and `sd = 1/(lambda*sqrt(40))^2`.

Simulations

Let's start with a randomly generated exponential distribution with rate parameter `lambda` equal to `0.2`.

```r
library("ggplot2")
library("tibble")

set.seed(1234)

lambda <- 0.2
hist(rexp(n = 40, rate = lambda))
```



Now we generate 1000 exponential distributions and calculate the mean of each to generate a distribution of mean exponentials.

```r
exps <- lapply(1:1000, function(x) rexp(n = 40, rate = lambda))
exp_mean <- sapply(exps, mean)
exp_df <- tibble(id = 1:1000, exp = exps, exp_mean = exp_mean)
```

Sample Mean versus Theoretical Mean

Let's compare the sample mean of the simulated exponential distributions to the theoretical mean which is defined as `1/lambda`.

The sample mean is very similar to the theoretical mean.

```r
mean_pop <- 1/lambda
mean_dis <- mean(exp_df$exp_mean)

abs(mean_pop - mean_dis)/mean_pop
```

```
## [1] 0.005271712
```

Sample Variance versus Theoretical Variance

Let's compare the sample variance of the simulated exponential distributions to the theoretical variance which is defined as `1/(lambda*sqrt(40))^2`.

The sample variance is very similar to the theoretical variance

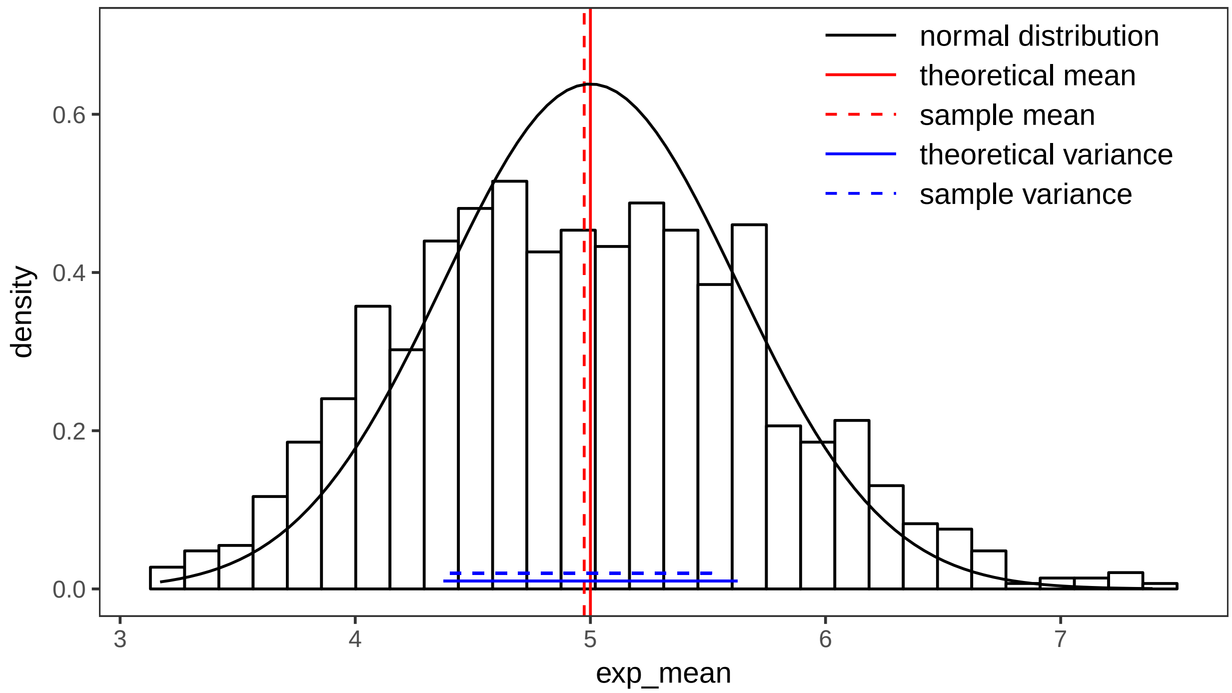
```r
var_pop <- 1/((lambda*sqrt(40))^2)
var_dis <- var(exp_df$exp_mean)

abs(var_pop - var_dis)/var_pop
```

```
## [1] 0.08637105
```

```r
ggplot(exp_df, aes(x = exp_mean)) +
  geom_histogram(aes(y = ..density..), fill = "white", col = "black") +
  geom_vline(xintercept = mean_pop, col = "red", lty = 1) +
  geom_vline(xintercept = mean_dis, col = "red", lty = 2) +
  geom_segment(x = mean_pop - var_pop, xend = mean_pop + var_pop, y = .01, yend = .01, col = "blue", lty
  geom_segment(x = mean_dis - var_dis, xend = mean_dis + var_dis, y = .02, yend = .02, col = "blue", lty
  stat_function(fun = dnorm, args = list(mean = 1/lambda, sd = ((lambda*sqrt(40))^-2))) +
  theme_bw() +
  theme(panel.grid = element_blank(), plot.caption = element_text(hjust = 0)) +
  annotate("segment", x = 6, xend = 6.3, y = 0.7, yend = 0.7, colour = "black") +
  annotate("segment", x = 6, xend = 6.3, y = 0.65, yend = 0.65, colour = "red") +
  annotate("segment", x = 6, xend = 6.3, y = 0.6, yend = 0.6, colour = "red", lty = 2) +
  annotate("segment", x = 6, xend = 6.3, y = 0.55, yend = 0.55, colour = "blue") +
  annotate("segment", x = 6, xend = 6.3, y = 0.5, yend = 0.5, colour = "blue", lty = 2) +
  annotate("text", x = 6.4, y = 0.7, label = "normal distribution", hjust = 0) +
  annotate("text", x = 6.4, y = 0.65, label = "theoretical mean", hjust = 0) +
  annotate("text", x = 6.4, y = 0.6, label = "sample mean", hjust = 0) +
  annotate("text", x = 6.4, y = 0.55, label = "theoretical variance", hjust = 0) +
  annotate("text", x = 6.4, y = 0.5, label = "sample variance", hjust = 0) +
  labs(caption = "Histogram of the distribution of mean exponentials. The black line indicates the true
distribution with parameters mean = 5 and sd = 0.625. The vertical red lines show the true
population mean (5, solid) and the mean of the simulated distributions (5.03, dashed). The
```

```
variance is indicates as horizontal lines arround the mean values for the true population
(0.625, solid) and the simulated distributions (0.622, dashed).")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Histogram of the distribution of mean exponentials. The black line indicates the true normal distribution with parameters mean = 5 and sd = 0.625. The vertical red lines show the true population mean (5, solid) and the mean of the simulated distributions (5.03, dashed). The variance is indicates as horizontal lines arround the mean values for the true population (0.625, solid) and the simulated distributions (0.622, dashed).