Investigating statistical inferences by simulation

LG

27 12 2021

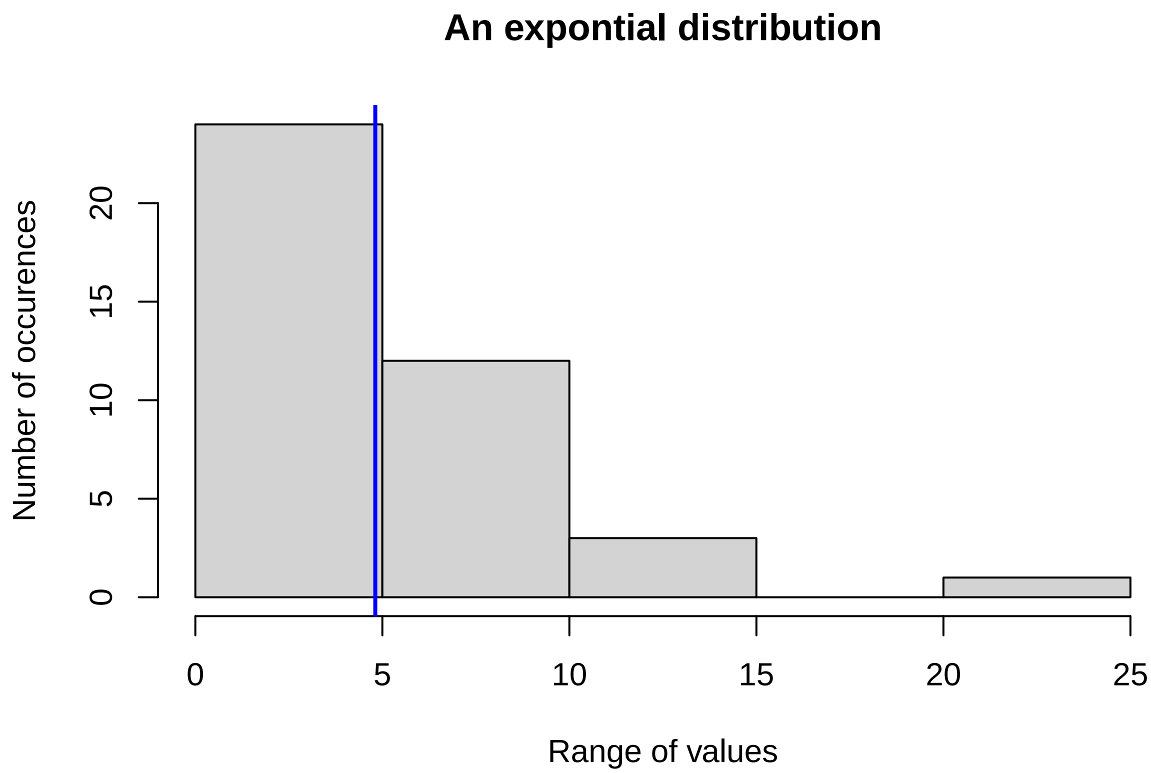
Overview:

Statistical inferences is a scientific field in which we can draw inferences from a sample to a whole population. One major theory is called the central limit theory. This says that for each sample the unknown population follows a normal distribution even if data is non-normal. To proof this, this document presents a simulation. 1000 sample means from exponential distributions are investigated in terms of mean, variability and distribution compared to a theoretical population. ## Simulation of exponential distribution

1000 experiments

First, let's have a look at a exponential distribution of 40 random variables at a rate lambda of 0.2. An example is shown in the histogram below. The blue line represents the mean of this distribution. Please note, that the distribution is highly skewed whereas normal distributions are bell-shaped.

```
set.seed(123)
#theoretical distribution
lambda <- 0.2
#creating exponential distribution with rate lambda
n <- 40
expdis <- rexp(n, lambda)
# showing in a graph
hist(expdis, ylab = "Number of occurences", xlab= "Range of values", main = "An expontial distribution")
abline(v=mean(expdis), lwd= 2, col="blue")
```



Imagine now, this simulation is performed a thousand times: 1000 samples revealing 1000 sample means. Ordering these 1000 sample means within a new distribution, according to the central limit theorem, this new distribution will behave normally. We store all 1000 sample means within one vector called `sample_means`.

```
# performing 1000 experiments
sample_means <- NULL
n_samples <- 1000
 for (i in 1 : n_samples) sample_means <- c(sample_means, mean(rexp(40, lambda)))
```

Sample mean vs theoretical mean

In theory, the mean as well as the standard deviation of an exponential distribution is calculated by 1/lambda.

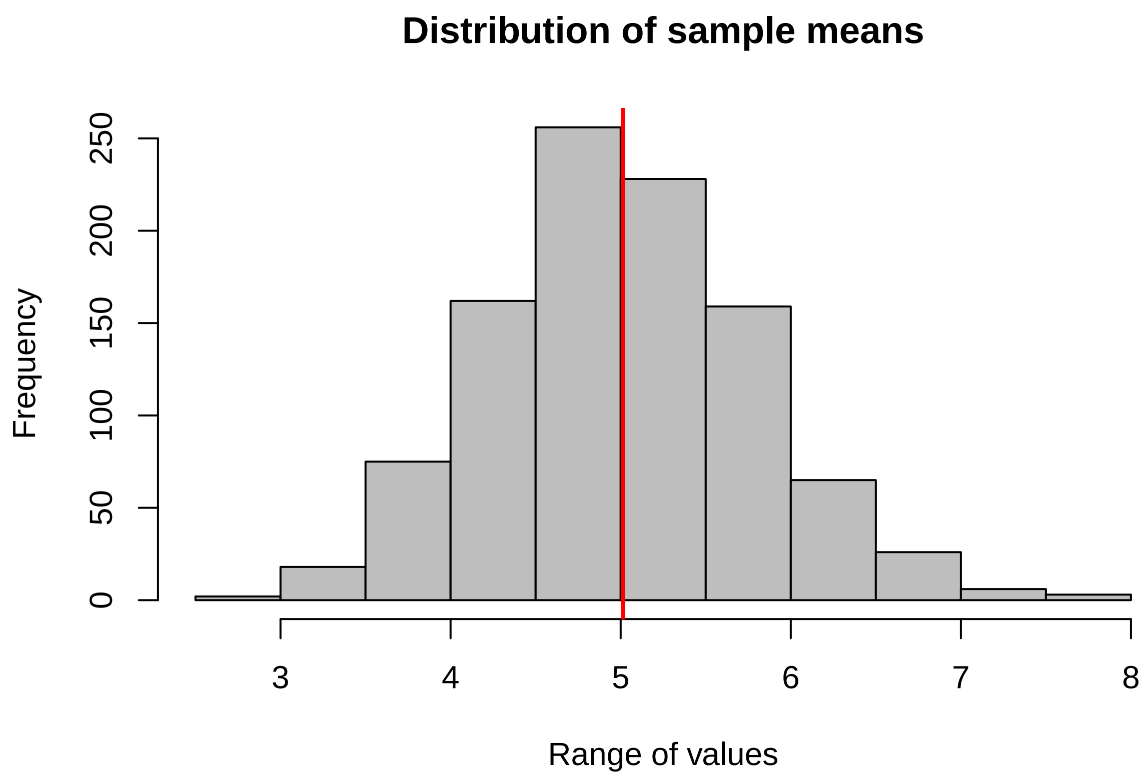
```
mu <- 1/lambda
sample_center <- mean(sample_means)
```

Thus a exponential distribution with rate of 0.2 results in a theoretical mean of 5. The mean of our sample means is 5.014. ### Sample variability vs theoretical variability Since the means are pretty close, let's look at variability.

```
theoretical_sd <- mu / sqrt(n)
sample_sd <- sd(sample_means)
theoretical_var <- (1/lambda^2)/n
sample_var <- var(sample_means)
```

Theoretical standard deviation results in 0.790569415042095 and the standard deviation of our simulation is 0.776. Theoretical variance results in 0.625 and variance of our simulation is 0.602. ### Distribution Exploring the distribution of our sample means follows accordingly to the central limit theorem: we clearly see a bell-shaped distribution with the red line illustrating the mean.

```
hist(sample_means, col= "gray", main = "Distribution of sample means", ylab = "Frequency", xlab="Range o
abline(v=mean(sample_means), lwd= 2, col="red")
```



All in all, the central limit theorem is proven correct, even if the population distribution is exponential, the means of 1000 samples follow normal distr