Simulations with exponential distribution

Xingmin Aaron Zhang

In this project, we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

We randomly sample the exponential distribution (lambda = 0.2) for 40 numbers (n = 40) and repeated do this for 1000 simulations.

```
n = 40
lambda = 0.2
N = 1000
set.seed(31237)
expMatrix <- matrix(rexp(n * N, rate = lambda), nrow = N, byrow = TRUE )
```

We calculate the mean of these 1000 samples, and create a histogram to show the distribution.

```
expMean <- rowMeans(expMatrix)
{ hist(expMean, xlab = "sample mean (n = 40)", ylab = "count",
      main = "histogram of means of exponentials")
  abline(v = mean(expMean), col = "royalblue", lwd = 3)
  text(x = 5, y = 160, labels = paste("mean = ", round(mean(expMean), 2)), pos = 4,
     col = "red")}
```

For exponential distributions, both the mean and standard deviation is 1/lambda. For a sample of size n, the mean is still 1/lamba and the standard deviation of the mean will be 1/(lambda * sqrt(n)). If we compare the number with the theoretical values, we can see that they match quite well.

```
mu = 1/lambda
sd = 1/lambda/sqrt(n)
meanAndSd <- data.frame(theoretical = c(mu, sd), observed = c(mean(expMean), sd(expMean)))
rownames(meanAndSd) = c("mean", "sd")
meanAndSd
```

```
##      theoretical  observed
## mean   5.0000000 5.0148683
## sd     0.7905694 0.8166727
```

We can make inference about population mean from the simulations, the 95% confidence interval is be [3.41 ~ 6.62]. Clearly, the theoretical mean (u = 5) is within this interval.

We can also fit the above histogram with a normal distribution curve and found that they match very well (left). The right side shows the distribution of original values for the simulations. One can see that while the original values follow the exponential distribution, the mean of samples follow normal distribution.

```
par(mfrow = c(1, 2))
{ hist(expMean, xlab = "mean of samples (n = 40)", ylab = "fraction",
      main = "sample mean (n = 40)", freq = FALSE)
```

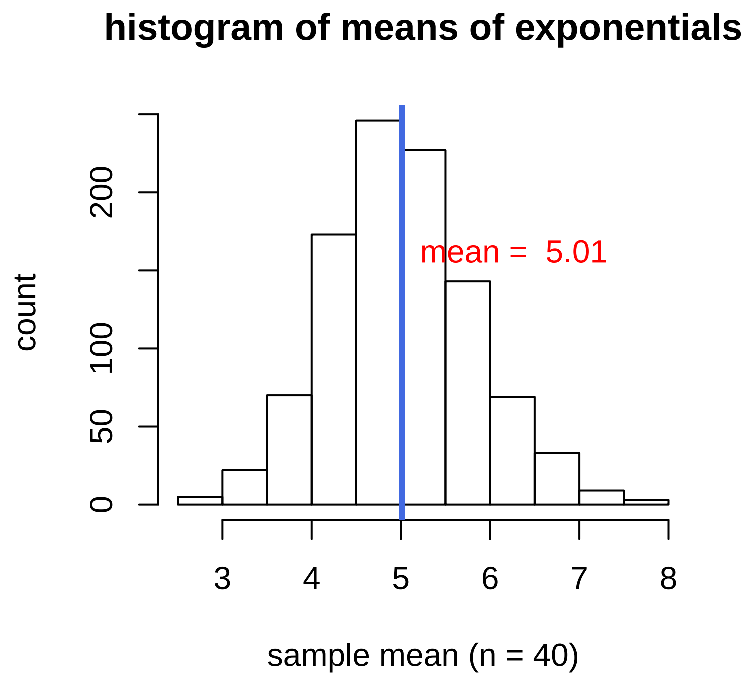

Figure 1: Figure 1 Distribution of means of 1000 simulations

```
  abline(v = mean(expMean), col = "royalblue", lwd = 3)
  text(x = 6, y = 0.3, labels = "theoretical\nnormal\ncurve", pos = 4, col = "red")
  x <- seq(2, 8, by = 0.2)
  curve(dnorm(x, mean = mean(expMean), sd = sd(expMean)),add = TRUE, col = "red", lwd = 2)
}

{
    rawNum <- as.vector(expMatrix)
    hist(rawNum, freq = FALSE, main = "distribution of raw values in simulation",
         xlab = "raw values of samples")
    x <- seq(min(rawNum), max(rawNum), by = 0.2)
    curve(dexp(x, rate = lambda), add = TRUE, col = "red", lwd = 2)
    text(x = 6, y = 0.08, labels = "theoretical exponential\ncurve", pos = 4, col = "red")
}
```

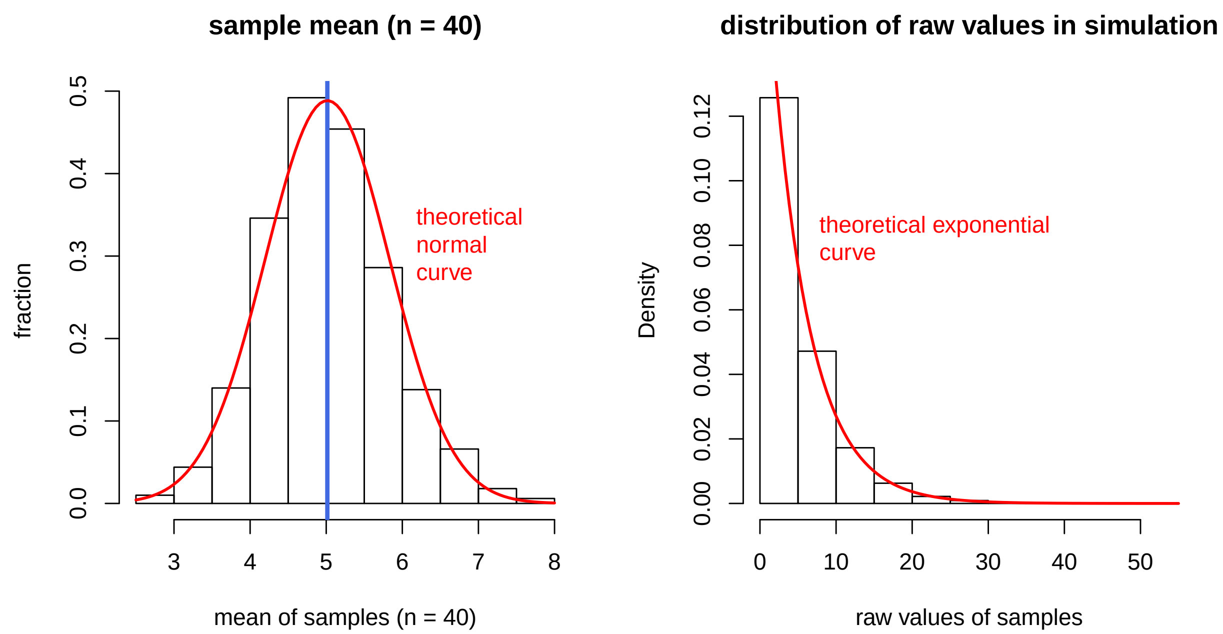

Figure 2: Figure 2 Distribution of sample means (left) and raw values of samples (right)