# *Exponential distribution, LLN, CLT*

*Bartek Skorulski*

*March 14th, 2014*

## *Introduction*

In this project you will investigate behavour of a sequence of random varaibles that follow the exponential distribution. We ilustrate Law of Large Numbers (LLN) nad Central Limit Theorem (CLT) via simpulations using R, a programming language especially design for statistical computing and graphics.

## *Exponential distribtion*

*Exponential distriubution* is given by the formula:

$$f(x) = \lambda e^{-\lambda x},$$

for $\lambda > 0$ where $x \geq 0$. Its mean is $\lambda^{-1}$, and its variance is $\lambda^{-2}$.

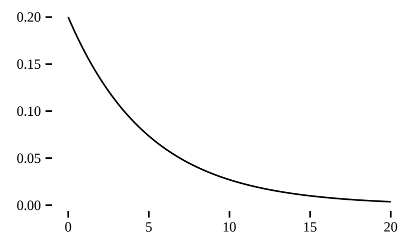


Figure 1: Exponental distribution with lambda equal to 0.2

## *Sample mean and variance*

Now we see how the sample mean and sample variance approximate the theoretical mean and variance. We use R function *rexp* that produces 40 samples. Then we calculate means and variances for first *k* samples for *k*=*1...40*.

```
N <- 40
set.seed(1110)
exps <- rexp(N, lambda)
means <- sapply(1:N, function(n) mean(exps[1:n]))
vars <- sapply(1:N, function(n) var(exps[1:n]))
```

Now we plot the results of our experiments.

```
g1 <- ggplot() + geom_line(aes(x = 1:N, y = means)) +
    geom_hline(yintercept = 1/lambda, color = "red") +
    labs(x = "") + theme_tufte()
g2 <- ggplot() + geom_line(aes(2:N, vars[2:N])) +
    geom_hline(yintercept = 1/lambda^2, color = "red") +
    labs(x = "", y = "variances") + theme_tufte()
grid.arrange(g1, g2, nrow = 2)
```

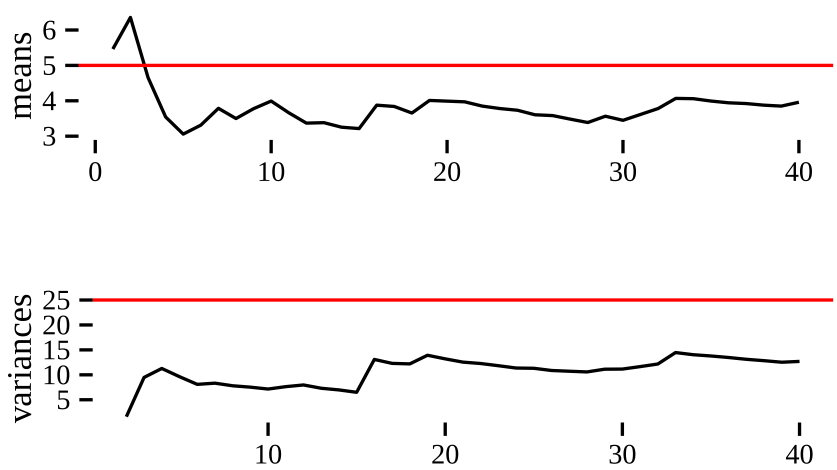


Figure 2: Sample means and variances converging to the theoretical ones reporesented by red lines.

## *Sum of independent random variable with exponential distribution*

Exponential distribution is a special case of *gamma distribution*

$$f_{\alpha,\nu}(x) = \frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}, \nu > 0, x > 0.$$

with $\alpha = \lambda$ and $\nu = 1$. Let us also recall that:

$$\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}dx.$$

It is also worth noting that if $X_1$ and $X_2$ are two independent random variable that have gamma distribution respectively $f_{\alpha,\mu}$, $f_{\alpha,\nu}$ with the same parameter $\alpha$, then the sum $X_1 + X_2$ has the distribution $f_{\alpha,\mu+\nu}$. Hence sum $k$ of identicaly distributed random variables with expotential distribribution is a gamma distribution $f_{\alpha,k}$. Moreover, if $X$ is a Gamma distribution with a parametrs $\lambda, \nu$, then $X/k$ is Gamma distribution with paramenter $k\lambda, \nu$. Therefore, the density of $(X_1 + \ldots + X_k)/k$ is $f_{k\lambda,k}$. So lets compare Gamma distribution $f_{8,40}$ that correspons to the distribution of the mean of 40 exponential samples with $\lambda = 0.2$ and the normal distribution that should approximate this mean.

```
x <- seq(2, 8, 0.01)
k <- 40
e40.m <- 1/lambda
e40.sd <- 1/(lambda * sqrt(40))
ggplot() + geom_path(aes(x, dgamma(x, shape = k,
    rate = k * lambda), colour = "blue")) + geom_path(aes(x = x,
    y = dnorm(x, e40.m, e40.sd), colour = "red")) +
    labs(x = "", y = "") + scale_colour_discrete(name = "Distribution",
    labels = c("Normal", "Gamma")) + theme_tufte()
```

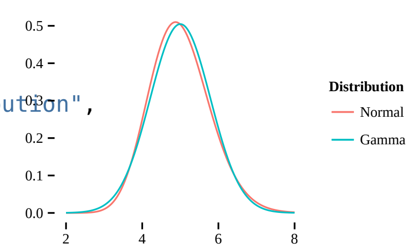


Figure 3: Gamma and normal distribution

## *Simulation*

```
N <- 1000
means <- c()
set.seed(10)
for (n in 1:N) {
    exps <- rexp(k, lambda)
    means <- c(means, mean(exps))
}
e40.dt <- data.table(means = means)
x <- seq(e40.m - 3.5 * e40.sd, e40.m + 3.5 * e40.sd,
    e40.sd/10)
ggplot() + geom_histogram(data = e40.dt, aes(x = means,
    y = ..density..), fill = "white", colour = "black") +
    geom_path(aes(x = x, y = dnorm(x, e40.m, e40.sd),
        colour = "red")) + geom_path(aes(x, k *
    dgamma(k * x, shape = k, scale = 5), colour = "blue")) +
    scale_colour_discrete(name = "Distribution",
        labels = c("Gamma", "Normal")) + theme_tufte() +
    theme(text = element_text(size = 5), line = element_line(size = 0.1),
        legend.background = element_rect(fill = "white"),
        legend.position = c(0.14, 0.8), panel.grid.major = element_line(colour = "grey40"),
        panel.grid.minor = element_blank())
```

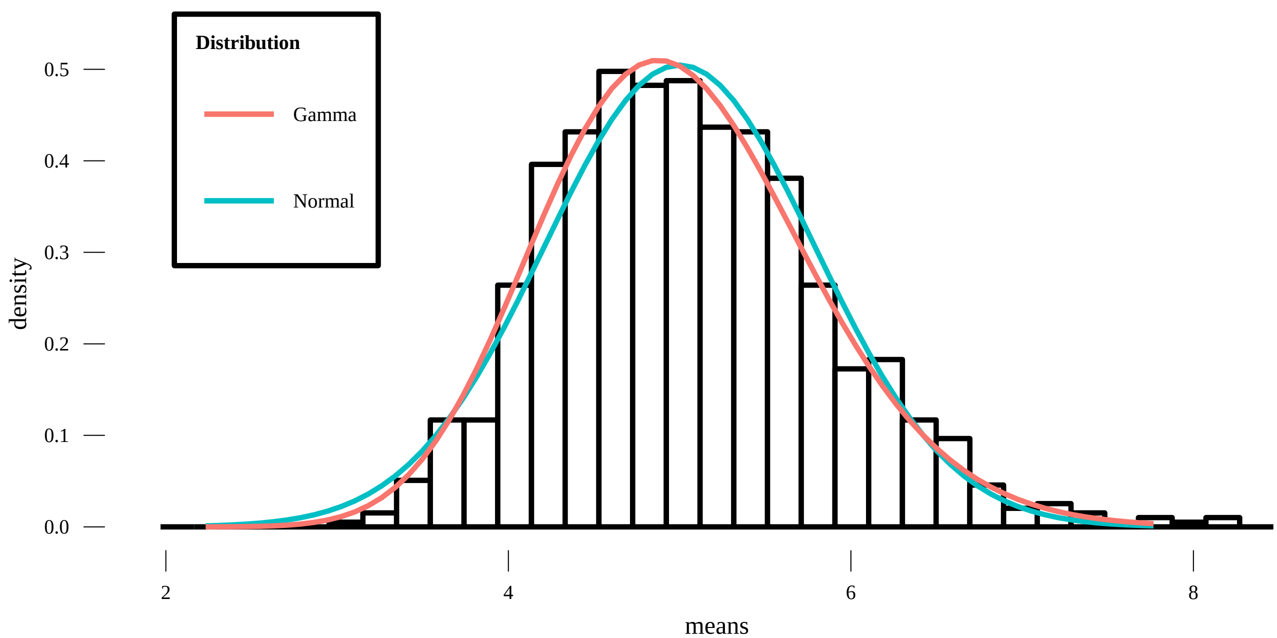


Figure 4: Histogram of 1000 sample means and corresponding normal and Gamma distribution.

```
# theme_tufte()+theme(=element_text(size=10))
```