Central Limit Theorem demo

Statistical Inference course project

Ilya Kochergin

22 nov 2015

```
library(ggplot2)
```

Overview

Let's demonstrate, that average computed on sample (sample mean) has distribution that tends to be asymptotically normal. Original random variable in this demo is exponentially distributed.

```
lambda=0.2 # rate parameter
n=40       # sample size
nosim=1000 # number of samples
```

Density of exponentially distributed random variable.

Theoretial $Std.deviation\,(\sigma) = Mean\,(\mu) = 1/\lambda{=}5$

Simulations

We generate `nosim=1000` rows of random numbers (samples). Sample size is `n=40`

```
# Random matrix. Each row is a sample (n=40)
rnd_exp_samples<-  matrix(data=rexp(nosim*n,rate=lambda),ncol = n, byrow = T)
```

Sample Mean versus Theoretical Mean

Theoretical mean $\mu = 1/\lambda$ Sample mean is a random variable. We have 1000 sample means

```
mu=1/lambda
# vector of sample means
rnd_exp_samples_means <-  apply(X = rnd_exp_samples,1 ,FUN = mean)
# range of sample means
summary(rnd_exp_samples_means)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.673   4.476   4.987   5.013   5.520   7.913
```

Sample means centered at 5.0127542 near theoretial 5

Sample Variance versus Theoretical Variance

Theoretical standard deviation is $\sigma = 1/\lambda = 5$ Theoretical variance is $\sigma = 25$ Sample variance is a random variable. We have 1000 sample variance

```
sigma=1/lambda
# vector of sample means
rnd_exp_samples_vars <-  apply(X = rnd_exp_samples,1 ,FUN = var)
# range of sample variances
summary(rnd_exp_samples_vars)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.162  17.610  23.030  25.280  30.980  79.400
```

Sample variance is centered at 25.2767121 near theoretial 25

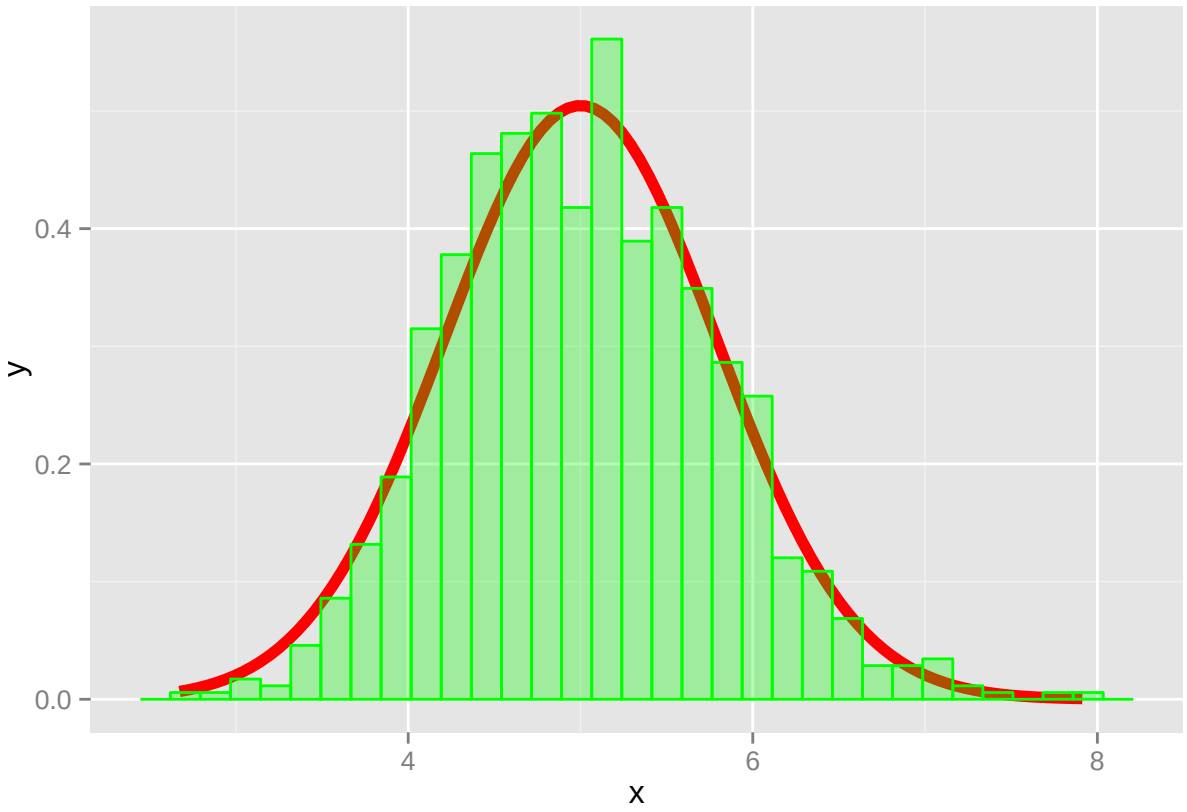
Distribution

Histogram of sample means approches to density of Normal distributon with the following parameters:

```
gaussian_mu = mu
gaussian_sigma = sigma/sqrt(n)
```

```
ggplot(data.frame(x=rnd_exp_samples_means),aes(x)) +
  stat_function(fun=dnorm,args=list(mean=gaussian_mu,sd=gaussian_sigma),col="red",size=2)+
  geom_histogram(aes(y=..density..),fill="green",alpha=0.3,col="green",stat_bin=1/30)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```


0.4
0.2
0.0
y
4
6
8
x