Estimation theory – Laboratory 1.

Agnieszka Szkutek, 208619
Marta Frankowska, 208581

October 19, 2017

Contents

1 Exercise 1

We generate a vector of $Y \sim N(\mu = 2, \sigma^2 = 4)$ with length $N = 1000$:

```
Y <- rnorm(n = 1000, mean = 2, sd = 2)
# transform Y to Y1
Y1 <- 3 * (Y - 1)
```

Y_1 has normal distribution, because it is a linear combination of Y. We can calculate analytical mean μ_1 and variance σ_1^2:

$$\mu_1 = E(Y_1) = E(3(Y-1)) = 3E(Y-1) = 3E(Y) - 3 = 3 \cdot 2 - 3 = 3$$

$$\sigma_1^2 = Var(Y_1) = Var(3(Y-1)) = 9Var(Y) = 9 \cdot 4 = 36$$

We can compare numerical and analytical results:

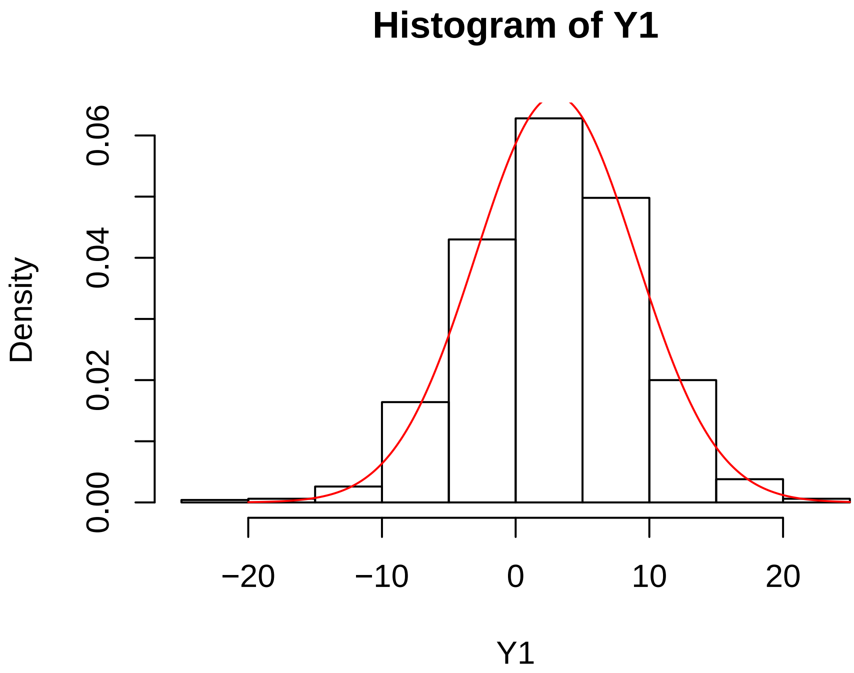
```
Y <- rnorm(n = 1000, mean = 2, sd = 2)
Y1 <- 3 * (Y - 1)
mean(Y1)

## [1] 3.047471

var(Y1)

## [1] 34.70353
```

Now we can plot the frequency historam of Y_1 and analytical normal density with $\mu = 3$ and $\sigma^2 = 36$.



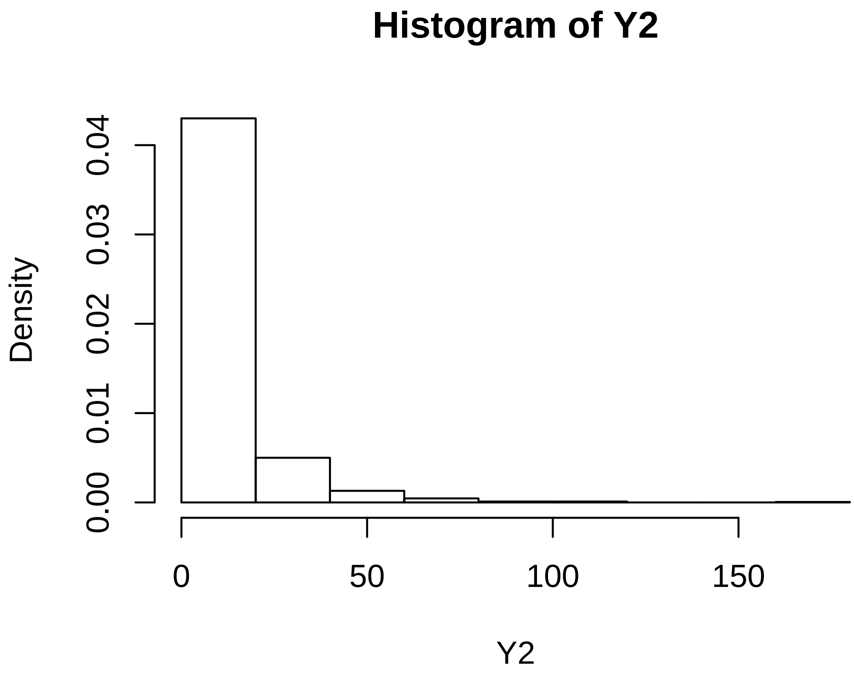
Next, we create variable Y_2

$$Y_2 = \left(\frac{Y_1 - 2}{2}\right)^2$$

and plot its frequency historam.

```
Y <- rnorm(n = 1000, mean = 2, sd = 2)
Y1 <- 3 * (Y - 1)
Y2 <- ((Y1 - 2) / 2) ^ 2

hist(Y2, freq = FALSE)
```



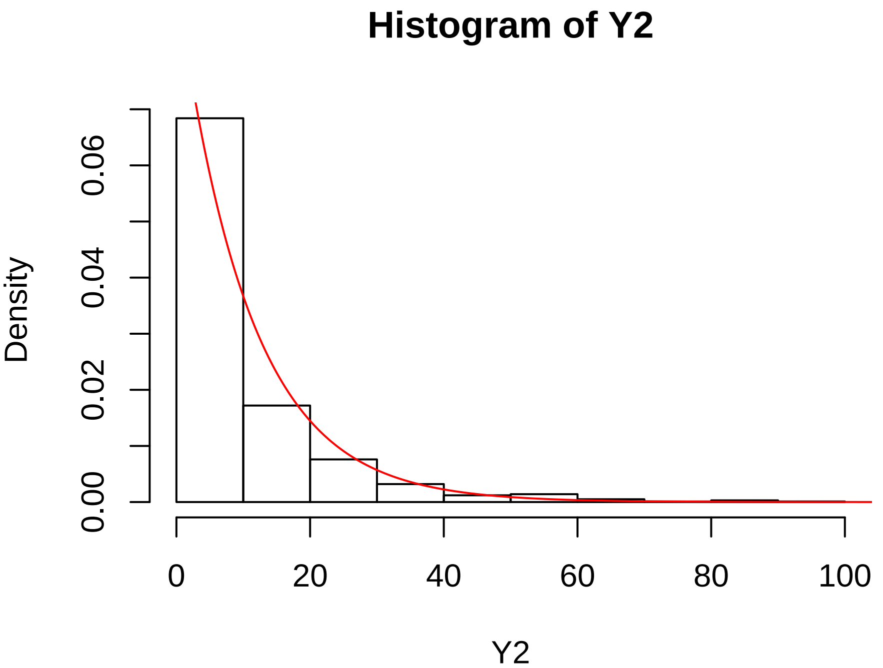
Y_2 is a quadratic function of Y_1, so we know that distribution of Y_2 is not normal. From the shape histogram we can assume that Y_2 has exponential distribution with

$$\lambda = \frac{1}{E(Y_2)}.$$

We can obtain the analytical value of $E(Y_2)$ in the following way

$$\begin{aligned} E(Y_2) =& E\left(\left(\frac{Y_1 - 2}{2}\right)^2\right) = \frac{1}{4}E(Y_1 - 2)^2 = \frac{1}{4}E(Y_1^2 - 2Y_1 + 4) = \\ =& \frac{1}{4}EY_1^2 - \frac{1}{2}EY_1 + 1 = \frac{1}{4}\left(VarY_1 + (EY_1)^2\right) - \frac{1}{2}EY_1 + 1 = \\ =& \frac{1}{4}(36 + 9) - \frac{3}{2} + 1 = 10.75, \end{aligned}$$

so $\lambda = \frac{1}{10.75} \approx 0.093$.

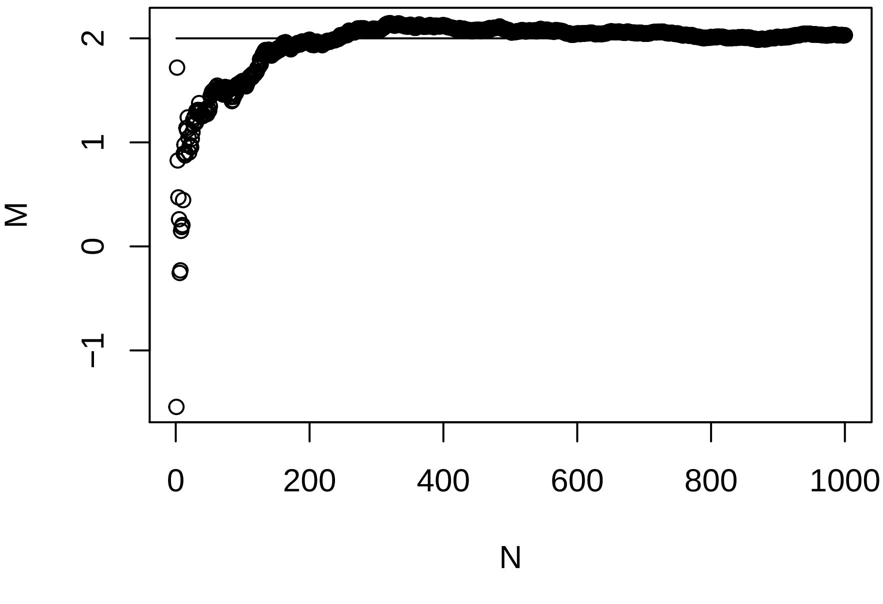


Next we will compute a sequence of means m_n and a sequence of variances σ_n^2 for the variable Y, where

$$m_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

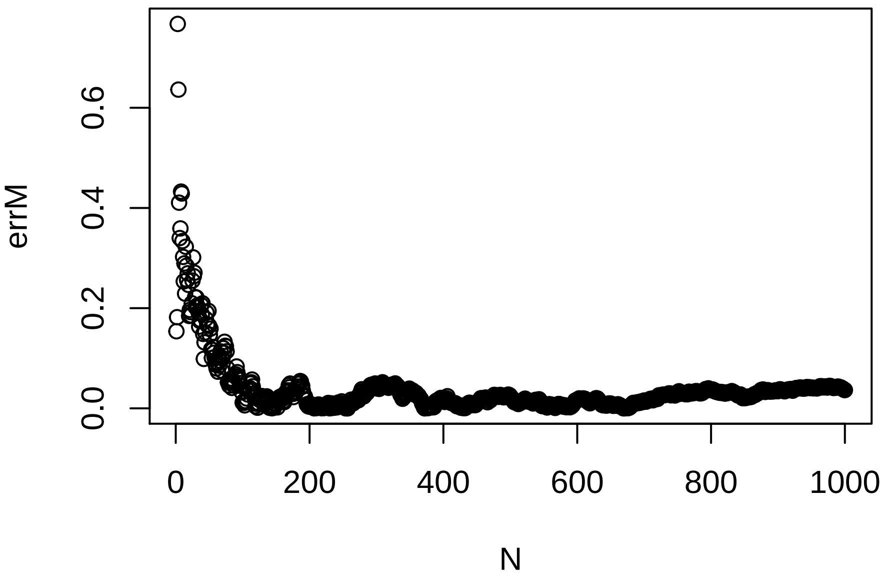
$$\sigma_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i - m_n)^2$$

and plot the results.



The sequences m_n and σ_n^2 converge to theoretical mean and variance, respectively. To examine the variability of the sequences we can calculate relative errors for both values.

$$err_{M_n} = \left| \frac{M_n - \mu_n}{\mu_n} \right|$$

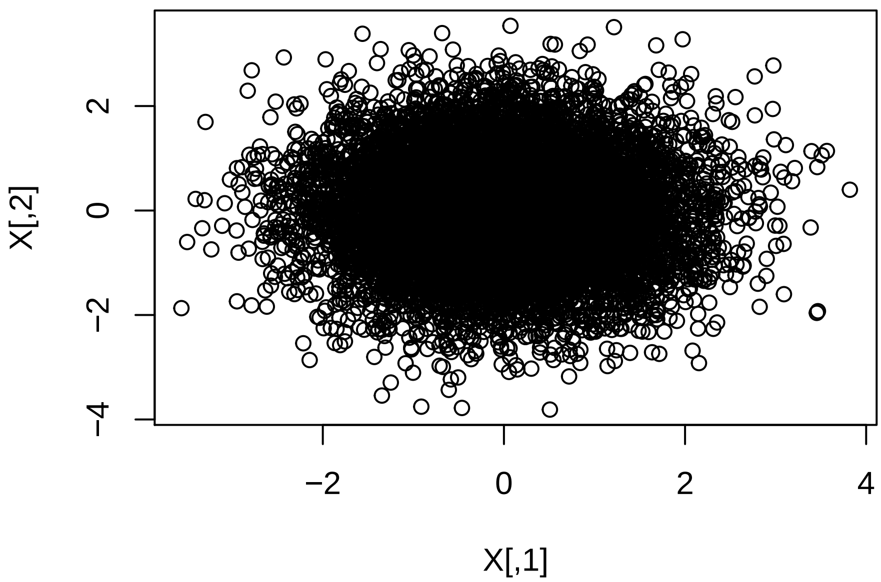


For $N > 200$ value of the error is less than 10% of theoretical mean.

2 Exercise 2

We simulate 10000 times and then plot 2-dimensional random variable $X \sim N(0, I_2)$.

```
n <- 10000
mu <- c(0, 0)
Sigma <- diag(2)
X <- rmvnorm(n = n, mean = mu, sigma =  Sigma)
plot(X)
```



We have to transform variable X into variable $Y \sim N(\mu, \Sigma)$, where

$$\mu = [0,\ 1] \quad \text{and} \quad \Sigma = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}$$

It means that we should find vector a and matrix A, such that $Y = AX + a$.

Expected value of Y is equal to

$$E(Y) = E(AX + a) = AE(X) + a = a,$$

and variance

$$Var(Y) = Var(AX + a) = Var(AX) = AVar(X)A' = AIA' = AA' = \Sigma = \Sigma^{0.5}\left(\Sigma^{0.5}\right)'.$$

From that we obtain $A = \Sigma^{0.5}$, so

$$Y = \Sigma^{0.5}X + \mu.$$

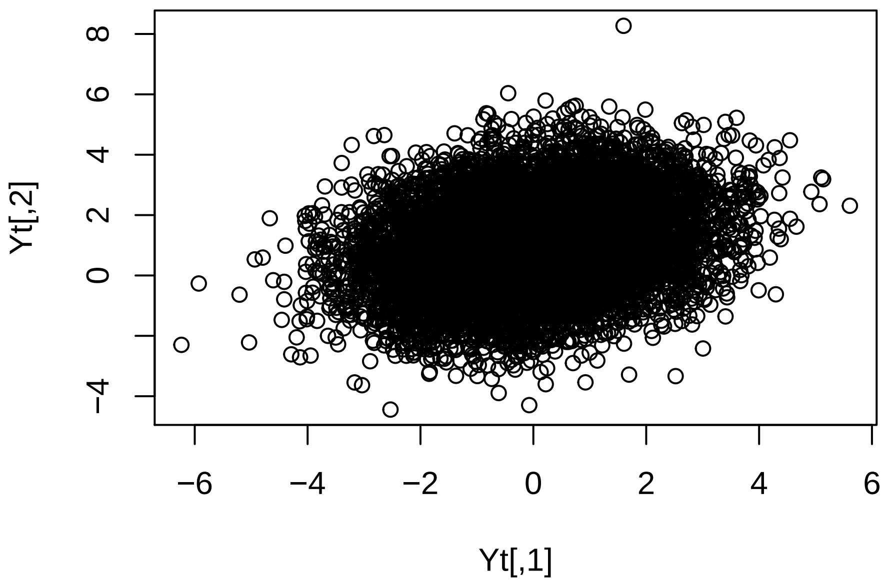
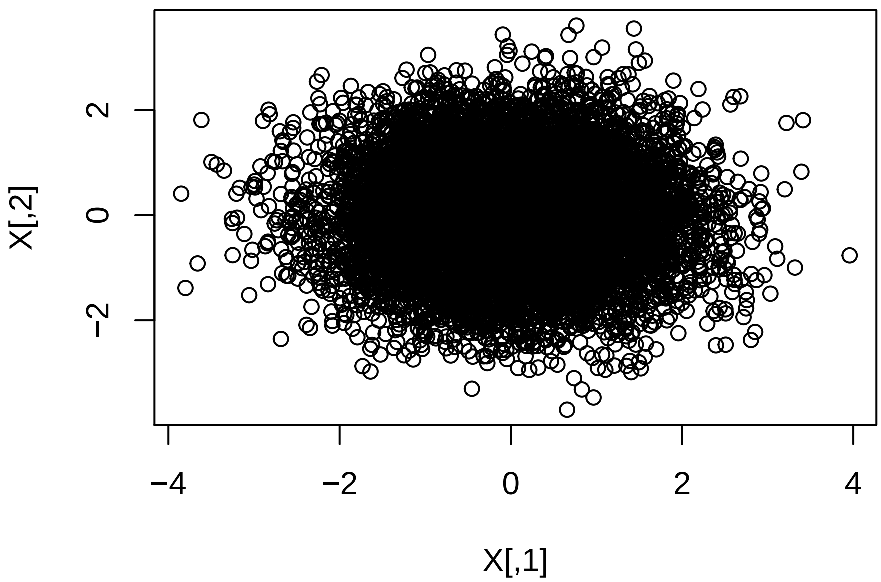


Figure 1: Random variable Y

```
## Error in eval(expr, envir, enclos):  could not find function "hist3D"
```

```
n <- 10000
mu <- c(0, 0)
Sigma <- diag(2)
X <- rmvnorm(n = n, mean = mu, sigma =  Sigma)
plot(X)
```



3 Exercise 3

When using additional source materials (books, links, etc.) do not forget to include appropriate references. For example, let us assume we want to cite Dalgard (2008). Then the bibliography section should contain:

References

[1] Peter Dalgaard, *Introductory Statistics with R*, Springer-Verlag New York, 2008.