

Estimation theory – Laboratory 1.

Agnieszka Szkutek, 208619

Marta Frankowska, 208581

October 18, 2017

Contents

1	Exercise 1	1
2	Exercise 2	6
3	Exercise 3	7
4	Exercise 4	7

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:

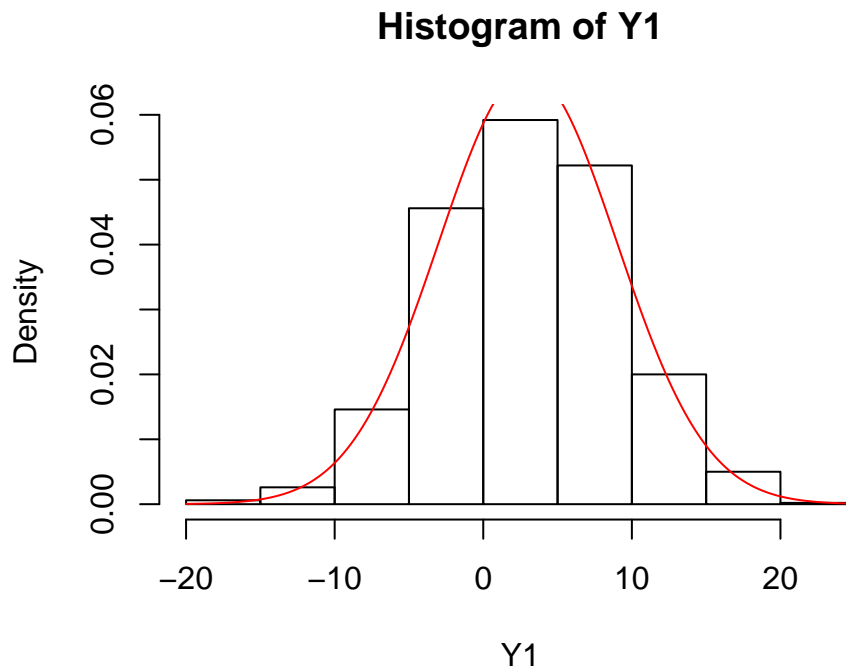
```
Y <- rnorm(n = 1000, mean = 2, sd = 2)
Y1 <- 3 * (Y - 1)
mean(Y1)

## [1] 3.361507

var(Y1)

## [1] 35.81307
```

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



Next, we create variable Y_2

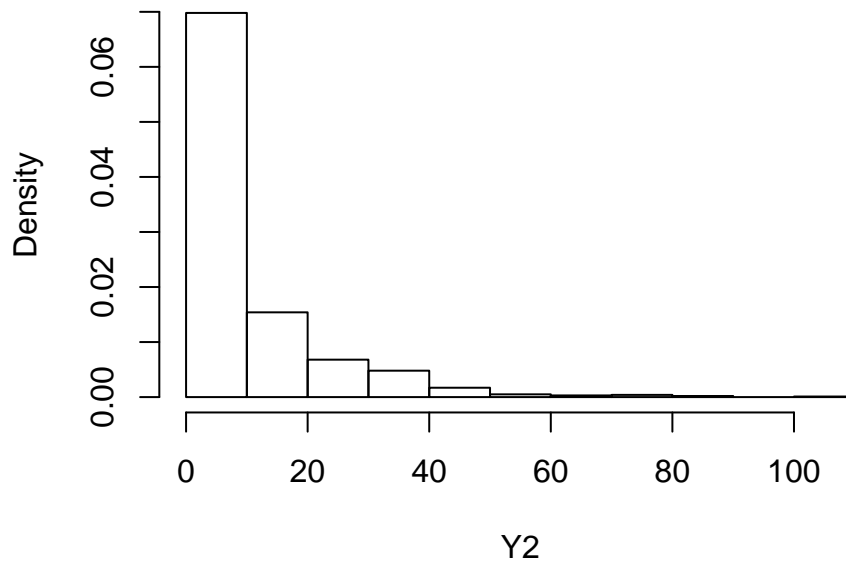
$$Y_2 = \left(\frac{Y_1 - 2}{2} \right)^2$$

and plot its frequency histogram.

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

hist(Y2, freq = FALSE)
```

Histogram of Y2



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

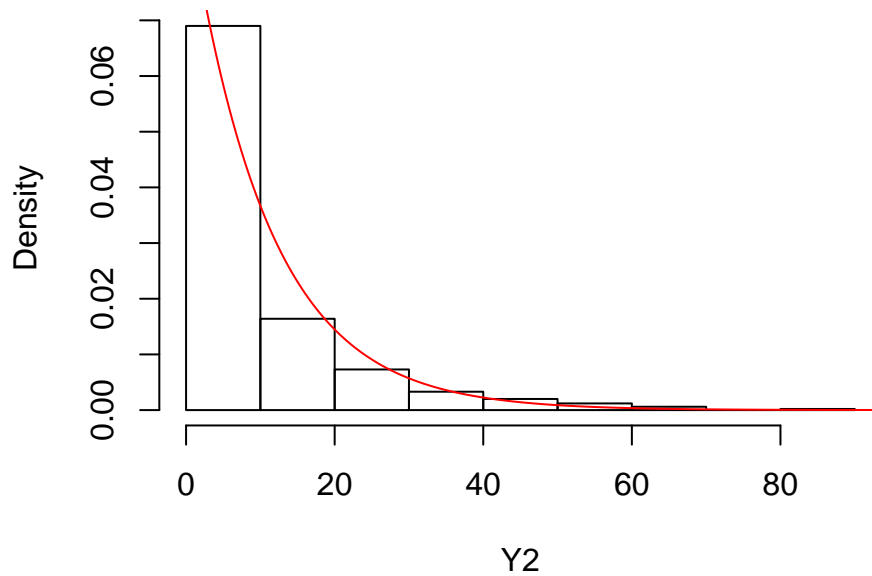
$$\lambda = \frac{1}{E(Y_2)}.$$

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

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

```
Y <- rnorm(n = 1000, mean = 2, sd = 2)
Y1 <- 3 * (Y - 1)
Y2 <- ((Y1 - 2) / 2) ^ 2
EY2 <- 10.75
hist(Y2, freq = FALSE)
x <- seq(0, 120, by = 0.01)
y <- dexp(x, rate = 1 / EY2)
lines(x, y, col = "red")
```

Histogram of Y2



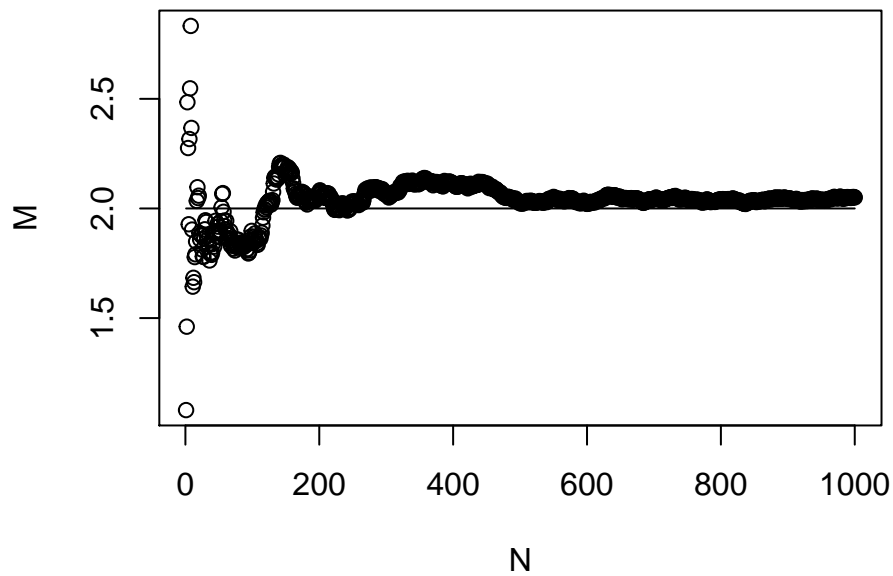
Next we will compute a sequence of means m_n and a sequence of variances σ_n^2 for the variable Y , where

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

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

and plot the results.

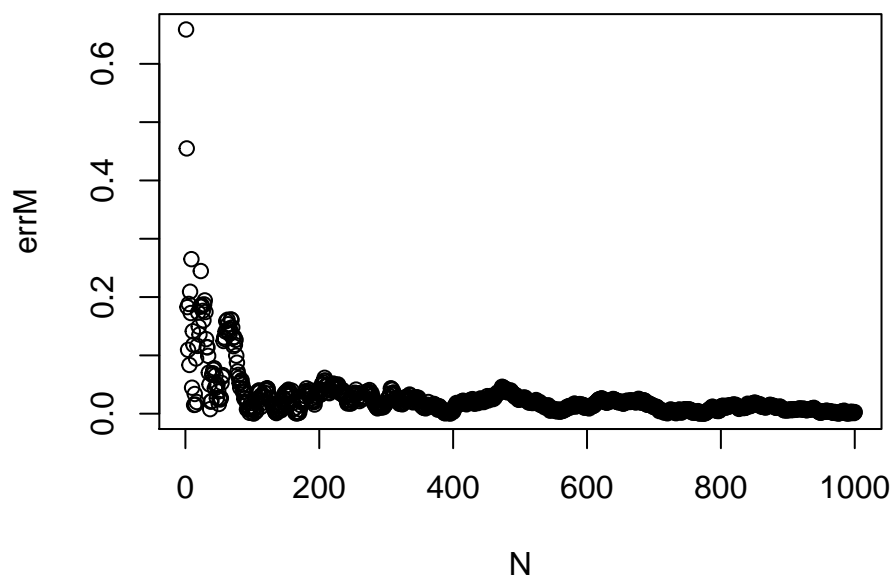
```
nn = 1000
N <- seq(1, nn, by=1)
Y <- rnorm(n = nn, mean = 2, sd = 2)
M <- cumsum(Y) / N
# V <- cumsum((Y - mean(Y)) ^ 2) / N
plot(N, M)
lines(N, rep(2, length(N), type = "l", col = "red"))
```



```
# plot(N, V)
# lines(N, rep(4, length(N), type = "l", col = "red"))
```

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.

$$err_M =$$

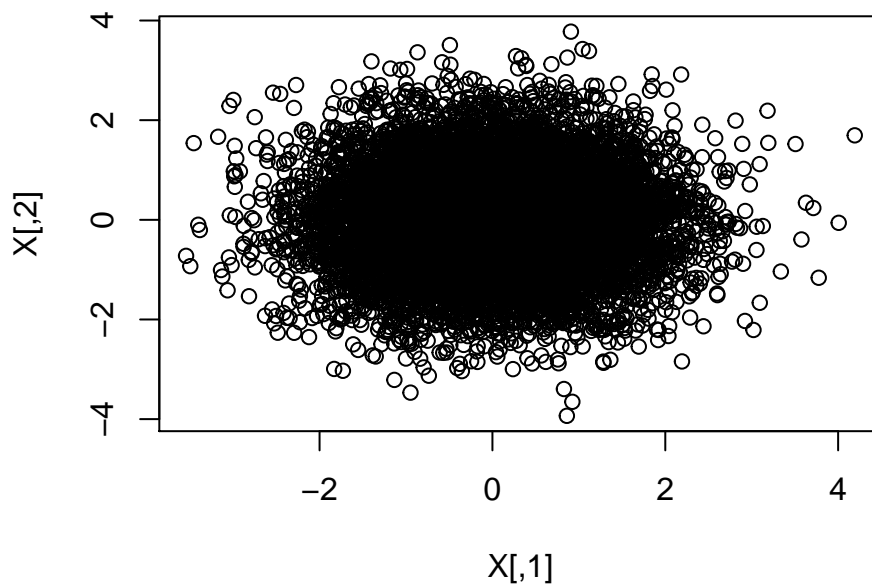


For $N > 200$ value of the error is less than 10% of theoretical mean.

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 $Y = AX + a$. From the lecture we know that if

$$X \sim N(0, I_2), \quad Y \sim N(\mu, \Sigma) \quad \text{and} \quad Y = AX + b,$$

then

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

$$Y = AX + a$$

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

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

It means that

$$A = \Sigma^{0.5} \quad \text{and} \quad a = \mu.$$

3 Exercise 3

If needed (e.g. when answering theoretical problems etc.), use L^AT_EX syntax to add mathematical formulas. For example, inline (within text) formulas can be included using: $a^2 + b^2 = c^2$. On the other hand, displayed equations can be added as follows:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \tag{1}$$

where n denotes sample size.

4 Exercise 4

A few final suggestions:

- The results of simulations or analysis should be presented in a clear format.
- When you plot the graphs, remember about axis descriptions, setting appropriate axis limits (e.g: `plot(..., xlim = c(x.min, x.max), ylim = c(y.min, y.max))`), adding legend (necessary when there are several lines in the graph), and adding a caption under the graph or table.
- If possible try to show the results in a condensed form (e.g. comparing several charts in one figure, comparing the results in a single table, etc.).
- It is worth to consider whether all R-codes should be visible in the report (option `echo = TRUE / FALSE`).
- Remember to carefully format the R-codes and write appropriate comments (also in the case of chunks that are invisible in the report. You can use the following option available in RStudio: *Code* → *Reformat Code*).

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 Dalgard, *Introductory Statistics with R*, Springer-Verlag New York, 2008.