Computational Numerical Statistics
PROJECT 1

Group G2 – TP1

Joao Camacho, nº 56861
Ana Mendes, nº 57144
Lia Schmid nº 57629
Emanuele Vivoli, nº 57284
Simao Goncalves, nº 54896

First Semester 2019–2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

References

Sheldon M Ross.
Simulation.
Academic press, 2002.

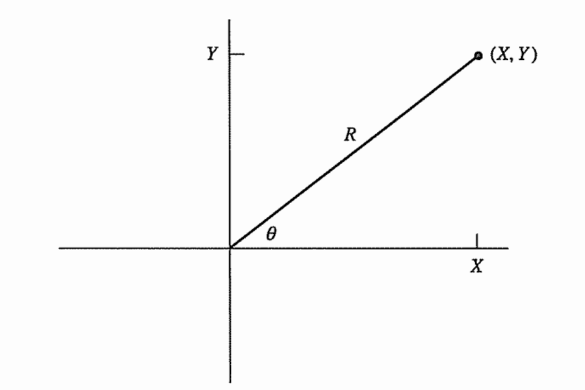
Sheldon M Ross.
Introduction to Probability Models, ISE.
Academic press, 2006.

Project task 1
Random variable generation

Random variable generation

Introduction

- Mervin Edgar Muller & George Edward Pelham Box, 19^{th} century
- Method for generating samples from two independent random variables X, Y $\sim N(0,1)$ from the $U(0,1)$ distribution
- Polar coordinates are created



[1]

Random variable generation

Derivation of the Box-Muller method

- $R^2 = X^2 + Y^2$ and $\theta = tan^{-1}(\frac{x}{y})$
- Proposition: Let $Z_1, ..., Z_v$ be v independent random variables such that $Z_i \sim N(0,1)$. Then $\sum_{i=1}^{v} Z_i^2 \sim \chi_v^2$
  so in our case $X^2 + Y^2 \sim \chi_2^2$

So we have the following p.d.f.

	χ_2^2	$\Gamma(1, \frac{1}{2})$	$Exp(\frac{1}{2})$
p.d.f.	$\frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{\frac{-x}{2}}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\lambda e^{-\lambda x}$
	$\frac{1}{2} e^{-\frac{1}{2}x}$		

$\Rightarrow R^2 \sim \frac{1}{2} e^{-\frac{1}{2}x}$

Let's consider the joined density:

$$f(x, y) \overset{X, Y \sim N(0,1)}{=} f(x) * f(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

$\rightarrow$ Joined density of $g(R^2, \theta)$?

given $J_{f(x,y)} = 2$ and $R^2 = X^2 + Y^2$:

$$g(R^2, \theta) = \frac{1}{|J_{f(x,y)}|} f(x, y) = \underbrace{\frac{1}{2\pi}}_{I} \underbrace{\frac{1}{2} e^{-\frac{1}{2} R^2}}_{II} \qquad [2]$$

Random variable generation

Derivation of the Box-Muller method: ITM

- Now we want to sample R^2 and θ from $U(0, 1)$
- **Inverse Transformation Method:** generate a random variable R^2 from the continuous distribution function F by generating a random number U and then setting $R^2 = F^{-1}(U)$

I. $u = \frac{1}{2}e^{-\frac{x}{2}} \Leftrightarrow x = -2log(\underbrace{2u}_{*})$ Remember: $R^2 \sim \frac{1}{2}e^{-\frac{1}{2}x}$

* $\sim U(0, 1)$ as well as -2log(u)

II. $g(R^2, \theta) = \frac{1}{2\pi}\frac{1}{2}e^{-\frac{1}{2}R^2} \Rightarrow \theta \in [0, 2\pi] \Rightarrow \theta = 2\pi u_2$

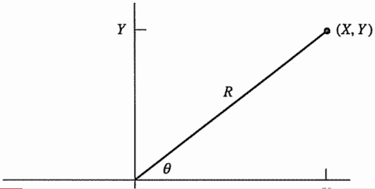
- once we have the radius R of the coordinates, the angle θ is uniformly distributed over the circumference with radius R.

From I. and II. we get

$$R = \sqrt{-2log(u_1)} \qquad X = Rcos(\theta)$$
$$\theta = 2\pi u_2 \qquad Y = Rsin(\theta)$$

$$\Rightarrow X = \sqrt{-2log(u_1)}cos(2\pi u_2)$$
$$\Rightarrow Y = \sqrt{-2log(u_1)}sin(2\pi u_2)$$

with $u_1, u_2 \sim U(0, 1)$



Q.: How to generate samples from $Z \sim N(\mu, \sigma^2)$?

$$y = ax + b \quad , \quad x \sim N(0,1)$$

$E[Y] = E[a \cdot X + b] = aE[X] + E[b] = a \cdot \mu + b$
Var[Y] = Var[a $\cdot X + b] = a^2 * Var[X] = a^2 \cdot \sigma^2$
$\sigma[Y] = \sqrt{Var[Y]} = a \cdot \sigma$

Note: $Y \sim N(\mu^*, \sigma^{2*})$:

$$p.d.f._Y = \frac{1}{\sigma^* \cdot \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y-\mu^*}{\sigma^*})^2}$$

$\rightarrow$ Find transformation that converts it to a $N(\mu^*,\sigma^{2*})$:

$$y = \sigma^* \cdot X + \mu^*$$

- Generate two observations u_1 and u_2 from U (0, 1).
- Set

$$\theta = 2\pi u_1$$

$$R = \sqrt{-2log(u_2)}$$

- Set

$$x = Rcos(\theta)$$

$$y = Rsin(\theta)$$

- Iterate the previous steps until you reach the desired sample size.

Random variable generation

Algorithm

(a) Implement the Box-Muller method in R.

```
maxLik(logLik=loglik.b,start=mme.graph.b)

# Maximum Likelihood estimation
# Newton-Raphson maximisation, 4 iterations
# Return code 1: gradient close to zero
# Log-Likelihood: -12.55405
# (1 free parameter)
# Estimate(s): 0.497838
```

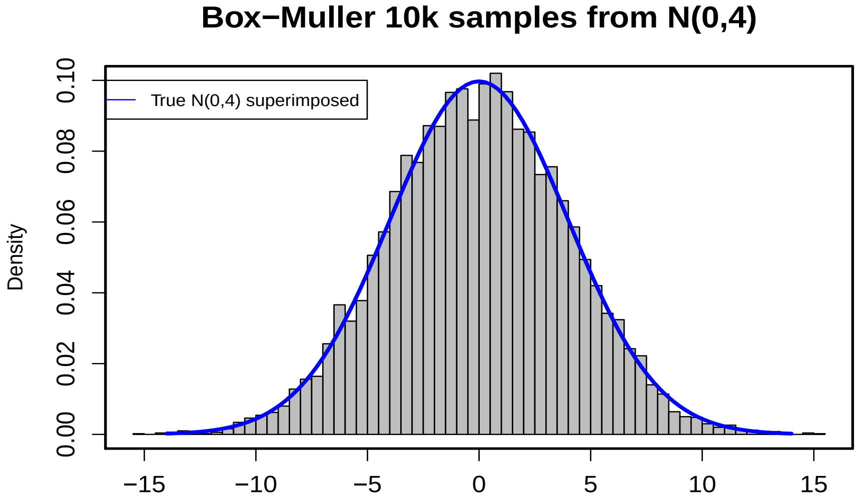
(b) Use routine sim.norm() to generate a sample of size m = 10000 from $X \frown N(0, 4)$.

```
#Exercise B - generate 10k samples from X~(0,4)
set.seed(123)
mu = 0
std = 4
n = 10 * 1000

# n/2 because we can merge X and Y since their both are independent.
samples = sim.norm(n/2,0,4)

# merge the 5k samples from X with the 5k samples from Y
samples = c(samples[,"X"], samples[,"Y"])
```

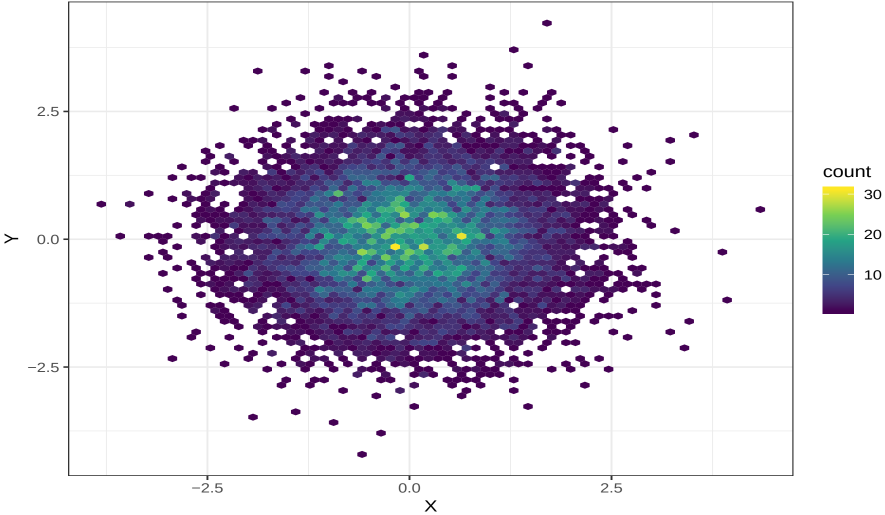
(c) Plot the histogram referring to the previous exercise.



Random variable generation

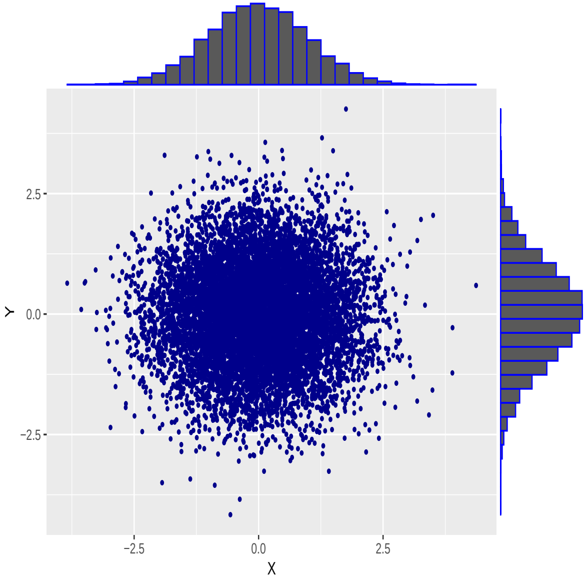
In-depth Visualization of the samples

Let´s see the density of the point cloud of X and Y.



Random variable generation

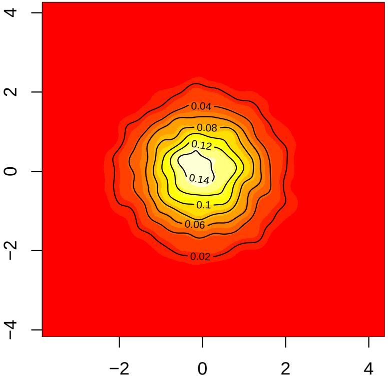
In-depth Visualization of the samples



As most points are located near the origin they increase the density of N(0,1) around the mean of 0.

Outlier points are barely accounted for since there are 10000 samples.

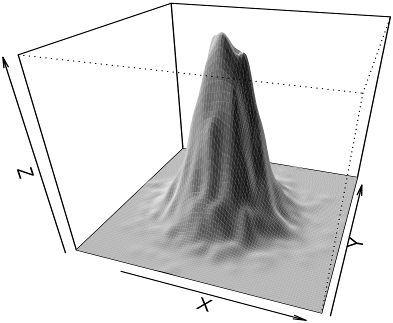
The different confidence regions, note that none of them are elliptical.



Calculating the covariance matrix and the mean of X and Y we get:

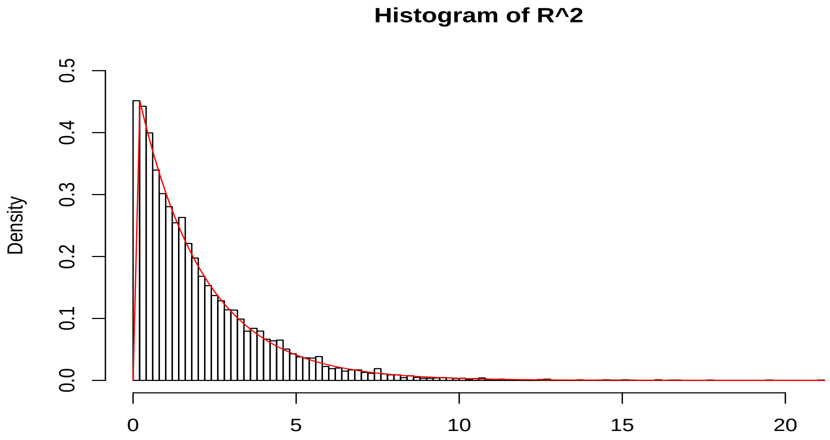
- for 10 samples, S= $\begin{bmatrix} 1.543 & -0.468 \\ -0.468 & 0.452 \end{bmatrix}$; mean = $\begin{bmatrix} 0.1303 \\ -0.4074 \end{bmatrix}$
- for 100 samples, S= $\begin{bmatrix} 0.8660 & -0.0221 \\ -0.0221 & 0.8487 \end{bmatrix}$; mean = $\begin{bmatrix} -0.1112 \\ 0.0059 \end{bmatrix}$
- for 1000 samples, S= $\begin{bmatrix} 1.02431 & 0.00116 \\ 0.00116 & 0.97704 \end{bmatrix}$; mean= $\begin{bmatrix} -0.0237 \\ -0.0341 \end{bmatrix}$
- for 10000 samples, S= $\begin{bmatrix} 1.00123 & 0.00103 \\ 0.00103 & 1.01359 \end{bmatrix}$; mean= $\begin{bmatrix} -0.0134 \\ -0.0041 \end{bmatrix}$

Bivariate Normal KDE:



Random variable generation
Visualization

Since $X, Y \frown N(\mu, \sigma^2)$, and $R^2 = X^2 + Y^2$ then $R^2 \frown \chi^2_2$
In the special case of two degrees of freedom, the chi-squared distribution coincides with the exponential distribution.



Random variable generation

This is the end

Questions?

If we finished the presentation on time, of course :-)