Dear Christine & Sophia
students

Could you please make sure that the recording is properly managed? We had major losses on the last one. Rec is currently ON BUT ON PAUSE

Also, if any assignements are given written instructions are put on Moodle.

Thanks Seb

Theorem: law of large numbers

let $X_1, \dots, X_n$ be n iid random variables

We denote by $\mu := \mathbb{E}[X_1]$ and $\sigma^2 = V[X_1]$.

Then

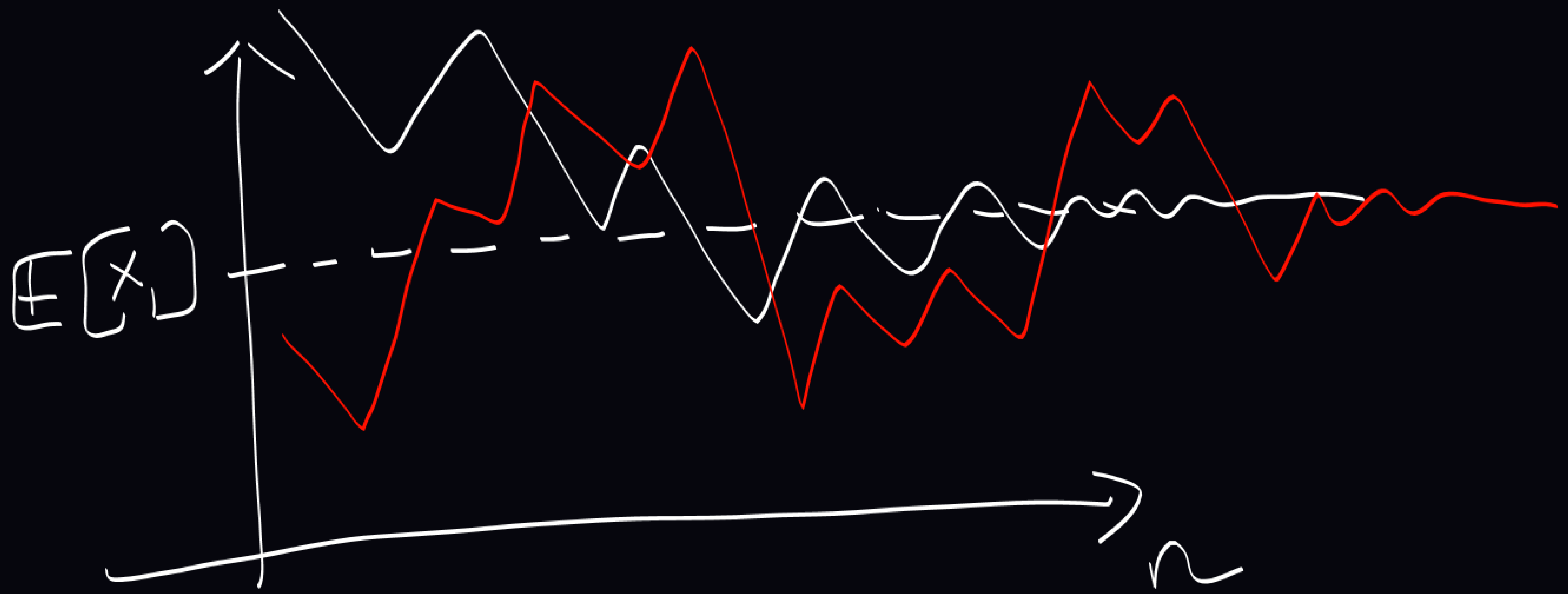
$$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[P]{n \to +\infty} \mu$$

definition of the convergence in probability ($\xrightarrow[P]{}$)

$X_n \xrightarrow[P]{} a$ if $\forall \varepsilon > 0$,

$$P(|X_n - a| > \varepsilon) \xrightarrow[n \to +\infty]{} 0$$

? How to show with the R software this theorem



Ex :

1) Simulate 5000 observations from an exponential with $\lambda = 4$. Plot the trajectory associated to the empirical mean.

I simulate $X_1 \longrightarrow$ I compute $\overline{X}_1 = X_1$

I simulate X_1 and $X_2 \longrightarrow$ I compute

$$\overline{X}_2 = \frac{1}{2}\left(X_1 + X_2\right)$$

This is FALSE!

2) Do the same but to obtain 10 trajectories of $\overline{X}_n$ on a same graphic!

```
cy = y[1]
for (i in 2:5000)
  {
    cy = c(cy, cy[i-1] + y[i])
  }
y2 = cy/(1:5000)
```

$$\overline{X}_n = \frac{1}{n} \underbrace{\sum_{i=1}^{n} X_i}_{cy}$$

Rk:

Markov inequality:

Let X be a random random variable with μ for expectation and σ^2 for variance:

$$\forall \varepsilon > 0, \quad P(|X-\mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

We apply this inequality to

$\overline{X}_n$

$\forall \varepsilon > 0, \quad P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) \leq \dfrac{Var(\overline{X}_n)}{\varepsilon^2}$

$$\mu = E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu = \mu = E[X_1]$$

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

$\hookrightarrow \forall \varepsilon > 0$

$$P\left(\left|\overline{X}_n - \mathbb{E}[X_1]\right| > \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

Ex: if we take $\varepsilon = 0.01$

the upper limit is $\dfrac{\sigma^2}{10^{-4} \times n} \xrightarrow[n \to +\infty]{} 0$

if we want that this term is less than 10^{-3}

Theorem: Central limit

let $X_1, \dots, X_n$ be iid r.v with expectation μ and variance σ^2.

let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\overline{X}_n \underset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Then $\sqrt{n} \, \frac{\overline{X}_n - \mu}{\sigma} \underset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, 1) \leftarrow$

We say that

$$X_n \xrightarrow[\mathcal{L}]{} X$$

if $\forall t \in \mathbb{R}$, $F_n(t) \xrightarrow[n \to +\infty]{} F(t)$

F_n : distribution function of X_n

F : ——————————————— X

? How to see this result

Ex:

1) let n be a integer

let $X_1, \dots, X_n$ be $\mathcal{U}([0;3])$

Estimate with a histogram the density of $\sqrt{n}\,\dfrac{\bar{X}_n - \mu}{\sigma}$ and compare it to $\mathcal{N}(0;1)$

$\left(\begin{array}{l} \mu = \mathbb{E}[X_1] \\ \sigma^2 = \mathbb{V}[X_1] \end{array} \right)$

Rk :

n

K $\qquad \mathcal{U}([0;3])$

$$\longrightarrow \overline{X}_n^{(1)} \longrightarrow \sqrt{n}\,\frac{\overline{X}_n^{(1)}-\mu}{\sigma}$$

$$\longrightarrow \overline{X}_n^{(2)} \longrightarrow \sqrt{n}\,\frac{\overline{X}_n^{(2)}-\mu}{\sigma}$$

$$\longrightarrow \overline{X}_n^{(K)} \longrightarrow \sqrt{n}\,\frac{\overline{X}_n^{(K)}-\mu}{\sigma}$$

K times (for example $K = 500$) → for loop

a) We simulate n observations from a uniform on $[0;3]$ denoted $u_1, \dots, u_n$

b) with the n observations, we compute

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} u_i$$

→ At the end of the for loop K observations of $\bar{X}_n$.

Let M be the vector that contains the K values of $\overline{X_n}$.

We compute $(M - \mu) \times \sqrt{n} / \sigma \to N$

RK: If $X \sim \mathcal{U}([0,3])$ $\quad \mu = 1.5$

$\sigma^2 = \dfrac{9}{12}$

We plot the histogram of N + density of $\mathcal{N}(0,1)$

Rk:

Thanks to this theorem, we prove that $\mathcal{B}(n;p) \underset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(np;npq)$

In practice, we consider the approximation when $n \geq 30$, $np > 5$ and $npq > 5$

```r
TCL <- function(n, K)
{
  U = matrix(data=3*runif(n*K),
             ncol= n, nrow=K)

  M = apply(U, 1, mean)
  N = sqrt(n)*(M-1.5)/sqrt(9/12)
```

```
H = hist(N, freq=FALSE)
limit = H$breaks
ll = min(limit)
up = max(limit)
x = seq(ll, up, 0.01)
y = dnorm(x)
lines(x, y, col = "red")
```

TCL(5, 500)

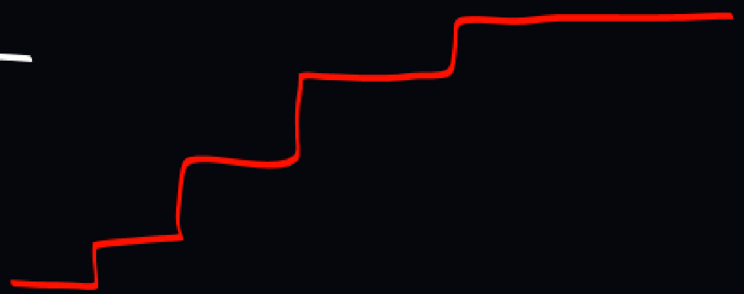
TCL(30, 500)

With histograms, is it difficult to see all the things on a same graph

$\rightarrow$ distribution function

$\Rightarrow$ on a same graph, plot the distribution functions associated to several values of n

and the theoretical limit

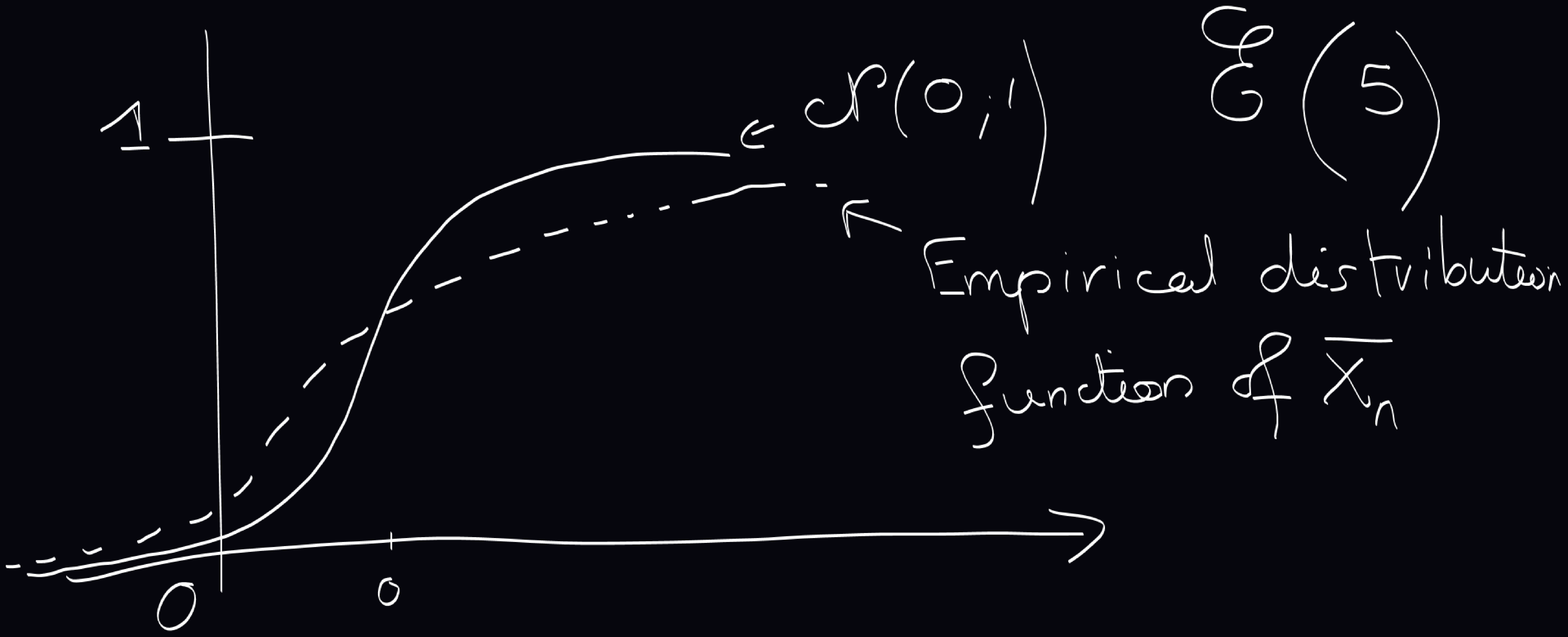


$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{x_i < t}$$

$$\bar{x}_n^{(1)} < \bar{x}_n^{(2)} < \cdots < \bar{x}_n^{(k)}$$

(empirical) distribution function

(approximation of $P(X_n < t)$)



$\mathcal{E}(5)$
$\mathcal{N}(0,1)$
1
0
0
Empirical distribution function of $\overline{X}_n$

How to do simulations of not common distribution?

- first technic based on the distribution function

Let X be a continuous r.v.
let f be the density of X and F the distribution f^{ct} of X.

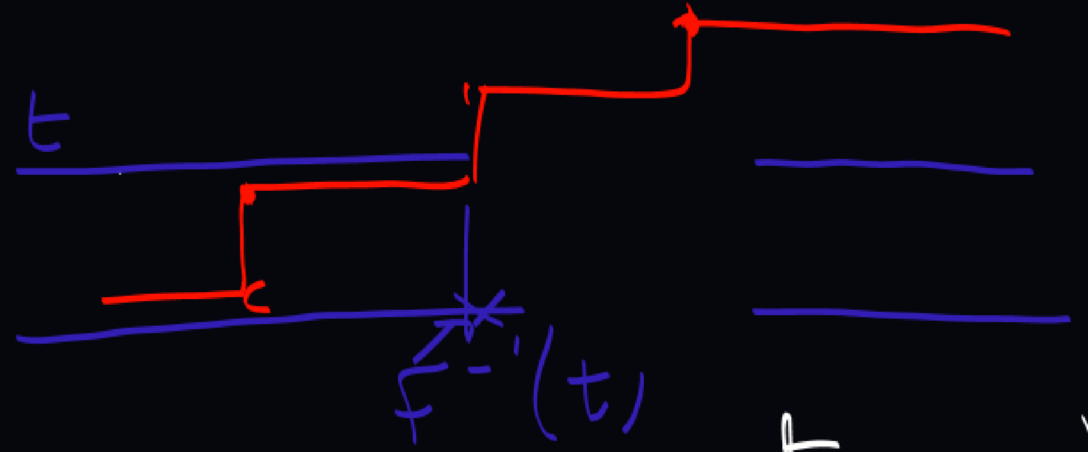
Compute F^{-1} $\left(\begin{array}{l} F \circ F^{-1}(x) = x \\ F^{-1} \circ F(x) = x \end{array} \right)$

Then $Y = F^{-1}(U)$ with $U \sim \mathcal{U}([0,1])$

has the same distribution

than X.

Ex: Apply this method to simulate an $\mathcal{E}(\lambda)$



Rk: we can replace the inverse f^{ct} F^{-1} by the pseudo inverse F^{-1}

$$F^{-1}(t) = \inf\{u \in \mathbb{R} / F(u) \geq t\}$$

Rk:

$]0,3]$ 0 infimum

not a minimum!

- <u>Second method: reject</u>

Let X be a random variable with density f. (we are not able to perform simulations from f).

We assume that there exists $c>0$ and a density f^{ct} g such that $f \leq c \times g$

g should be a density from which we are able to make simulations.

let z be a simulation from g.

let u ———————— $\mathcal{U}([0,1])$.

If $u \times c \times g(z) < f(z)$ then we keep z

otherwise we forget z and we begin again!

<u>Ex :</u> for monday :

To simulate a $\mathcal{N}(0,1)$, we can use this method.

$$g(x) = \frac{1}{2}\exp(-|x|) \qquad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

and $c = \sqrt{\frac{2\exp(1)}{\pi}}$

1) Proof of

$$\forall x \in \mathbb{R}, \ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \leq c \times \frac{\exp(-|x|)}{2}$$

2) Write a f^{ct} to program the reject method for $\mathcal{N}(0;1)$

c) Influence of c