TP 1 : Fonction de répartition empirique avec R.

Hamel Elhadj

Departement de Mathematiques

Université Hassiba Benbouali-Chlef

Ce TP est distiné aux étudiants Master2 mathématiques

Option :Mathématique Apliquées et statistique

2021/2022

6 novembre 2021

Introduction

Le but de ce TP est de fournir les éléments de base permettant une prise en main rapide du logiciel R afin de faciliter le déroulement de certains TD mais aussi de vous aider de faire le patique.

Un point fort de logiciel R réside dans le fait que ce logiciel est distribué librement. Son installation peut être mise en œuvre à partir du site internet du *Comprehensive R Archive Network* (CRAN) qui d'une part met à disposition les exécutables et d'autres part donne des informations relatives à la procédure d'installation.

Création de données

Création de données.

Il existe différentes manière de créer des données quelles soient régulières ou aléatoires

1. **a :b** $\Longrightarrow$ qui crée, à partir de la valeur a et par pas de 1, une suite de nombres inférieurs ou égaux à b.
2. **seq(a,b,p)** $\Longrightarrow$ où a est la valeur de départ, b la valeur maximale à ne pas dépasser et p le pas.
3. $c(n_1, n_2, n_3)$ $\Longrightarrow$ qui produit le vecteur contenant les valeurs n_1 ; n_2 ; n_3.
4. D'autres fonctions sont également disponibles telles :

   -**rep** qui permet de répéter un objet un certain nombre de fois ;

   - **sequence** qui crée une suie de séquences de nombres entiers se terminant chacune par les valeurs spécifiées en argument

   -**gl(k,n)** qui génère une série régulière dans un facteur comprenant k niveaux

Création de données

Création de données.

Il existe différentes manière de créer des données quelles soient régulières ou aléatoires

1. **a :b** $\Longrightarrow$ qui crée, à partir de la valeur a et par pas de 1, une suite de nombres inférieurs ou égaux à b.
2. **seq(a,b,p)** $\Longrightarrow$ où a est la valeur de départ, b la valeur maximale à ne pas dépasser et p le pas.
3. $c(n_1, n_2, n_3)$ $\Longrightarrow$ qui produit le vecteur contenant les valeurs n_1 ; n_2 ; n_3.
4. D'autres fonctions sont également disponibles telles :

   -**rep** qui permet de répéter un objet un certain nombre de fois ;

   - **sequence** qui crée une suie de séquences de nombres entiers se terminant chacune par les valeurs spécifiées en argument

   -**gl(k,n)** qui génère une série régulière dans un facteur comprenant k niveaux

Création de données

Création de données.

Il existe différentes manière de créer des données quelles soient régulières ou aléatoires

1. **a :b** $\Longrightarrow$ qui crée, à partir de la valeur a et par pas de 1, une suite de nombres inférieurs ou égaux à b.
2. **seq(a,b,p)** $\Longrightarrow$ où a est la valeur de départ, b la valeur maximale à ne pas dépasser et p le pas.
3. $c(n_1, n_2, n_3)$ $\Longrightarrow$ qui produit le vecteur contenant les valeurs n_1 ; n_2 ; n_3.
4. D'autres fonctions sont également disponibles telles :

   -**rep** qui permet de répéter un objet un certain nombre de fois ;

   - **sequence** qui crée une suie de séquences de nombres entiers se terminant chacune par les valeurs spécifiées en argument

   -**gl(k,n)** qui génère une série régulière dans un facteur comprenant k niveaux

Création de données

Création de données.

Il existe différentes manière de créer des données quelles soient régulières ou aléatoires

1. **a :b** $\Longrightarrow$ qui crée, à partir de la valeur a et par pas de 1, une suite de nombres inférieurs ou égaux à b.
2. **seq(a,b,p)** $\Longrightarrow$ où a est la valeur de départ, b la valeur maximale à ne pas dépasser et p le pas.
3. $c(n_1, n_2, n_3)$ $\Longrightarrow$ qui produit le vecteur contenant les valeurs n_1 ; n_2 ; n_3.
4. D'autres fonctions sont également disponibles telles :

   -**rep** qui permet de répéter un objet un certain nombre de fois ;

   - **sequence** qui crée une suie de séquences de nombres entiers se terminant chacune par les valeurs spécifiées en argument

   -**gl(k,n)** qui génère une série régulière dans un facteur comprenant k niveaux

Données aléatoires

Voici un tableau qui fait apparaître la dénomination et les paramètres de certaines lois.

nom de la loi	fonction dans R
Gauss(normale)	rnorm(n,mean=μ ,sd=σ)
exponentielle	rexp(n, rate=λ)
gamma	rgamma(n, shape = a, scale = s)
poisson	rpois(n, λ)
weibull	rweibull(n, shape = a, scale = s)
cauchy	rcauchy(n, location = a, scale = s)
student	rt(n, df)
fisher	rf(n,df1,df2)
binomiale	rbinom(n,size,prob)
géométrique	rgeom(n,prob)
uniforme	runif(n, min=a, max=b)

fonctions graphiques

Dans ce tableau ne sera mentionné que les fonctions qui pourront être utiles en statistique. Cependant, nous n'en donnerons qu'un petit aperçu, les détails et options figurant dans l'aide.

plot(x)	graphe des valeurs de x (sur l'axe des y) ordonnées sur l'axe x
plot(x,y)	graphe de y en fonction de x
boxplot(x)	boîte à moustaches
hist(x)	histogramme des fréquences de x
barplot(x)	histogramme des valeurs de x
qqnorm(x)	quantiles de x en fonction des valeurs attendues selon une loi normale
qqplot(x,y)	quantiles de y en fonction de ceux de x

TP01.Loi normale et densité

1. Simuler un échantillon de taille 1000 suivant une loi normale N(0,1).
2. Sur un même graphique, superposer la répartition empirique (histogramme) et la densité de la loi N(0,1).
3. Sur un même graphique, superposer la fonction de répartition empirique et la fonction de répartition de la loi N(0,1).
4. Reprendre les questions précédentes avec d'autres lois de votre choix.

TP01.Loi normale et densité

1. Simuler un échantillon de taille 1000 suivant une loi normale N(0,1).
2. Sur un même graphique, superposer la répartition empirique (histogramme) et la densité de la loi N(0,1).
3. Sur un même graphique, superposer la fonction de répartition empirique et la fonction de répartition de la loi N(0,1).
4. Reprendre les questions précédentes avec d'autres lois de votre choix.

TP01.Loi normale et densité

1. Simuler un échantillon de taille 1000 suivant une loi normale N(0,1).
2. Sur un même graphique, superposer la répartition empirique (histogramme) et la densité de la loi N(0,1).
3. Sur un même graphique, superposer la fonction de répartition empirique et la fonction de répartition de la loi N(0,1).
4. Reprendre les questions précédentes avec d'autres lois de votre choix.

TP01.Loi normale et densité

1. Simuler un échantillon de taille 1000 suivant une loi normale N(0,1).
2. Sur un même graphique, superposer la répartition empirique (histogramme) et la densité de la loi N(0,1).
3. Sur un même graphique, superposer la fonction de répartition empirique et la fonction de répartition de la loi N(0,1).
4. Reprendre les questions précédentes avec d'autres lois de votre choix.

TP01.Loi normale et densité

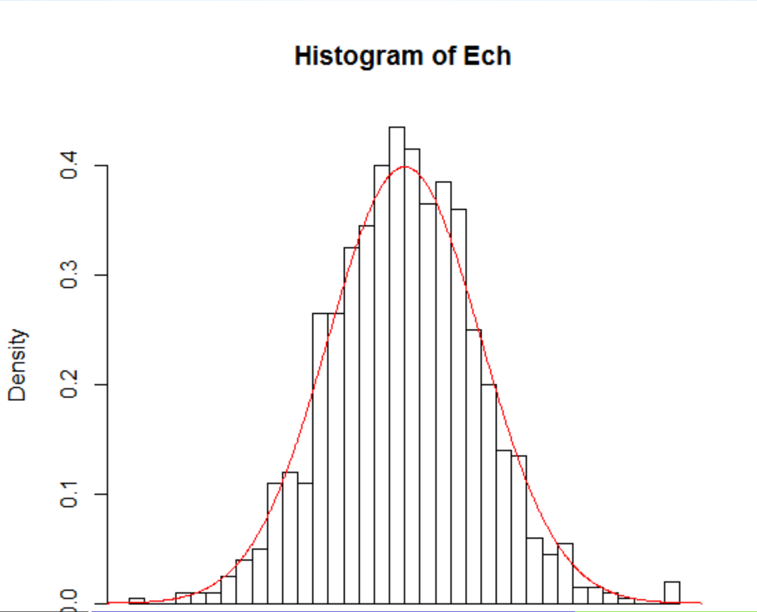
Question 1

```
Ech = rnorm(1000, 0, 1)
```

Question 2

```
hist(Ech, breaks = 50, freq = F)
points(seq(-4, 4, 0.01), dnorm(seq(-4, 4, 0.01)),
type = "l", col = "red")
```

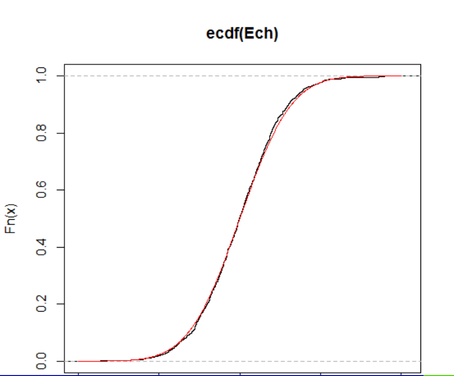
TP01.Loi normale et densité



TP01.Loi normale et densité

Question 3

```
?ecdf
plot(ecdf(Ech))
points(seq(-4, 4, 0.01), pnorm(seq(-4, 4, 0.01)), col
= "red", type = "l")
```



Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

Excercice

1. Tracer la densité d'une loi normale centrée réduite sur [-5, 5]. Pour cela :
   - i- Créer un vecteur x qui varie doucement de -5 a 5.
   - ii- Créer un vecteur y qui donne la densité de la loi normale centrée réduite pour tous les points de x.
   - iii- Tracer le graphique. Attention, il faut tracer une courbe, pas des points côte à côte.
2. Superposer la densité d'une loi normale centrée de variance 0.5. Attention : on veut voir entiérement les deux densités.
3. Superposer la densité d'une loi normale centrée de variance 2.

TP01.Loi uniforme et densité

```
 x = runif(20)
xseq = seq(0,1,length=100)
Fhat = ecdf(x)
plot(Fhat)
lines(xseq,punif(xseq), col="blue",lwd=2)
```

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

Fonction de répartition empirique

TP : fonction de répartition .

Le but de cet TP est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1. Générer un échantillon de taille n = 100 de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{E}$ de paramètre λ à choisir.
2. Soit F la fonction de répartition de la loi $\mathcal{E}$. Estimer F à partir de la fonction de répartition empirique $\widehat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Représenter graphiquement F et $\widehat{F}_n$.
4. Illustrer la convergence presque sûre de $\widehat{F}_n$ vers $F(x)$.
5. Illustrer également la normalité asymptotique associée à $\widehat{F}_n$.

TP01.Loi normale et densité

1. Question 1 :

```
n=100 ; lambda=2 ; X=rexp(n,lambda) ; plot(X)
```

2. Question 2

on découpe l'intervalle [0,3] en une grille de 3000 points (sequence)

```
xmin=0 ; xmax=3 ;nbpoints=3000
x = seq(xmin, xmax, length.out=nbpoints)
Fn = rep(0, nbpoints) for (i in 1:nbpoints) Fn[i]
= length(X[X <= x[i]])/n  plot(Fn)
plot(ecdf(X))
```

TP01.Loi normale et densité

1. Question 1 :

```
n=100 ; lambda=2 ; X=rexp(n,lambda) ; plot(X)
```

2. Question 2

on découpe l'intervalle [0,3] en une grille de 3000 points (sequence)

```
xmin=0 ; xmax=3 ;nbpoints=3000
x = seq(xmin, xmax, length.out=nbpoints)
Fn = rep(0, nbpoints) for (i in 1:nbpoints) Fn[i]
= length(X[X <= x[i]])/n  plot(Fn)
plot(ecdf(X))
```

TP01.Loi normale et densité

1. Question 1 :

```
n=100 ; lambda=2 ; X=rexp(n,lambda) ; plot(X)
```

2. Question 2 :

on découpe l'intervalle [0,3] en une grille de 3000 points (sequence)

```
xmin=0 ; xmax=3 ;nbpoints=3000
x = seq(xmin, xmax, length.out=nbpoints)
Fn = rep(0, nbpoints) for (i in 1:nbpoints) Fn[i]
= length(X[X <= x[i]])/n  plot(Fn)
plot(ecdf(X))
```

TP01.Loi normale et densité

1. Question 1 :

   ```
   n=100 ; lambda=2 ; X=rexp(n,lambda) ; plot(X)
   ```

2. Question 2 :

   on découpe l'intervalle [0,3] en une grille de 3000 points (sequence)

   ```
   xmin=0 ; xmax=3 ;nbpoints=3000
   x = seq(xmin, xmax, length.out=nbpoints)
   Fn = rep(0, nbpoints) for (i in 1:nbpoints) Fn[i]
   = length(X[X <= x[i]])/n  plot(Fn)
   plot(ecdf(X))
   ```

TP01.Loi normale et densité

1. Question 1 :

```
n=100 ; lambda=2 ; X=rexp(n,lambda) ; plot(X)
```

2. Question 2 :

   on découpe l'intervalle [0,3] en une grille de 3000 points (sequence)

```
xmin=0 ; xmax=3 ;nbpoints=3000
x = seq(xmin, xmax, length.out=nbpoints)
Fn = rep(0, nbpoints) for (i in 1:nbpoints) Fn[i]
= length(X[X <= x[i]])/n  plot(Fn)
plot(ecdf(X))
```

TP : fonction de répartition

- Question 3 :

  on calcule F fonction de répartition de la loi expo en chaque point de la grille

  et on trace F et Fn

```
F = pexp(x, lambda)
plot(x, Fn, type="s", col="blue")
lines(x, F, type="l", col="red")
```

  comparer avec `plot(ecdf(X))????`

TP : fonction de répartition

1. Question 3 :

   on calcule F fonction de répartition de la loi expo en chaque point de la grille

   et on trace F et Fn

   ```
   F = pexp(x, lambda)
   plot(x, Fn, type="s", col="blue")
   lines(x, F, type="l", col="red")
   ```

   comparer avec `plot(ecdf(X))????`

TP : fonction de répartition

1. Question 3 :

   on calcule F fonction de répartition de la loi expo en chaque point de la grille

   et on trace F et Fn

   ```
   F = pexp(x, lambda)
   plot(x, Fn, type="s", col="blue")
   lines(x, F, type="l", col="red")
   ```

   comparer avec `plot(ecdf(X))????`

TP : fonction de répartition

1. Question 4 : on fait varier n entre 1 et N mais on se fixe un x0

```
x0=0.2
N=3000
X2=rexp(N,lambda);Fchap = rep(0, N)
for (n in 1:N) Xcoupe = X2[1:n]
Fchap[n] = length(Xcoupe[Xcoupe <= x0])/n
plot(1:N, Fchap, type="l", col="blue")
abline(h=pexp(x0, lambda), lty=2, col="red") %
F(x0)
```

TP : fonction de répartition

1. Question 5 : On fixe n assez grand, on fixe x0, et on simule N réalisations de
$sqrt(n) * [Fchapeau_n(x0) - F(x0)]$
lambda=2 ;n = 100 ; x0 = 0.6 ; N = 1000 ;Dn = rep(0, N)
for (i in 1 :N) X = rexp(n, lambda) Dn[i] = sqrt(n)*(length(X[X <= x0])/n - pexp(x0, 2))
on trace l'histogramme de ces réalisations et on superpose la densité de la loi normale attendue centrée de variance $F(x0)(1 - F(x0))$
IntTLC = seq(-2, 2, 0.1) ; variance = pexp(x0, lambda)*(1-pexp(x0, lambda))
hist(Dn, col='light blue', xlim=c(-3,3), breaks=seq(min(Dn)-0.5, max(Dn)+0.5, 0.2), freq=FALSE)
plot(function(IntTLC) dnorm(IntTLC, 0, sqrt(variance)), xlim=c(-2,2), col=2, add=TRUE, lwd=2)

TP : fonction de répartition

1. Question 6 : on fait varier n et on prend aă chaque fois le sup de $|\hat{F}_n(x) - F(x)|$ sur tous les points de la grille

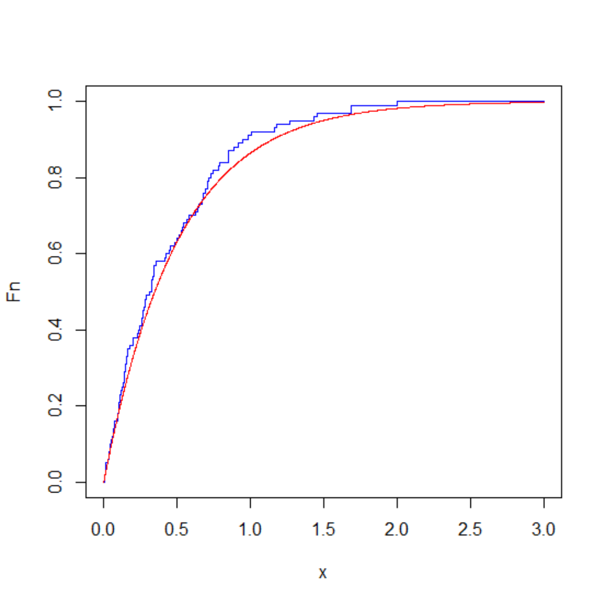
   N = 300 ; supCumul = rep(0, N) ; X = rexp(N, lambda) ;F = pexp(x, lambda)

   for (n in 1 :N) Xcoupe = X[1 :n] Fnx = rep(0, nbpoints) for (j in 1 :nbpoints)

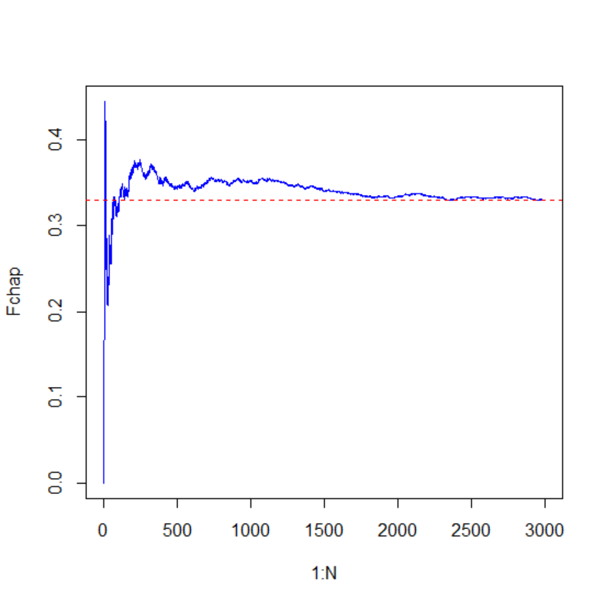
   Fnx[j] = length(Xcoupe[Xcoupe <= x[j]])/n supCumul[n] = max(abs(Fnx-F))

   plot(1 :N, supCumul, type="l", col="blue")

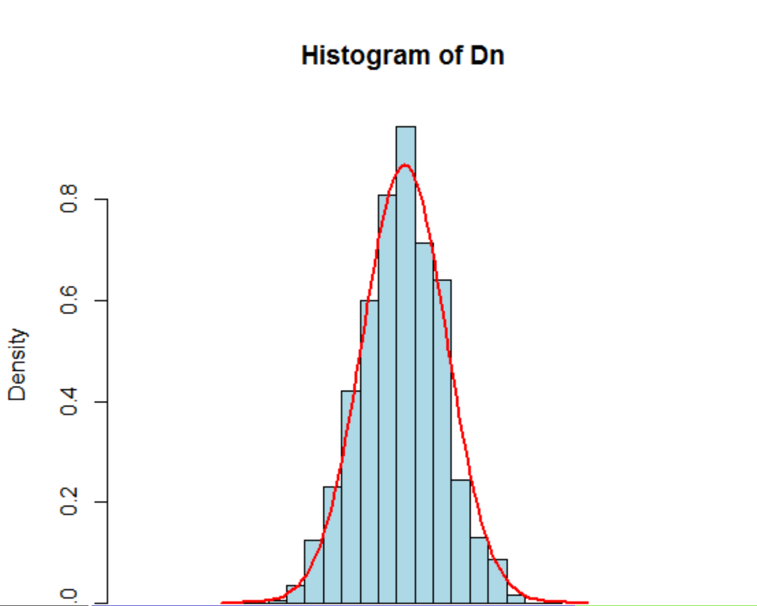
TP : fonction de répartition



TP : fonction de répartition



TP : fonction de répartition



TP : fonction de répartition