

# **Simulation et Estimation Loi de Pareto**

**Tom Boumba et Angelo Castino**

Encadré par : Pr. Bruno Portier, Mr. Michel Bobia  
PROJET GM3 / S2V2 - MSRO



## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Partie théorique : étude mathématique</b>	<b>3</b>
2.1	Étude probabiliste . . . . .	3
2.2	Simulation d'une valeur d'une variable de loi de Pareto . . . . .	6
2.3	Étude statistique . . . . .	7
2.4	Récapitulatif de la partie mathématique . . . . .	10
<b>3</b>	<b>Résultats numériques</b>	<b>11</b>
3.1	Simulation de la loi de Pareto . . . . .	11
3.2	Estimation du paramètre $\alpha$ . . . . .	13
3.2.1	Graphique d'évolution de l'estimateur . . . . .	13
3.2.2	Boîtes à moustaches . . . . .	15
<b>4</b>	<b>Conclusion</b>	<b>18</b>
<b>5</b>	<b>Bibliographie</b>	<b>18</b>
<b>6</b>	<b>Annexe</b>	<b>19</b>
6.1	Figures . . . . .	19
6.2	Listing du programme R . . . . .	26

# 1 Introduction

A la fin du  $XIX^e$  siècle, l'économiste italien Vilfredo Pareto a mis au point un principe qui porte aujourd'hui son nom : Le principe de Pareto. En effet, en analysant les données fiscales des plus puissants pays européens, il a remarqué que nonobstant une certaine variabilité selon le pays, le pourcentage de la population possédant une richesse supérieure ou égale à un  $x$  donné était proportionnel à  $\frac{A}{x^\alpha}$  ( $A$  étant une constante, et le coefficient  $\alpha$  variant selon les pays). Ses travaux ont inspiré beaucoup de ses collègues qui ont mené des recherches approfondies sur ce phénomène, ce qui a mené à la formalisation de la loi de Pareto, aujourd'hui très importante en économie, et permettant par exemple d'optimiser la gestion de stocks, de clients ou de temps. Une loi de Pareto est donc aujourd'hui formalisée par la théorie des probabilités comme étant une loi à densité hyperbolique permettant de modéliser probabilistiquement une relation de cause à effet asymétrique. Dans ce projet, nous étudions tout d'abord les caractéristiques théoriques d'une variable aléatoire suivant une loi de Pareto (espérance, variance, fonction de répartition inverse, estimateur de maximum de vraisemblance). On trouve notamment que le paramètre  $\alpha$  de la loi doit être strictement supérieur à 2 pour que la variable admette une espérance et une variance finies, et que l'estimateur de maximum de vraisemblance du paramètre  $\beta = \frac{1}{\alpha}$  est sans biais et convergent. Ensuite, nous nous intéressons à la simulation d'une telle variable aléatoire par la méthode de l'inversion, qui est applicable car la fonction de répartition de la variable est continue et strictement croissante. Enfin, en nous basant sur un échantillon de causes observées, qui sont supposément des réalisations indépendantes d'une variable aléatoire suivant une loi de Pareto, on cherche à déterminer le paramètre  $\alpha$  de la loi en question par la méthode de l'estimation par maximum de vraisemblance. Cela nous permet de mettre de évidence le caractère sans biais et convergent de l'estimateur de maximum de vraisemblance.

## 2 Partie théorique : étude mathématique

### 2.1 Étude probabiliste

1. Soient  $\alpha$  et  $\theta$  deux réels strictement positifs. Déterminer le réel  $C$  pour que la fonction :  $x \rightarrow \begin{cases} \frac{C}{x^{\alpha+1}} & \text{si } x \geq \theta \\ 0 & \text{sinon.} \end{cases}$  soit une densité de probabilité.

On sait que  $f$  est une densité si :

$$\begin{aligned} \int_{-\infty}^{+\infty} f(t)dt &= 1 \Leftrightarrow \int_{\theta}^{+\infty} \frac{C}{x^{\alpha+1}} dx = 1 \\ &\Leftrightarrow C \int_{\theta}^{+\infty} \frac{1}{x^{\alpha+1}} dx = 1 \\ &\Leftrightarrow C \left[ \frac{-1}{\alpha x^{\alpha}} \right]_{\theta}^{+\infty} = 1 \\ &\Leftrightarrow \frac{C}{\alpha \theta^{\alpha}} = 1 \\ &\Leftrightarrow C = \alpha \theta^{\alpha} \end{aligned}$$

On obtient que :  $f(x) = \begin{cases} \frac{\alpha}{\theta} \left( \frac{\theta}{x} \right)^{\alpha+1} & \text{si } x \geq \theta \\ 0 & \text{sinon} \end{cases}$

Dans toute la suite, on notera cette densité  $f$ .

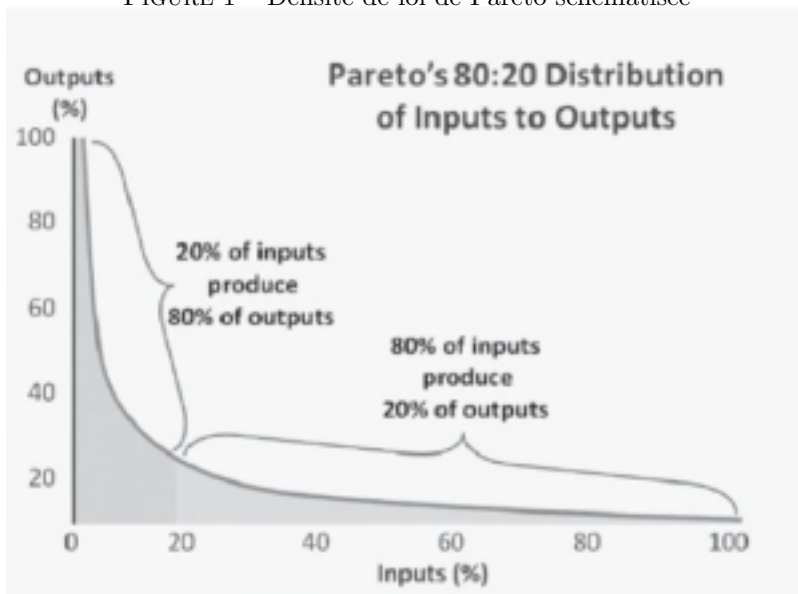
**Définition :** On dit que la variable aléatoire  $X$  suit une loi de Pareto de paramètres  $\alpha$  et  $\theta$  si et seulement si sa densité est  $f$ . On notera alors  $X \sim P(\alpha, \theta)$ .

Dans ce qui suit,  $X$  désignera une variable aléatoire de loi  $P(\alpha, \theta)$

2. Dans quelle contexte cette loi est-elle utilisée ? Donner des exemples de modélisation avec une telle loi.

Après quelques recherches, on apprend que cette loi est utilisée en économie principalement. Elle est néanmoins très discutée car beaucoup la trouvent trop imprécise. A l'aide de la loi de Pareto on peut notamment modéliser de la gestion de stock, de ventes, de projet ou de production. On appelle aussi cette loi la loi des 80-20 car elle permet d'observer que 80% des effets sont le produit de seulement 20% des causes. En effet, si on modélise un problème comme étant causé par les valeurs que peuvent prendre  $X$ , on voit qu'après  $\theta$ , approximativement 20% des valeurs de cause constituent 80% de l'intégrale de la densité de probabilité du problème (voir figure 1). On peut généraliser ce raisonnement pour différentes valeurs de  $\alpha$  (on a alors des proportions différentes de 80-20).

FIGURE 1 – Densité de loi de Pareto schématisée



3. A quelle condition la variable  $X$  possède-t-elle une espérance? une variance? On se placera sous cette dernière condition dans la suite.

—  $X$  admet une espérance si :

$$\begin{aligned} & \int_{-\infty}^{+\infty} |x|f(x)dx \text{ converge} \\ \Leftrightarrow & \int_{\theta}^{+\infty} |x|^{\frac{\alpha}{\theta}} \left(\frac{\theta}{x}\right)^{\alpha+1} dx = \theta^{\alpha} \alpha \int_{\theta}^{+\infty} \frac{1}{x^{\alpha}} dx \quad , (\theta > 0 \Leftrightarrow |x| = x) \text{ converge} \\ \Leftrightarrow & \alpha > 1 \text{ (Intégrale de Riemann)} \end{aligned}$$

$X$  admet une espérance si  $\alpha > 1$  et cette espérance vaut :

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

— X admet une variance si X admet une espérance et si :

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx \text{ converge} \\
 \Leftrightarrow & \int_{\theta}^{+\infty} x^2 f(x) dx + E(X)^2 \int_{\theta}^{+\infty} f(x) dx - 2E(X) \underbrace{\int_{\theta}^{+\infty} x f(x) dx}_{E(X)} = \int_{\theta}^{+\infty} x^2 f(x) dx - E(X)^2 \text{ converge} \\
 \Leftrightarrow & \int_{\theta}^{+\infty} x^2 f(x) dx \text{ converge} \\
 \Leftrightarrow & \theta^{\alpha} \alpha \int_{\theta}^{+\infty} \frac{1}{x^{\alpha-1}} \text{ converge} \\
 \Leftrightarrow & \alpha - 1 > 1 \text{ (Intégrale de Riemman)} \\
 \Leftrightarrow & \alpha > 2
 \end{aligned}$$

X admet une variance si  $\alpha > 2$  et cette variance vaut :

$$V(X) = \int_{-\infty}^{+\infty} (x - E(x))^2 f(x) dx = E(X^2) - E(X)^2$$

4. Calculer l'espérance et la variance de X.

— Calcul de l'espérance :

$$\begin{aligned}
 E(X) &= \theta^{\alpha} \alpha \int_{\theta}^{+\infty} \frac{1}{x^{\alpha}} dx \\
 &= \theta^{\alpha} \alpha \left[ \frac{-1}{(\alpha-1)x^{\alpha-1}} \right]_{\theta}^{+\infty} \\
 &= \theta^{\alpha} \alpha \frac{1}{(\alpha-1)\theta^{\alpha-1}} \\
 &= \frac{\theta \alpha}{(\alpha-1)}
 \end{aligned}$$

— Calcul de la variance :

$$\begin{aligned}
 V(x) &= E(X^2) - E(X)^2 \\
 E(X^2) &= \frac{\theta^2 \alpha}{(\alpha-2)} \\
 V(X) &= \frac{\theta^2 \alpha}{(\alpha-2)} - \frac{\theta^2 \alpha^2}{(\alpha-1)^2} \\
 V(X) &= \frac{\theta^2 \alpha (\alpha-1)^2 - \theta^2 \alpha^2 (\alpha-2)}{(\alpha-1)^2 (\alpha-2)} \\
 V(X) &= \frac{\alpha \theta^2}{(\alpha-1)^2 (\alpha-2)}
 \end{aligned}$$

5. Calculer la fonction de répartition F de X.

La fonction de répartition, notée  $F_X(x)$  se calcule suivant la formule suivante :

$$\begin{aligned}
 F_X(x) &= P(X \leq x) = \int_{-\infty}^x f(t) dt \\
 &= \begin{cases} 0 & \text{si } x < \theta \\ \int_{\theta}^x \frac{\alpha}{t} \left(\frac{\theta}{t}\right)^{\alpha+1} dt & \text{si } x \geq \theta \end{cases} \\
 &= \begin{cases} 0 & \text{si } x < \theta \\ \alpha \theta^{\alpha} \left[ \frac{-1}{\alpha t^{\alpha}} \right]_{\theta}^x = \frac{\alpha \theta^{\alpha}}{\alpha \theta^{\alpha}} - \frac{\alpha \theta^{\alpha}}{\alpha x^{\alpha}} & \text{si } x \geq \theta \end{cases} \\
 &= \begin{cases} 0 & \text{si } x < \theta \\ 1 - \left(\frac{\theta}{x}\right)^{\alpha} & \text{si } x \geq \theta \end{cases}
 \end{aligned}$$

6. Déterminer la fonction réciproque de F.

$$\begin{aligned}
 y &= 1 - \left(\frac{\theta}{x}\right)^{\alpha} \\
 \left(\frac{\theta}{x}\right)^{\alpha} &= 1 - y \\
 x^{\alpha} &= \frac{\theta^{\alpha}}{1 - y} \\
 x &= \frac{\theta}{(1 - y)^{1/\alpha}}
 \end{aligned}$$

On a donc :

$$F^{-1}(x) = \frac{\theta}{(1 - x)^{1/\alpha}}$$

## 2.2 Simulation d'une valeur d'une variable de loi de Pareto

L'objet de ce paragraphe est de proposer une méthode de construction d'une valeur d'une variable aléatoire de loi  $P(\alpha, \theta)$ .

7. Expliquer de manière formelle comment simuler une valeur d'une variable aléatoire de loi  $P(\alpha, \theta)$ .

On va appliquer la méthode d'inversion.

Soit X une variable réelle de loi continue et strictement croissante. Alors si U est de loi uniforme sur  $[0,1]$ , la variable  $F^{-1}(U)$  a la même loi que X.

Dans le cadre de la loi de Pareto :

$$\begin{cases} F(\theta^-) = 0 \\ F(\theta) = 1 - \left(\frac{\theta}{\theta}\right)^{\alpha} = 0 \end{cases}$$

Donc  $F$  est continue en  $\theta$ , de plus  $F$  est continue de manière triviale sur  $\mathbb{R} \setminus \{\theta\}$ , donc  $F$  est continue sur  $\mathbb{R}$ .

$F$  est également strictement croissante sur  $[\theta, +\infty[$  de manière évidente.

On peut donc appliquer la méthode d'inversion à la loi de Pareto pour en simuler une valeur  $x$  en simulant une réalisation  $u$  de  $U$  et en posant  $x = F^{-1}(u) = \frac{\theta}{(1-u)^{1/\alpha}}$ .

## 2.3 Étude statistique

On considère  $n$  données réelles  $x_1, x_2, \dots, x_n$ . On fait l'hypothèse que ces données sont les réalisations de  $n$  variables aléatoires  $X_1, X_2, \dots, X_n$ , indépendantes et de même loi  $P(\alpha, \theta = 1)$ . On suppose que le paramètre  $\alpha$  est inconnu et on s'intéresse au problème de son estimation par la méthode du maximum de vraisemblance.

8. Ecrire la log-vraisemblance de l'échantillon et trouver l'estimation du maximum de vraisemblance de  $\alpha$ . En déduire alors l'estimateur du maximum de vraisemblance du paramètre  $\alpha$ . On le notera  $\hat{\alpha}_n$ .

$$\begin{aligned} L(x, \alpha) &= \prod_{j=1}^n \alpha \left(\frac{1}{x_j}\right)^{\alpha+1} \\ \log(L(x, \alpha)) &= n \ln(\alpha) - \sum_{j=1}^n (\alpha + 1) \ln(x_j) \\ &= n \ln(\alpha) - (\alpha + 1) s_n \text{ avec } s_n = \sum_{j=1}^n \ln(x_j) \\ t_n &= \arg \max_{z \in [1, +\infty]} (LL(x, z)) \\ g(z) &= n \ln(z) - (z + 1) s_n \\ g'(z) &= \frac{n}{z} - s_n \\ g''(z) &= \frac{-n}{z^2} \end{aligned}$$

Donc  $g'(z) = 0 \Leftrightarrow z = \frac{n}{s_n}$  et  $g''(z) < 0$ . L'unique valeur  $t_n$  qui maximise la vraisemblance sachant  $x_1, x_2, \dots, x_n$  est donc égale à  $t_n = \frac{n}{s_n}$ .

On peut alors construire un estimateur du maximum de vraisemblance de  $\alpha$  noté  $\hat{\alpha}_n : \hat{\alpha}_n = \frac{n}{\sum_{j=1}^n \log(X_j)}$ .

9. Calculer  $E[\log(X_1)]$  et  $Var(\log(X_1))$ .



$$\begin{aligned}
E(\log(X_1)) &= \int_1^{+\infty} \alpha \ln(x) \left(\frac{1}{x}\right)^{\alpha+1} dx \\
&= \alpha \int_1^{+\infty} \ln(x) \left(\frac{1}{x}\right)^{\alpha+1} dx
\end{aligned}$$

On va procéder par intégration par parties :  $u(x) = \ln(x)$ ,  $u'(x) = \frac{1}{x}$ ,  $v'(x) = \frac{1}{x^{\alpha+1}}$ ,  $v(x) = \frac{-1}{\alpha x^{\alpha}}$

$$\begin{aligned}
\int_1^{+\infty} \ln(x) \left(\frac{1}{x}\right)^{\alpha+1} dx &= \left[ \frac{-\ln(x)}{\alpha x^{\alpha}} \right]_1^{+\infty} + \int_1^{+\infty} \frac{1}{\alpha x^{\alpha+1}} dx \\
&= 0 + \left[ \frac{-1}{\alpha^2 x^{\alpha}} \right]_1^{+\infty} = \frac{1}{\alpha^2}
\end{aligned}$$

$$\text{donc } E(\log(X_1)) = \frac{1}{\alpha}$$

Calcul de  $Var(\log(X_1))$  :

$$\begin{aligned}
Var(\log(X_1)) &= E((\log(X_1))^2) - E(\log(X_1))^2 \\
E((\log(X_1))^2) &= \alpha \int_1^{+\infty} \ln^2(x) \left(\frac{1}{x}\right)^{\alpha+1} dx \\
&= \alpha \left( \underbrace{\left[ \frac{-\ln^2(x)}{\alpha x^{\alpha}} \right]_1^{+\infty}}_0 + \int_1^{+\infty} \frac{2\ln(x)}{\alpha x^{\alpha+1}} dx \right) \\
&= \frac{2}{\alpha^2} \\
Var(\log(X_1)) &= \frac{1}{\alpha^2}
\end{aligned}$$

10. Proposer un estimateur sans biais et convergent du paramètre  $\beta = \frac{1}{\alpha}$ . On notera cet estimateur  $B_n$  et  $b_n$  sa réalisation.

$$B_n = \frac{1}{n} \sum_{i=1}^{+\infty} \log(X_i)$$

$B_n$  est un estimateur du paramètre  $\beta = \frac{1}{\alpha}$  sans biais et convergent. En effet,  $E(\log(X_1)) = \frac{1}{\alpha}$  donc par linéarité de l'espérance, on a

$$E(B_n) = \frac{1}{n} \sum_{i=1}^n E(\log(X_i)) = \beta$$

De même, puisque les variables  $X_i$  indépendantes, on a la linéarité de la variance de leur somme, donc

$$Var(B_n) = \frac{n}{n^2} \beta^2 \rightarrow 0$$

11. Enoncer le TLC que satisfait cet estimateur.

$E(\log(X_1)^2) < \infty$  car la variance de  $\log(X_i)$  existe. Cette variable admet donc un momet d'ordre 2.  $B_n$  vérifie donc le Théorème limite centrale suivant :

$$\underbrace{\sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n \log(X_i) - E(\log(X_i)) \right)}_{B_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, Var(\log(X_i)))$$

$$\sqrt{n}(B_n - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \beta^2)$$

12. Proposer un estimateur du paramètre  $\alpha$ . On le notera  $A_n$ .

$$A_n = B_n^{-1} = \frac{n}{\sum_{i=1}^n \log(X_i)} = \widehat{\alpha_n}$$

13. En déduire la convergence de  $A_n$  vers  $\alpha$  et énoncer le TLC que satisfait  $A_n$ .

D'après la loi forte des grands nombres, on a :

$$B_n \xrightarrow[n \rightarrow \infty]{p.s} \beta$$

$$\text{donc } \frac{1}{B_n} = A_n \xrightarrow[n \rightarrow \infty]{p.s} \frac{1}{\beta} = \alpha$$

$A_n$  converge donc bien vers  $\alpha$ .

On a :

$$\sqrt{n}(B_n - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \beta^2)$$

On utilise la delta-méthode en utilisant  $g(x) = \frac{1}{x}$ , dérivable en  $\alpha$  de manière évidente ( $\alpha > 0$ )

$$\sqrt{n} \left( \frac{1}{B_n} - \frac{1}{\beta} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \left(-\frac{1}{\beta^2}\right)^2 \beta^2)$$

On obtient donc finalement :

$$\sqrt{n}(A_n - \alpha) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \alpha^2)$$

## 2.4 Récapitulatif de la partie mathématique

Fonction de densité	$f(x)$	$\begin{cases} \frac{\alpha}{\theta} (\frac{\theta}{x})^{\alpha+1} & \text{si } x \geq \theta \\ 0 & \text{sinon} \end{cases}$
Fonction de répartition	$F(x)$	$\begin{cases} 0 & \text{si } x < \theta \\ 1 - (\frac{\theta}{x})^\alpha & \text{si } x \geq \theta \end{cases}$
Fonction inverse de la fonction de répartition	$F^{-1}(x)$	$F^{-1}(x) = \frac{\theta}{(1-x)^{1/\alpha}}$
Esperance	$E(X)$	$\frac{\theta\alpha}{(\alpha-1)} \quad (\alpha > 1)$
Variance	$V(x)$	$\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)} \quad (\alpha > 2)$
Estimateur du paramètre $\alpha$	$A_n$	$\frac{n}{\sum_{i=1}^n \log(X_i)}$

### 3 Résultats numériques

Nous avons implémenté la méthode d'inversion pour la simulation et l'estimation par maximum de vraisemblance sous R.

#### 3.1 Simulation de la loi de Pareto

Nous avons simulé un échantillon de 400 réalisations de la loi de Pareto  $\mathcal{P}(\alpha, \theta = 1)$  par la méthode de l'inversion, puis affiché les fréquences des valeurs obtenues sous la forme d'un histogramme de 20 classes. Nous avons ensuite superposé la courbe représentative de la densité de la loi  $\mathcal{P}(\alpha, \theta = 1)$  sur le graphique.

Voici l'un des graphiques que nous avons obtenus (figures 2 et 3, pour  $\alpha = 2$ ).

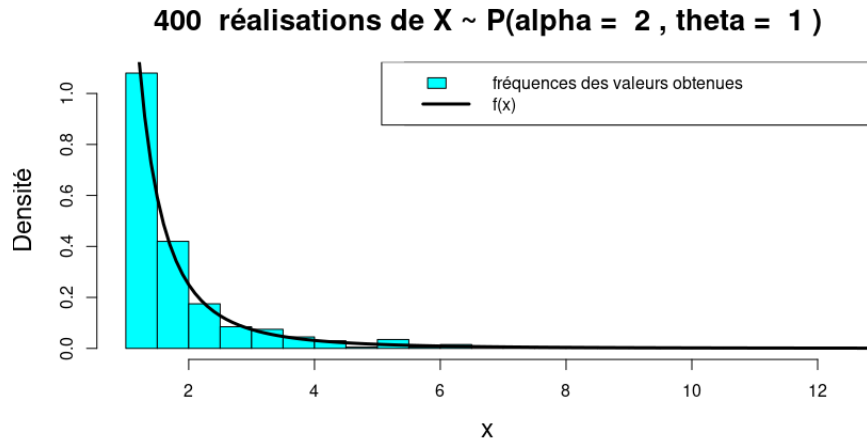


FIGURE 2 – Affichage de l’histogramme et de la courbe de densité de  $\mathcal{P}(\alpha, \theta = 1)$

```
# mise à jour de la graine de génération aléatoire
RNGkind()
set.seed(222)
sample(1:10, 5, replace=T)

# Taille de l'échantillon simulé
n = 400

# Paramètres de la loi de Pareto simulée
theta = 1
alpha = 10

# Simulation de l'échantillon X
U<-runif(n)
X<-(theta/((1-U)^(1/alpha)))

# Tracé de l'histogramme de fréquences des valeurs de l'échantillon
# (les fréquences de ces valeurs modélisent la densité de probabilité de la loi de X)
titre = paste(as.character(n)," réalisations de X ~ P(alpha = ", alpha, ", theta = ", theta, ")")
hist(X, proba=TRUE, col="cyan", cex.lab=1.5,
     cex.main=1.7, main=titre, nclass=20, xlab="x", ylab="Densité")

# Tracé de la densité de probabilité exacte de la loi
t = seq(0, 80, 0.1)
lines(t, (alpha/theta)*(theta/t)^(alpha+1), lwd=3)
```

FIGURE 3 – Code R permettant de générer les simulations de  $\mathcal{P}(\alpha, \theta = 1)$

Les résultats graphiques confirment que la simulation par la méthode de l'inversion fonctionne correctement car l'histogramme suit la courbe de la densité de  $\mathcal{P}(\alpha, \theta = 1)$ . De plus, la valeur de  $\alpha$  choisie a une influence sur la magnitude des valeurs obtenues, mais pas sur leur répartition (voir annexe figures 9 et 10). On remarque toutefois que puisque les valeurs simulées le sont de manière aléatoire (car la variable  $X$  est aléatoire), on obtient un histogramme différent si on change la graine de génération aléatoire, ce qui illustre le phénomène de

la fluctuation d'échantillonnage (voir annexe figure 8).

## **3.2 Estimation du paramètre $\alpha$**

### **3.2.1 Graphique d'évolution de l'estimateur**

Tout d'abord, nous avons simulé un échantillon de 400 réalisations de  $X \sim \mathcal{P}(\alpha, \theta = 1)$ . Ensuite, nous avons utilisé cet échantillon pour calculer les valeurs de l'estimation de  $\alpha$  par maximum de vraisemblance  $A_n$  pour  $n$  allant de 1 à 40, puis pour  $n$  allant de 1 à 400.

Nous avons obtenu ces graphiques pour l'une des exécutions (figures 4 et 5, pour  $\alpha = 2$ ).

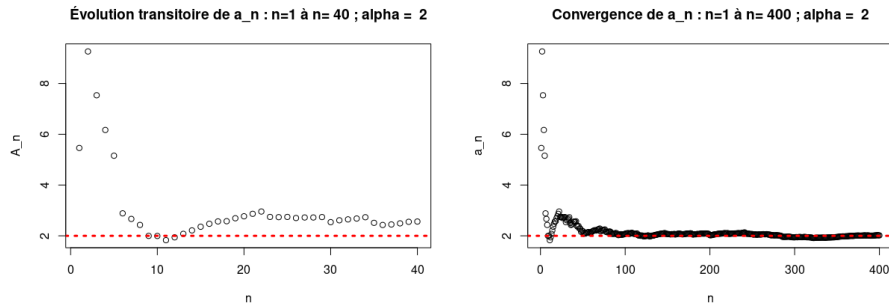


FIGURE 4 – Affichage des courbes d'évolution de  $a_n$

```
# Évolution de l'estimation de alpha par la méthode du maximum de vraisemblance
# en fonction de l'évolution de la taille de l'échantillon

# Graphique démontrant l'évolution transitoire de A_n
transi=40
A<-matrix(ncol=1,nrow=transi) # Estimateur A_n
s=0
# réalisation d'estimations sur des échantillons de taille grandissante
for(i in 1:transi){
  s=s+log(X[i])
  A[i]=i/s
}
titre = paste("Évolution transitoire de a_n : n=1 à n=",as.character(transi)," ; alpha = ",alpha)
plot(A, main=titre, xlab="n", ylab="A_n")

# Graphique démontrant la convergence de a_n
A<-matrix(ncol=1,nrow=n)
s=0
# réalisation d'estimations sur des échantillons de taille grandissante
for(i in 1:n){
  s=s+log(X[i])
  A[i]=i/s
}
titre = paste("Convergence de a_n : n=1 à n=", as.character(n), " ; alpha = ", alpha)
plot(A, main=titre, xlab="n", ylab="a_n")
```

FIGURE 5 – Code R permettant de générer les valeurs de  $a_n$

On peut voir que l'estimateur est bien convergent car au fur et à mesure que  $n$  augmente,  $a_n$  se rapproche asymptotiquement de  $\alpha$  (ici 2). Toutefois, pour des petites valeurs de  $n$ , on peut voir un manque de performance de l'estimateur, ce qui illustre son caractère transitoire : il faut que  $n$  dépasse un certain "seuil" (qui est empirique, on pourrait dire  $n = 50$ ) pour que l'estimateur devienne performant. Ici encore, selon les exécutions, en changeant la graine de génération aléatoire, on obtient des résultats différents (la fluctuation d'échantillonnage) mais tous nous permettent d'obtenir cette même conclusion (voir annexe figure 11), et la valeur d' $\alpha$  n'importe pas (voir annexe figures 11 et 12).

### 3.2.2 Boîtes à moustaches

Tout d'abord, nous avons simulé 400 échantillons de 1000 simulations de  $X \sim \mathcal{P}(\alpha, \theta = 1)$ . Nous avons ensuite calculé les valeurs de  $a_n$  pour  $n$  allant de 1 à 1000 pour chacun des 400 échantillons (nous avons appelé "échantillon d'estimations de  $\alpha$ " l'ensemble des 400 valeurs  $a_n$  générées pour un  $n$  donné). Enfin, nous avons représenté les échantillons d'estimations par des boîtes à moustaches pour  $n$  prenant certaines valeurs arbitraires (25, 50, 100, 200, 500, et 1000).

Voici les boîtes à moustaches que nous avons obtenues pour l'une des exécutions (voir figures 6 et 7, pour  $\alpha = 2$ ).



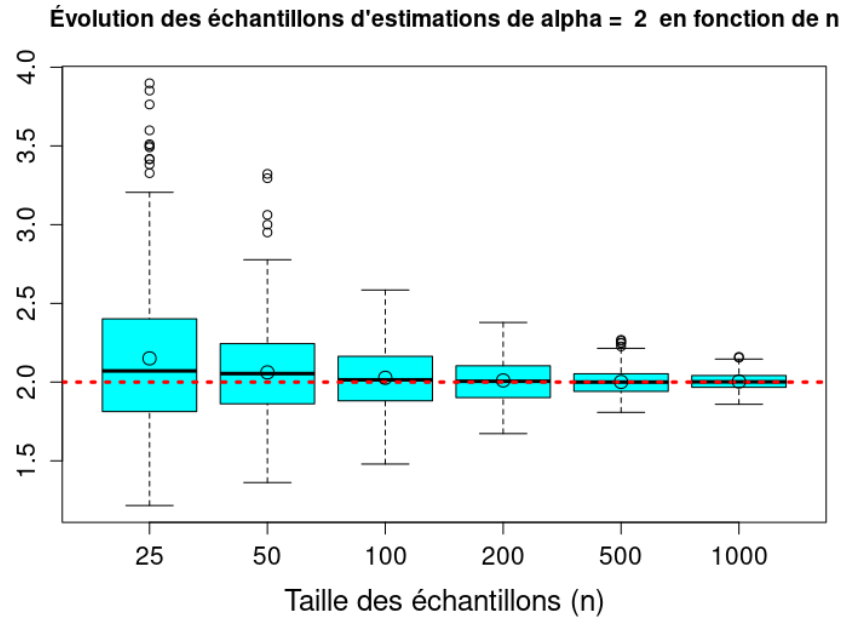


FIGURE 6 – Représentation des échantillons d'estimations de  $\alpha$  par des boîtes à moustaches

```
# 400 échantillons de taille 1000 de X
Data<-matrix(ncol=400,nrow=1000)
U<-runif(400000)
Data[,]<-(theta/((1-U)^(1/alpha)))

# 400 estimations de alpha par le maximum de vraisemblance sur les 400 échantillons
# (on calcule l'estimation pour les tailles d'échantillon de 1 à 1000)
Data2<-matrix(ncol=400,nrow=1000)
for (j in 1:400){
  s=0
  for(i in 1:1000){
    s=s+log(Data[i,j])
    Data2[i,j]=1/s
  }
}

# Affichage de la boîte à moustaches des 400 estimations pour différentes tailles d'échantillon
val_n = c(25,50,100,200,500,1000)
titre = paste("Évolution des échantillons d'estimations de alpha = ",alpha," en fonction de n")
boxplot(t(Data2)[,val_n],col = "cyan",names=c("25","50","100","200","500","1000"),
        cex.axis=1.3,cex.lab = 1.5, xlab="Taille des échantillons (n)",
        outline = TRUE, main=titre)
abline(h = alpha, col="red", lwd=3, lty=3)

# Affichage des moyennes des 400 estimations sur chaque boîte à moustaches
points(apply(t(Data2)[,val_n],2,mean),cex=1.5)
```

FIGURE 7 – Code R permettant de générer les valeurs de  $a_n$

On observe le caractère sans biais de l'estimateur  $a_n$ , car même pour des valeurs de  $n$  "petites", la moyenne des 400 estimations calculées est proche de  $\alpha$ . De plus, on observe la convergence de l'estimateur car lorsque  $n$  augmente, la taille des boîtes à moustaches diminue, ce qui illustre une baisse de l'influence des fluctuations d'échantillonnage sur les résultats obtenus lorsque  $n$  augmente, ce qui nous confirme que la variance de l'estimateur tend vers 0 lorsque  $n$  augmente. Les résultats obtenus lors d'exécutions différentes (à cause des fluctuations d'échantillonnage, voir annexe figure 13) nous permettent de tirer des conclusions similaires, et la valeur de  $\alpha$  n'a pas d'influence sur nos conclusions (voir annexe figure 14).

## 4 Conclusion

Après avoir étudié les caractéristiques théoriques d'une variable aléatoire suivant une loi de Pareto, nous avons donc pu simuler une telle variable aléatoire, et estimer son paramètre  $\alpha$  par la méthode du maximum de vraisemblance. Nous avons constaté une variation des résultats obtenus selon les exécutions effectuées causée par la fluctuation d'échantillonnage (que ce soit pour la simulation ou pour l'estimation de  $\alpha$ ). Mais malgré ces fluctuations, nous avons constaté par l'expérience d'une part la performance de la méthode de simulation par inversion d'une variable aléatoire de loi de Pareto, et d'autre part le caractère sans biais et la convergence de l'estimateur de maximum de vraisemblance du paramètre  $\alpha$ .

## 5 Bibliographie

- Cours magistral, travaux dirigés, et travaux pratiques encadrés de Statistique Inférentielle de GM3, Pr. Bruno PORTIER, 2022

## 6 Annexe

### 6.1 Figures

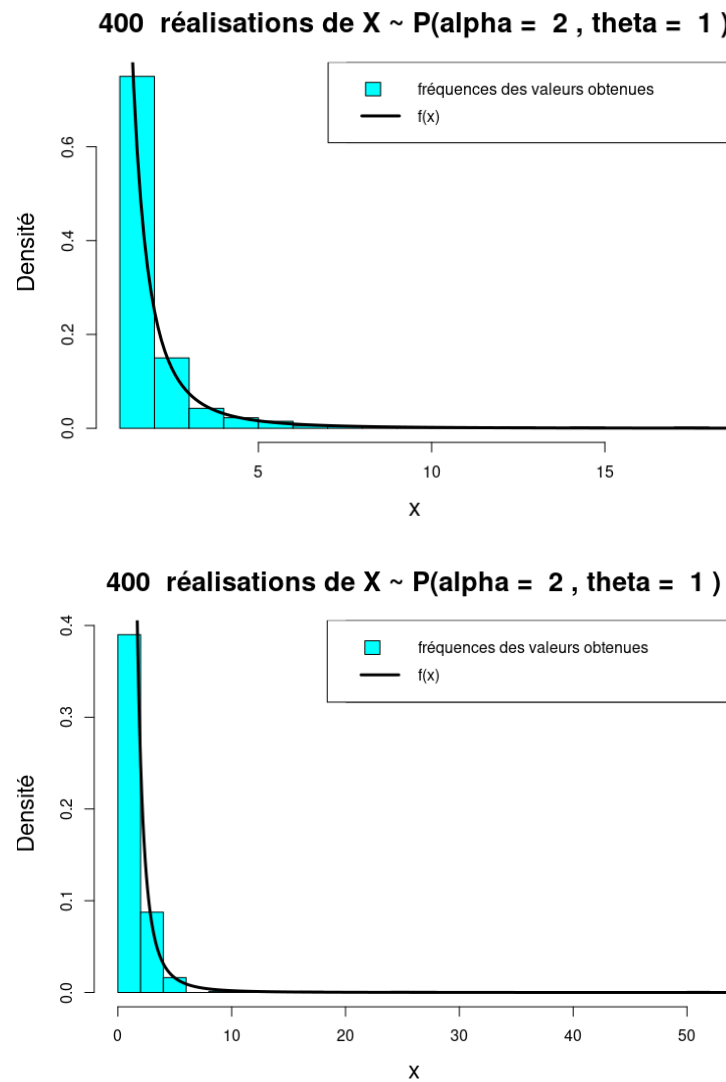


FIGURE 8 – D'autres simulations d'une variable  $X \sim \mathcal{P}(\alpha = 2, \theta = 1)$

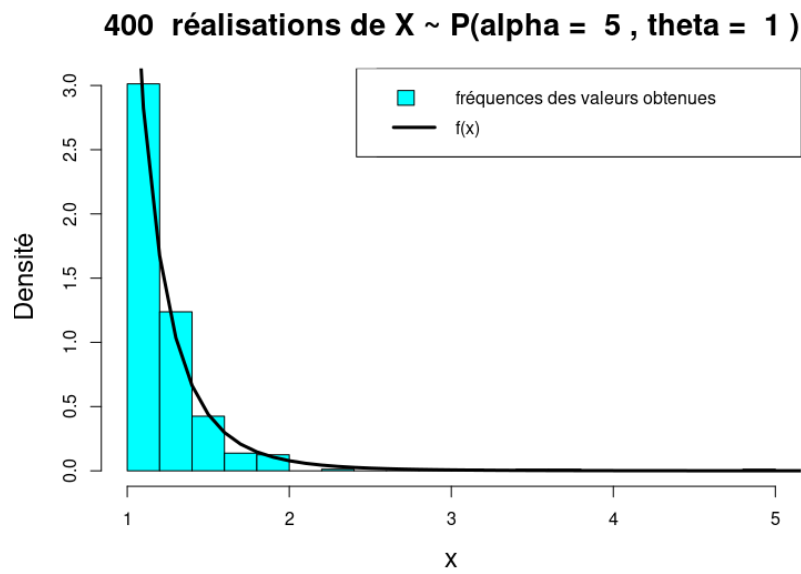
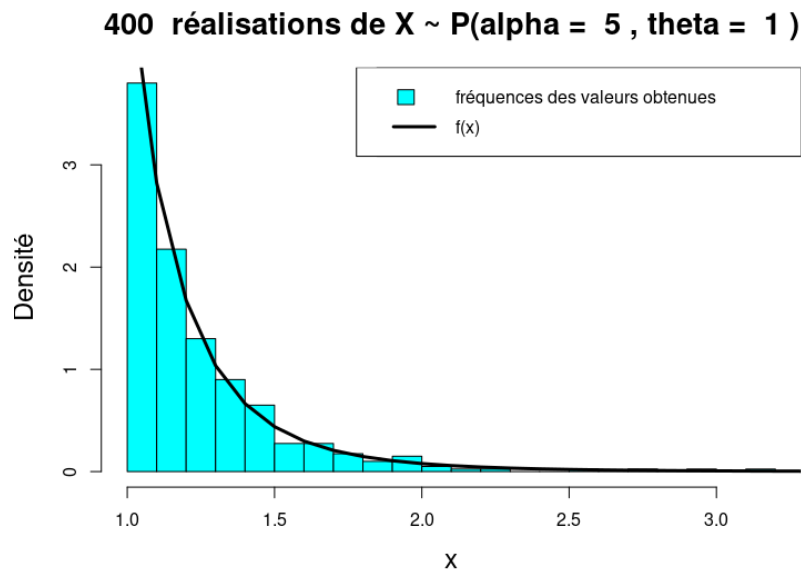
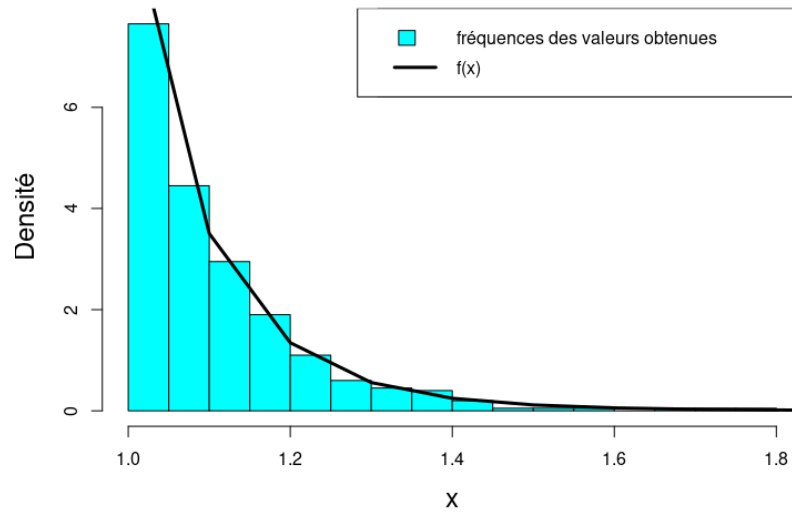


FIGURE 9 – D'autres simulations d'une variable  $X \sim \mathcal{P}(\alpha = 5, \theta = 1)$

400 réalisations de  $X \sim \mathcal{P}(\alpha = 10, \theta = 1)$



400 réalisations de  $X \sim \mathcal{P}(\alpha = 10, \theta = 1)$

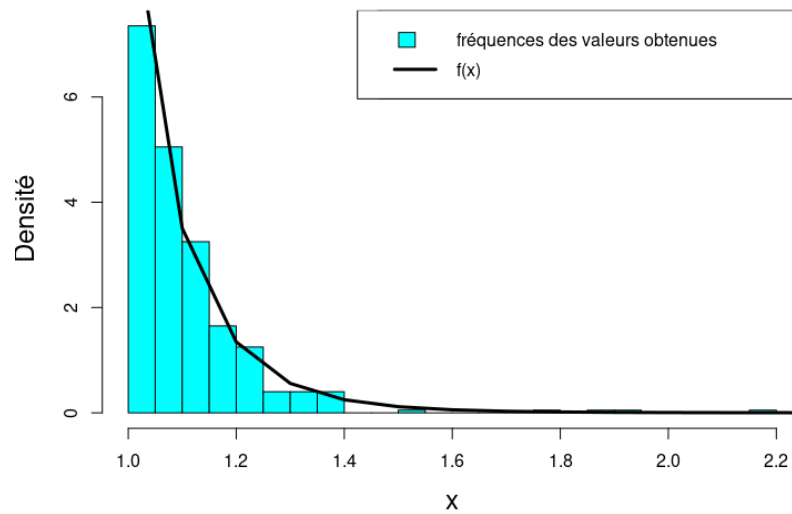


FIGURE 10 – D'autres simulations d'une variable  $X \sim \mathcal{P}(\alpha = 10, \theta = 1)$

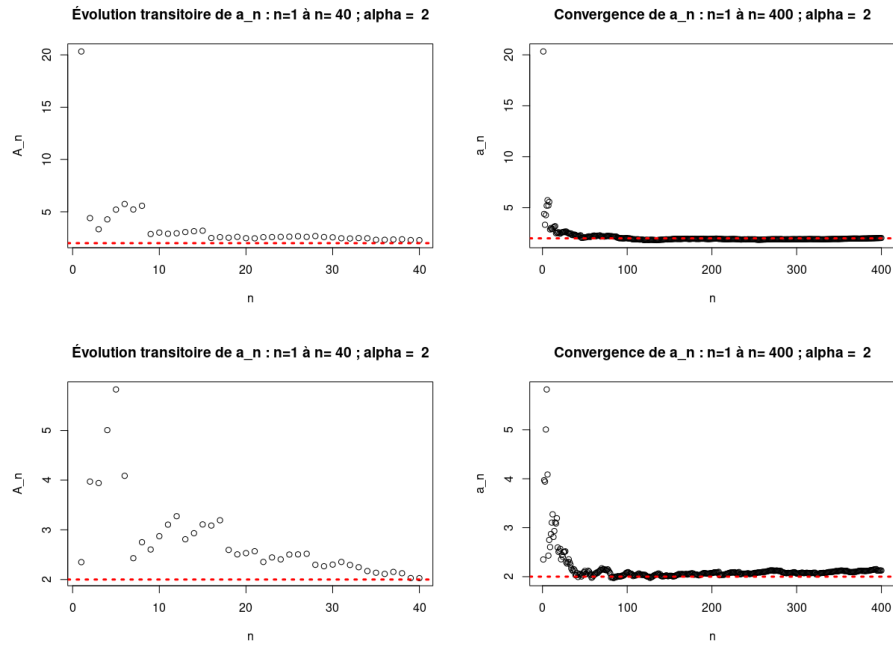


FIGURE 11 – Affichage des courbes d'évolution de  $a_n$  lors d'exécutions différentes pour  $\alpha = 2$

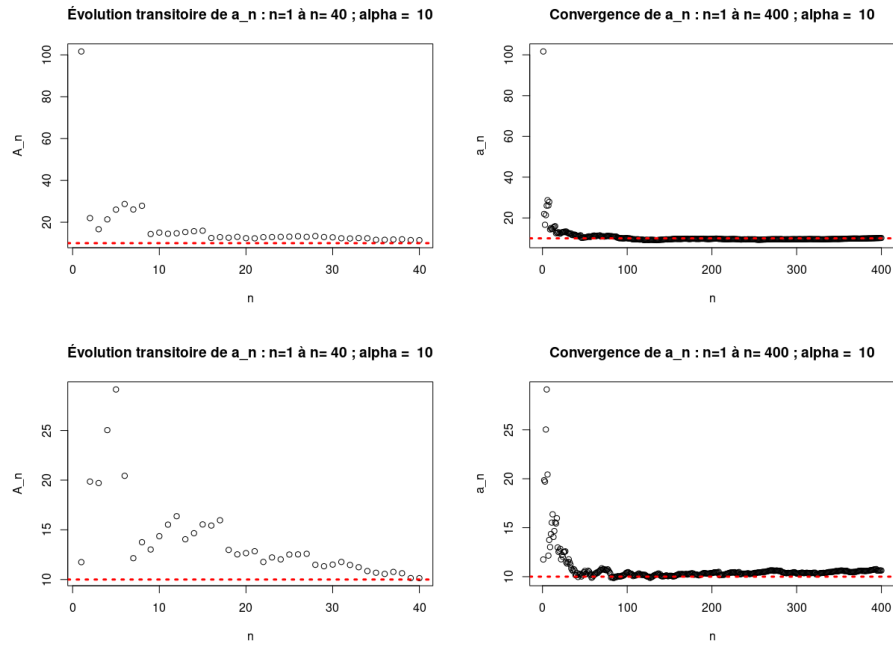
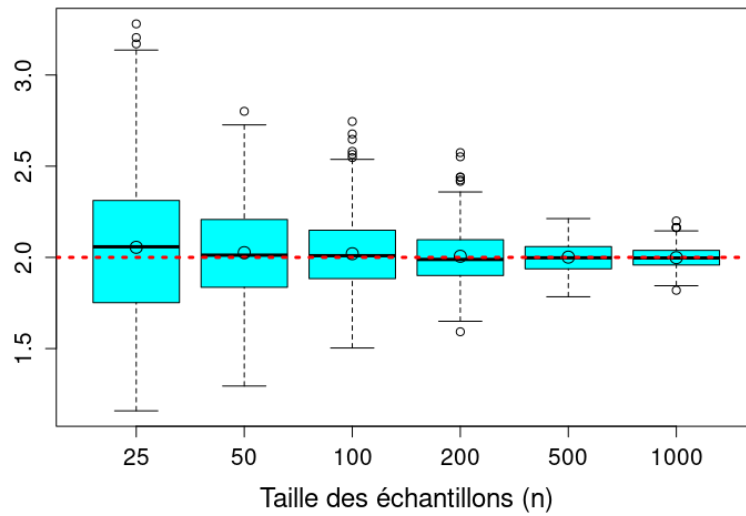


FIGURE 12 – Affichage des courbes d'évolution de  $a_n$  lors d'exécutions différentes pour  $\alpha = 10$



Évolution des échantillons d'estimations de  $\alpha = 2$  en fonction de  $n$



Évolution des échantillons d'estimations de  $\alpha = 2$  en fonction de  $n$

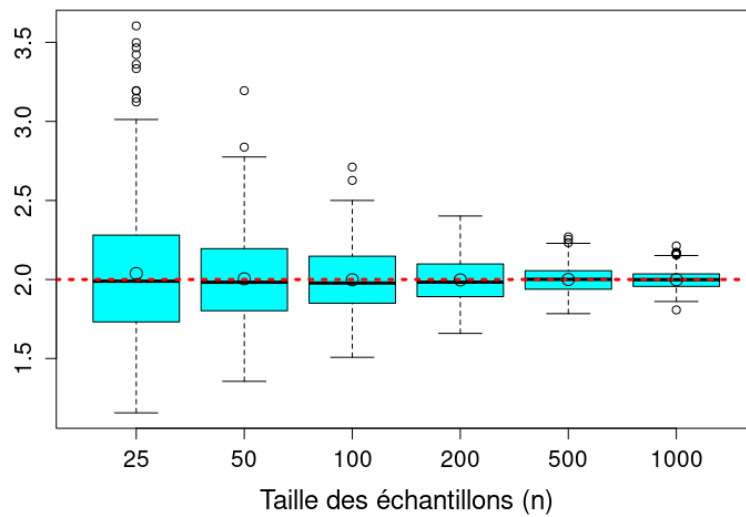
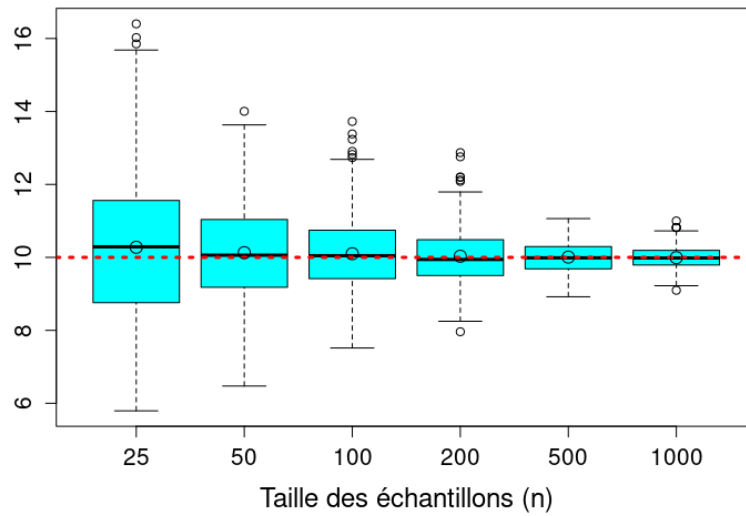


FIGURE 13 – Représentation d'autres échantillons d'estimations de  $\alpha = 2$  par des boîtes à moustaches

Évolution des échantillons d'estimations de  $\alpha = 10$  en fonction de  $n$



Évolution des échantillons d'estimations de  $\alpha = 10$  en fonction de  $n$

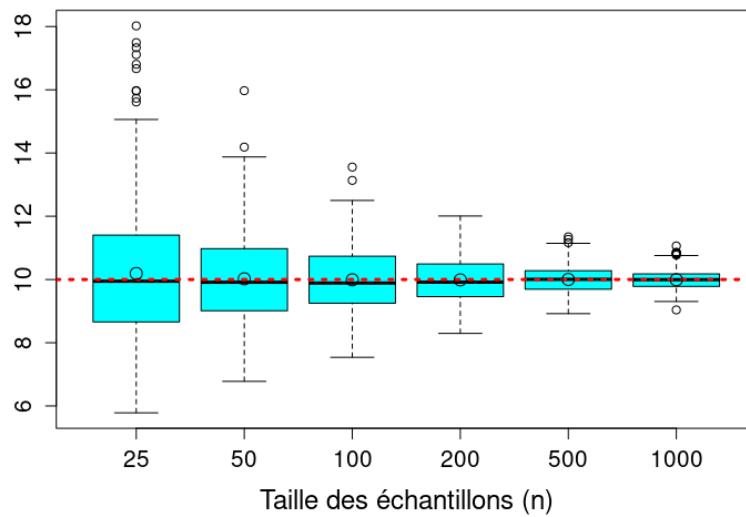


FIGURE 14 – Représentation d'autres échantillons d'estimations de  $\alpha = 10$  par des boîtes à moustaches

## 6.2 Listing du programme R

- `RNGkind()`, `set.seed(int)` : initialisation de la graine de génération aléatoire de nombres
- `runif( $n$ )` : échantillon de  $n$  simulations de la loi uniforme sur  $[0, 1]$
- `paste(chaines)` : concaténation de chaînes de caractères
- `hist()` : construction d'un histogramme
- `seq(deb,fin,pas)` : génération d'un tableau de nombres de "deb" à "fin" espacés de "pas"
- `lines( $t, f(t)$ )` : tracé des valeurs  $f(t)$  sur les points  $t$  ( $t$  est un tableau d'abscisses)
- `legend()` : génération d'une légende sur le graphique obtenu
- `matrix()` : création d'une matrice
- `plot( $A$ )` : affichage des valeurs contenues dans la matrice  $A$  sous la forme d'un graphique de points
- `abline()` : tracé d'une ligne sur le graphique
- `boxplot()` : affichage d'un histogramme
- `points()` : affichage de points sur un graphique