Cours 2 – Simulation d'échantillons.

Eya ZOUGAR * Institut National des Sciences appliqués-INSA

Génie mathématiques GM3 Thursday 26th January, 2023



^{*}Basé sur le cours de Bruno PORTIFR

Contents

Introduction

2 Variables discrétes

Variables continues

1. Introduction 1.1. L'objectif du cours

Dans ce cours, on va s'intéresser au problème de la simulation d'une réalisation d'une variable aléatoire X qui suit une loi donnée.

On va, alors, présenter un certain nombre de méthodes qui permettront de répondre à ce problème.

Ce cours a été construit à partir de plusieurs supports trouvés sur le net, et en particulier d'un chapitre du livre [1].

[†]B. Bercu et D. Chafaï, Modélisation stochastique et simulations, (2007).

1.2. La brique de base.

La brique de base pour résoudre la plupart des problèmes auxquels nous allons nous intéresser, est le générateur aléatoire de réalisations d'une loi uniforme sur [0, 1].

Avec le logiciel R, on utilisera la fonction runif.

En général, ce générateur n'est que "pseudo-aléatoire" au sens où il est construit sur un algorithme déterministe. Mais on supposera qu'il est parfait, et que différents appels à cet algorithme fournissent des réalisations indépendantes d'une variable aléatoire de loi uniforme sur $[0\,,\,1]$.

A partir de cette brique de base, on peut simuler toutes les lois, par des considérations assez simples.

2. Simulation de variables discrètes. 2.1. Introduction.

On s'intéresse dans cette partie à la simulation de réalisations de quelques variables aléatoires discrètes.

Nous allons considérer le cas de : \square la loi de Bernouilli $\mathcal{B}(p)$ le cas d'une loi discrète quelconque à nombre de valeurs fini \square la loi binômiale $\mathcal{B}(n,p)$ \square la loi géométrique $\mathcal{G}(p)$

Dans la suite, on désignera par U une variable aléatoire de loi uniforme[‡] sur [0, 1].

$$U \hookrightarrow \mathcal{U}([0,1])$$

 $^{^{\}ddagger}$ La fct de densité associée à U est: $f_{\mathcal{U}}(t) = \mathbf{1}_{[0,1]}(t)$ E. ZOUGAR [INSA] Génie mathématiques GM3

2.2. Loi de Bernouilli.

Considérons une variable aléatoire $X \hookrightarrow \mathcal{B}(p)$ de loi de Bernouilli de paramètre p avec $p \in]0, 1[$.

ALors,
$$X(\Omega)=\{0,1\}$$
 et $egin{cases} \mathbb{P}\left[X=1
ight]=p \ \mathbb{P}\left[X=0
ight]=1-p. \end{cases}$

Comment simuler une réalisation de X?

C'est très simple avec la remarque suivante.

Pour tout $p \in]0, 1[$, on a

$$\mathbb{P}\left[U \leq p\right] = \int_0^p f_U(t) \, \mathrm{d}t = \int_0^p \mathrm{d}t = p = \mathbb{P}\left[X = 1\right]$$

Ainsi, pour simuler une réalisation x de X, il suffit de simuler une réalisation u de U et de la comparer à p :

- Si u < p alors x = 1
- Sinon x = 0.

Autrement, $X = \mathbf{1}_{\{U \le n\}}$

2.3. Loi discrète à nombre de valeurs finies. 2.3.1. Le problème et le principe.

Soit X une variable discrète à valeurs dans un ensemble fini $X(\Omega) = \{x_1, x_2, \dots, x_K\}$. On note p_1, p_2, \dots, p_K les probabilités associées, c'est à dire que

$$\mathbb{P}\left[X=x_{j}\right] = p_{j} \text{ pour tout } j=1,2,\ldots,K$$

Comment simuler une réalisation x de X?

1.On construit la suite des probabilité cumulées $(P_i)_{0 \le i \le K}$ définies

par :
$$\begin{cases} P_0 = 0. \\ P_j = \sum_{i=1}^j p_i, \quad \forall 1 \leq j \leq k \end{cases}$$

E. ZOUGAR [INSA] Génie mathématiques GM3 2023-01-26

2.3. Loi discrète à nombre de valeurs finies. 2.3.1. Le problème et le principe.

2. Puis en remarquant que si U désigne une variable aléatoire de loi uniforme sur]0, 1[,

$$\mathbb{P}[P_{j-1} \le U < P_j] = \int_{P_{j-1}}^{P_j} dt = P_j - P_{j-1} = p_j = \mathbb{P}[X = x_j]$$

alors, il est très facile de simuler une réalisation x de X. En effet, il suffit de simuler une réalisation u de U, puis de trouver l'indice J tel que $u \in [P_{J-1}, P_J]$ et de poser alors $x = x_J$.

2.3.2. Exemple.

On considère la variable aléatoire discrète X de loi de probabilité

X	1	2	3	4	5
™[∧ – √]	1	1	1	1	1
$\mathbb{P}\left[X=x\right]$	3	4	5	6	20

Table: Loi de probabilité de la variable X.

La suite des probabilités cumulées est égale à :

	P_0	P_1	P_2	P_3	P_4	P_5
Ì	0	<u>20</u>	<u>35</u>	47 60	<u>57</u>	1

Table: Probabilités cumulées de la variable X.

2.3.3. Illustration avec R: le code.

Pour simuler une réalisation de X, on peut utiliser le code suivant :

```
# Instruction pour que tout le monde ait le même aléa de départ
RNGkind(sample.kind = "Rounding")
prob = c(1/3, 1/4, 1/5, 1/6, 1/20), P = cumsum(prob)
pp=[0 prob(1:(length(P)-1))], PPc = cumsum(PP)
U=runif(1);
J=which(cpp<= U)&(U<cp))
x=x[J]</pre>
```

Instructions pour en avoir N:

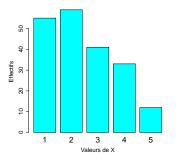
```
N = 200 ; x=rep(NA,N)
for(i in 1:N) x[i] = .....
```

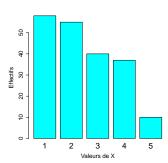
Instructions pour visualiser la répartition des réalisations sous forme d'un diagramme en batons :

```
barplot(table(x), col="cyan", cex = 2, cex.axis = 1.5,
cex.lab=1.5, xlab="Valeurs de X", ylab = "Effectifs")
```

2.3.4. Le résultat graphique.

On trouvera ci-dessous les diagrammes en batons de deux échantillons de 200 réalisations de la variable X.





2.4. Loi binômiale. 2.4.1. Le problème et le principe.

Soit X une variable aléatoire de loi Binômiale $\mathcal{B}(n,p)$, alors

$$\mathbb{P}[X = x] = \mathbf{C}_n^x p^x (1 - p)^{n-x} \quad \forall x \in \{0, 1, 2, \dots, n\}.$$

Comment simuler une réalisation x de X?

Solution. Comme une variable aléatoire x de loi Binômiale $\mathcal{B}(n,p)$ s'écrit sous la forme d'une somme de n variables aléatoires $(Z_i)_{i=1,\dots,n}$, indépendantes et de même loi de Bernouilli de paramètre p, c.à.d:

$$X=\sum_{i=1}^n Z_j,$$

il est très facile de simuler une réalisation x de X.

En effet, il suffit de simuler n réalisations indépendantes d'une variables de Bernouilli de paramètres p et de sommer les résultats.

2.4.2. Illustration avec R: le code.

Avec le logiciel R, on peut simuler une réalisation d'une variable aléatoire de loi (n = 10, p = 0.5) avec les instructions suivantes :

```
n = 10 ; p = 0.5
U = runif(n)
x = sum((U < p))</pre>
```

On peut simuler N = 200 réalisations en utilisant une boucle for ou bien en utilisant les instructions suivantes :

```
U = matrix(runif(n*N), nrow=N, ncol=n)
x = apply(U < p, 1, sum)</pre>
```

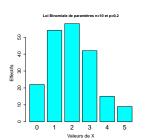
Enfin, on peut visualiser la répartition des réalisations avec l'instruction suivante :

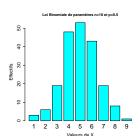
```
barplot(table(x), col="cyan", cex = 2, cex.axis = 1.5,
  cex.lab=1.5, xlab="Valeurs de X", ylab = "Effectifs",
  main=paste0("Loi Binomiale de paramètres n=",n,"
```

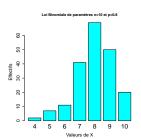
N = 200

2.4.3. Le résultat graphiques.

On trouvera ci-dessous 3 graphiques présentant chacun un diagramme en batons permettant de visualiser les 200 réalisations obtenues dans le cas d'une variable de loi $\mathcal{B}(10,0.2)$ (à gauche), de loi $\mathcal{B}(10,0.5)$ (au centre) et de loi $\mathcal{B}(10,0.8)$ (à droite).







E. ZOUGAR [INSA] Génie mathématiques GM3 2023-01-26

2.5. Loi Géométrique. 2.5.1. Le problème.

Soit X une variable aléatoire de loi géométrique de paramètre p, avec $p \in]0,1[$.

On notera $X \sim \mathcal{G}(p)$. Alors, pour tout entier $n \geq 1$, on a :

$$\mathbb{P}\left[X=n\right] = p(1-p)^{n-1}.$$

On peut montrer que

$$\mathbb{E}\left[X\right] = rac{1}{p}$$
 et $\mathbb{V}\operatorname{ar}(X) = rac{(1-p)}{p^2}$

Comment simuler une réalisation x de X?

2.5.2. Le principe.

On va utiliser le lemme suivant:

Lemma: Si $(Z_n)_{n\geq 1}$ est une suite de variable aléatoires indép. et de même loi de Bernouilli de paramètre p, alors la variable aléatoire T définie par:

$$T = \inf \{ n \geq 1 \text{ tel que } Z_n = 1 \}$$

suit la loi Géométrique de paramètre p.

Proof: On a $T(\Omega) = \mathbb{N}^*$ et pour tout $n \geq 1$

$$\mathbb{P}(T=n) = \mathbb{P}(\forall 1 \le i \le n-1, Z_i = 0 \text{ et } Z_n = 1)
= \prod_{i=1}^{n-1} \mathbb{P}(Z_i = 0).\mathbb{P}(Z_n = 1)
= (1-p)^{n-1}.p$$

Pour simuler une réalisation x d'une variable aléatoire X de loi géométrique de paramètre p, on simule des réalisations indépendantes d'une variable de loi de Bernouilli de paramètre p, jusqu'à obtenir un succès (ou 1), et on comptabilise le nombre n de réalisations nécessaires à l'obtention du 1, puis on pose x = n.

2.5.3. Illustration avec R: le code.

Avec le logiciel R, on simule une réalisation x de $X \sim \mathcal{G}(p)$ avec les instruction suivantes :

On peut simuler N = 200 réalisations en utilisant une boucle for et les

```
instructions suivantes:
N = 200
X = rep(NA,N)
for (j in 1:N){
    x = 1
    while (runif(1) >= p)    x= x + 1
    X[j] = x
```

while (runif(1) > p) x = x + 1

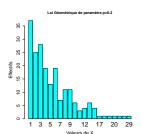
Enfin, on peut visualiser la répartition des réalisations avec l'instruction suivante:

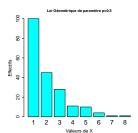
```
barplot(table(X), col="cyan", cex = 2, cex.axis = 1.5, cex.lab=1.5,
    xlab="Valeurs de X", ylab = "Effectifs",
    main=paste0("Loi Géométrique de paramètre p=",p))
```

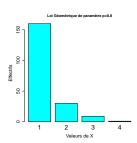
p = 0.5x = 1

2.5.4. Le résultat graphique.

On trouvera ci-dessous 3 graphiques présentant chacun un diagramme en batons permettant de visualiser la répartition des 200 réalisations obtenues dans le cas d'une variable de loi $\mathcal{G}(0.2)$ (à gauche), de loi $\mathcal{G}(0.5)$ (au centre) et de loi $\mathcal{G}(0.8)$ (à droite).







3. Lois continues. 3.1. La méthode d'inversion.

Cette méthode repose sur la proposition suivante.

Proposition. Soit X une variable aléatoire réelle de loi F 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.

Proof: Exercice E.1 (Série 1)

Autrement dit, pour simuler une réalisation x de X, il suffit de simuler une réalisation u de U et de poser $x = F^{-1}(u)$.

Cette méthode permet de simuler des réalisations de n'importe quelle variable aléatoire réelle, à partir du moment où la réciproque de sa fonction de répartition F est explicite.

3.2. Exemple : loi de Cauchy. 3.2.1. Le principe.

Considérons une variable aléatoire X de loi de Cauchy.

Sa densité est définie pour tout réel x par :

$$f(x) = \frac{1}{\pi(x^2+1)}$$

Sa fonction de répartition est définie pour tout $x \in \mathbb{R}$ par :

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$$

Puisque F est strictement croissante sur \mathbb{R} , la fonction F réalise une bijection de \mathbb{R} sur]0, 1[, et admet donc une fonction réciproque, notée F^{-1} et définie pour tout $u \in]0$, 1[par :

$$F^{-1}(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$$

Ainsi, pour simuler une réalisation x de X de loi de Cauchy, on simule une réalisation u de U et on pose $x = F^{-1}(u)$.

3.2.2. Illustration avec le logiciel R.

On construit un échantillon de 200 réalisations indépendantes d'une loi de Cauchy avec les instructions suivantes :

```
n = 200
x = tan(pi * (runif(n) - 0.5))
```

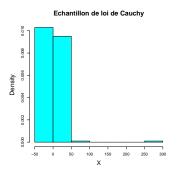
On représente la distribution des données simulées à l'aide d'un histogramme en fréquences.

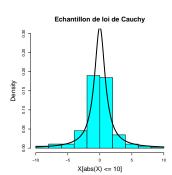
```
t = seq(-10, 10, 0.1)
Titre = "Echantillon de loi de Cauchy"
hist(x, proba=TRUE, col="cyan", cex.lab=1.6,
   cex.main=1.7, main=Titre)
lines(t, 1/(pi * (t^2+1)), lwd =4)
```

3.2.3. Le résultat graphique.

On trouvera ci-dessous les 200 réalisations obtenues, représentées à l'aide d'un histogramme en fréquences sur lequel on a superposé la densité de la loi de Cauchy.

On trouve à gauche, l'ensemble des données et à droite, un zoom sur les données comprises entre -10 et 10.





3.3. Loi uniforme sur [a, b].

On considère une variable aléatoire X de loi uniforme sur [a, b]. Sa densité f est définie par :

$$f(x) = \frac{1}{(b-a)} \mathbf{1}_{[a,b]}(x)$$

Il est facile de montrer à l'aide du théorème d'inversion, que pour simuler une réalisation x de X, il suffit de simuler une réalisation u d'une variable U de loi uniforme sur $[0\,,\,1]$ et de lui appliquer la transformation suivante :

$$x = (b - a)u + a$$

En effet, la fonction de répartition F de X est définie par :

$$F(x) = \frac{x-a}{b-a}$$
 pour tout $x \in [a, b]$

et par conséquent $F^{-1}(u) = (b-a)u + a$.

3.4. Loi normale $\mathcal{N}(\mu, \sigma^2)$.

Soit X une variable aléatoire réelle de loi $\mathcal{N}(0,1)$.

Soit Y une variable aléatoire réelle de loi $\mathcal{N}(\mu, \sigma^2)$, où σ^2 désigne la variance (et donc σ désigne l'écart-type) et μ l'espérance.

Problème. Comment simuler une réalisation de Y à partir d'une réalisation de X?

Solution. C'est très facile, puisque l'on peut exprimer Y en fonction de X :

$$Y = \sigma X + \mu$$

Ainsi, à partir du moment où l'on peut simuler une réalisation x de X, on construit une réalisation y de Y en posant tout simplement $y = \sigma x + \mu$.

Avec le logiciel R, on simule une réalisation de la loi normale avec la fonction rnorm.

4. La méthode du rejet. 4.1. Introduction.

Soit X une variable aléatoire de loi F.

On ne sait pas simuler de manière directe une réalisation x de X.

Comment simuler une réalisation x de X?

Il existe une méthode qui consiste à simuler selon la loi d'une autre variable aléatoire et à accepter une réalisation de cette variable aléatoire comme réalisation de X, lorsqu'une condition n'est pas rejetée.

C'est la méthode dite du rejet.

Cette idée remonte à Von Neu-mann (1947).

Il en existe plusieurs versions, et nous en verrons 2 ici.

4.2. Une première version. 4.2.1. Le problème.

Soient A et B deux boréliens de \mathbb{R}^2 tels que $B \subset A$, de mesure de Lebesgue respectives $\lambda(B)$ et $\lambda(A)$, supposées finies tels que

$$0 < \lambda(B) \le \lambda(A)$$

Soit X une variable aléatoire de loi uniforme sur B. On souhaite simuler une réalisation x de la variable X sachant que l'on sait seulement simuler une réalisation y de la variable Y de loi uniforme sur A.

Comment simuler une réalisation x de la variable X?

4.2.2. Le résultat théorique.

Soient $A, B \in \mathcal{B}(\mathbb{R}^2)$ tels que $B \subset A$, de mesure de Lebesgue respectives $\lambda(B)$ et $\lambda(A)$, tels que $0 < \lambda(B) \le \lambda(A) < \infty$.

Soit X un vecteur aléatoire de loi uniforme sur B et soit (Y_n) une suite de vecteurs aléatoires de \mathbb{R}^2 indépendants et de même loi uniforme sur A.

Soit T la var. a. d. définie par : $T = \inf \{n \ge 1 \text{ tel que } Y_n \in B\}$ On a alors,

- □ la variable T suit la loi géométrique $\mathcal{G}(p)$, avec $p = \lambda(B)/\lambda(A)$.
- \square les variables Y_T et T sont indépendantes et Y_T a la même loi que la variable X.

Ainsi à partir d'une loi uniforme sur un borélien A de \mathbb{R}^2 , il est facile de simuler une loi uniforme sur un borélien $B \subset A$.

4.2.3. Remarques.

- □ Le nombre d'essais nécessaires pour obtenir un succès étant de loi géométrique de paramètre p, la moyenne du nombre d'essais est de $1/p = \lambda(A)/\lambda(B)$.
- \square II est donc crucial de minimiser la surface de $A \setminus B$.
- ☐ Plus cette surface sera petite, plus la méthode du rejet sera compétitive.
- On prend souvent pour A un pavé de \mathbb{R}^2 de la forme $[a, b] \times [c, d]$.

En effet, simuler une loi uniforme sur ce pavé, revient à simuler chaque coordonnées de manière indépendante : on simule l'abscisse du point selon la loi $\mathcal{U}_{[a,b]}$, et l'ordonnée selon la loi $\mathcal{U}_{[c,d]}$.

4.2.4. Exemple.

4.2.4.1. Le problème et l'algorithme.

On souhaite simuler un échantillon de n valeurs d'une loi uniforme sur le quart supérieur droit du disque unité.

On met en oeuvre le résultat du théorème précédent.

Ainsi pour simuler une réalisation (x_1, x_2) d'une variable aléatoire X de loi uniforme sur le quart supérieur droit du disque unité, on simule des réalisations indépendantes (u_1, u_2) d'une variable aléatoire U de loi uniforme sur le carré $[0,1] \times [0,1]$ jusqu'a trouver un couple (u_1, u_2) qui appartienne au quart supérieur du disque unité.

4.2.4.2. Le code R.

```
disque =function(x,y){
    return(ifelse(x \ge 0 \& y \ge 0 \& (x^2 + y^2) \le 1, TRUE, FALSE))
N = 200
X = matrix(NA, ncol = 2, nrow = N)
for (j in 1:N) {
    x = runif(1,0,1) ; y = runif(1,0,1)
    while(disque(x,y)==FALSE) {
        x = runif(1,0,1); y = runif(1,0,1)
    }
    X[i,] = c(x,y)
```

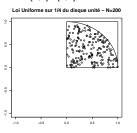
Titre = paste0("Loi Uniforme sur 1/4 du disque unité - N=",N) plot(X,xlim=c(0,1),ylim=c(0,1),xlab="",ylab="",main =Titre, cex.main=1.6) radius $\leftarrow 1$; theta $\leftarrow seq(0, pi/2, length = 200)$ lines(x = radius * cos(theta), y = radius * sin(theta))

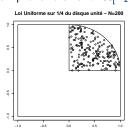
rect(0,0,1,1,lwd=2)

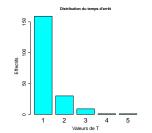
}

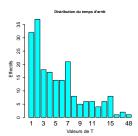
4.2.4.3. Le résultat graphique.

On trouvera ci-dessous le résultat de deux simulations, à gauche en prenant une loi $\mathcal{U}_{[0,1]\times[0,1]}$, et à droite en prenant une loi $\mathcal{U}_{[-1,1]\times[-1,1]}$.









E. ZOUGAR [INSA] 2023-01-26 Génie mathématiques GM3

4.2.6. Commentaires.

- On constate que, dans les deux cas, tous les points simulés se trouvent bien dans le quart de disque souhaité.
- □ On constate en examinant la distribution de la variable T qu'il faut effectivement simuler plus de réalisations dans le cas de la loi uniforme sur $[-1,1] \times [-1,1]$.
- Dans le cas de la loi uniforme sur $[0,1] \times [0,1]$, il faut en moyenne $4/\pi$ réalisations, alors que dans le cas de la loi uniforme $[-1,1] \times [-1,1]$, il faut $16/\pi$ réalisations

4.3. Méthode du rejet "comparatif". 4.3.1. Le résultat théorique

Ce théorème généralise la méthode du rejet pour les variables à densités. **Théorème.** Soit X une variable aléatoire réelle de densité f. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires réelles i.i.d de densité g. On suppose qu' \exists une constante $c\geq 1$ telle que $\forall x\in \mathbb{R}$,

$$f(x) \leq c g(x).$$

Pour tout réel x, on pose h(x) = f(x)/(cg(x)). Soit (U_n) une suite de variables aléatoires indépendantes de loi uniforme sur [0,1], indépendantes de la suite (Y_n) .

Alors, la variable aléatoire $T = \inf \{n \ge 1 \text{ tel que } U_n < h(Y_n)\}$ suit la loi géométrique de paramètre 1/c, les variables aléatoires T et Y_T sont indépendantes et Y_T a la même loi que X.

Ainsi, pour simuler une réalisation de X, il suffit de trouver une constante c > 0 et une densité g ("que l'on sait simuler") telle que $f(x) \le c g(x)$.

4.3.2. Remarques.

- \square La constante c et la densité f peuvent dépendre de valeurs inconnues, ce qui est important, c'est que l'on puisse calculer h(x) en tout point x.
- ☐ Plus la constante c sera grande, moins la méthode sera compétitive.

En effet, comme T suit une loi géométrique de paramètre 1/C, alors $\mathbb{E}[T] = c$, ce qui représente le nombbre moyen d'essais réalisés avant l'obtention d'une réalisation de X.

4.3.3. Exemple.

On considère une variable aléatoire réelle X dont la densité est définie pour tout $x \in \mathbb{R}$ par :

$$f(x) = \frac{1}{M} \exp\left(-\frac{x^4}{4}\right)$$
 avec $M = \int_{-\infty}^{+\infty} \exp\left(-\frac{x^4}{4}\right) dx \simeq 2,563693$

On peut montrer que pour tout $x \in \mathbb{R}$, on a la majoration $(1+x^2)\exp(-x^4/4) \le 2\exp(-1/4)$.

On peut ainsi via la méthode du rejet et l'utilisation de la densité de Cauchy simuler des réalisations de X.

En effet, si on note g la densité de la loi de Cauchy, on a :

$$f(x) = \frac{1}{M} \exp\left(-\frac{x^4}{4}\right) \le \frac{2 \exp(-1/4)}{M(1+x^2)} = Cg(x) \text{ avec } C = \frac{2\pi \exp(-1/4)}{M}$$

Ainsi, en posant $h(x) = \frac{f(x)}{Cg(x)} = \frac{1}{2}(1+x^2) \exp(-(x^4-1)/4)$, on peut

obtenir une réalisation x en simulant des réalisations indépendantes y de Y et u de U jusqu'à ce que la condition (u < h(y)) soit satisfaite. On pose ensuite x = y.

4.3.4. Illustration avec R.

```
4.3.4.1 Le code R.
g = function(x)\{g = exp(-x^4 / 4) \}
M = integrate(g, lower=-20, upper=20)$value
h = function(x){
  h = (1 + x^2) * exp(-x^4 / 4) / (2 * exp(-1/4)) }
N = 200
X = nrejet = rep(NA, N)
for(j in 1:N){
  n = 1
  u = runif(1)
  y = rcauchy(1)
  while (u > h(y)) {
    u = runif(1)
    y = rcauchy(1)
    n = n+1
  X[j] = y
  nrejet[j] = n
```

Introduction Variables discrétes Variables continues

4.3.4.2. Le résultat graphique.

On trouve dans le graphique ci-dessous un histogramme en fréquences des 200 données simulées par la méthode du rejet, sur lequel on a superposé le graphe de la densité f.

