

Modélisation et Simulations stochastiques

Sylvain Rubenthaler, Roland Diel

Table des matières

| | |
|---|-----|
| Préface | iii |
| Chapitre 1. Simulation de variables aléatoires | 1 |
| 1.1. Générateur de nombres pseudo-aléatoires | 1 |
| 1.2. Fonction de répartition et fonction quantile | 2 |
| 1.3. Simulation de variables gaussiennes (algorithme de Box-Müller) | 5 |
| 1.4. Simulation d'une variable aléatoire poissonienne | 6 |
| 1.5. Méthode de rejet | 7 |
| 1.6. Vérification par histogramme | 10 |
| Chapitre 2. Méthodes de Monte Carlo | 13 |
| 2.1. Description de la méthode | 13 |
| 2.2. Théorèmes de convergence | 14 |
| 2.3. Réduction de variance | 17 |
| Chapitre 3. Simulation d'événements rares | 27 |
| 3.1. Simulation d'événements rares pour les sommes de v.a.i.i.d à queue légère. | 28 |
| 3.2. Simulation d'événements rares pour les sommes de v.a.i.i.d à queue lourde | 31 |
| Chapitre 4. Méthodes de Monte-Carlo par chaînes de Markov | 33 |
| 4.1. Simulation d'une chaîne de Markov | 33 |
| 4.2. Rappels sur le comportement asymptotique des chaînes de Markov | 35 |
| 4.3. Algorithme de Hastings-Metropolis | 37 |
| 4.4. Algorithme de Metropolis simple | 39 |

Préface

Ce polycopié est une introduction aux méthodes de Monte-Carlo. Les prérequis sont : un cours de L3 de probabilités et des notions de **Python** .

Voici deux références que vous pouvez consulter pour approfondir les notions vues en cours :

- (1) Kroese, Taimre, Botev, *Handbook of Monte Carlo methods*
- (2) Asmussen et Glynn, *Stochastic Simulation : Algorithms and Analysis*

Chapitre 1

Simulation de variables aléatoires

Nous allons étudier ici comment simuler des réalisations d'une variable aléatoire suivant une loi donnée. Il existe de nombreuses techniques pour obtenir des simulations « exactes » ou approchées mais elles se basent toutes sur la donnée de réalisations de suites de v.a.i.i.d. de loi uniforme $\mathcal{U}([0; 1])$.

Les ordinateurs ne disposent pas de composant au comportement aléatoire¹. Quand on demande à un programme de générer des nombres tirés aléatoirement et indépendamment selon la loi $\mathcal{U}([0; 1])$, il renvoie des nombres pris dans une liste à la suite les uns des autres. En général, cette suite est suffisamment désordonnée et longue pour avoir un comportement similaire à celui d'une suite aléatoire. Avant de nous intéresser aux différentes techniques de simulation d'une loi générale, nous allons dire quelques mots sur la génération de ces suites de nombres. Les étudiants intéressés par ce problème pour consulter les références proposées dans la préface pour approfondir le sujet.

1.1. Générateur de nombres pseudo-aléatoires

Les *générateurs de nombres pseudo-aléatoires* sont des algorithmes qui, partant d'une *graine* x_0 renvoient une suite de nombres x_1, x_2, \dots qui « paraissent » être une réalisation d'une suite de variables aléatoires indépendantes et identiquement distribuées (ce que nous noterons i.i.d. dans la suite) de même loi $\mathcal{U}([0; 1])$. Un ordinateur classique fonctionnant de manière déterministe ne peut obtenir de « véritable aléa », c'est pourquoi on parle de nombres pseudo-aléatoires. Il existe maintenant des méthodes quantiques qui sont plus proches de cet objectif, mais nous n'aborderons pas ce point dans ce cours et nous nous contenterons de présenter un type de générateurs pseudo-aléatoires très utilisés actuellement, les générateurs congruentiels linéaires introduits en 1948 par le mathématicien D. Lehmer. La version unidimensionnelle est définie de la manière suivante :

$$\begin{cases} x_0 \in \mathbb{N} \\ \forall n \in \mathbb{N}, x_{n+1} = ax_n + c \bmod m \end{cases}.$$

Comme indiqué plus haut, $x_0 \in \mathbb{N}$ est appelé la *graine*, $m \in \mathbb{N}$ est le *module*, $a \in \mathbb{N}$ le *multiplicateur* et $c \in \mathbb{N}_m = \{0, \dots, m - 1\}$ l'*incrément*. Du fait de l'utilisation du modulo, pour $n \geq 1$, x_n est à valeurs dans \mathbb{N}_m .

On renvoie alors le nombre $u_n = x_n/m$ qui est un rationnel appartenant au sous-ensemble

$$\left\{ 0, \frac{1}{m}, \dots, \frac{m-1}{m} \right\}.$$

Pour obtenir un résultat raisonnable, il faut donc choisir m « grand ». Typiquement, pour obtenir des valeurs à un niveau proche de la précision machine, on choisit souvent m dans l'ensemble $\{2^{32}, 2^{64}, 2^{32} - 1, 2^{64} - 1, 2^{31} - 1, 2^{31}\}$. Le choix de a et de c est alors essentiel pour obtenir notamment une répartition des x_n proche de réalisations indépendantes de v.a. suivant la loi uniforme $\mathcal{U}([0; 1])$; les paramètres $a = 7^5$, $c = 0$ et $m = 2^{31} - 1$ étaient utilisés dans un certain nombre de générateurs de nombres pseudo-aléatoires d'ancienne génération.

Nous ne rentrerons pas ici plus dans les détails de l'efficacité de ces générateurs. Remarquons qu'on ne fixe pas ici la graine x_0 dans le choix des paramètres, celle-ci n'est en effet pas considérée comme un paramètre du modèle et le générateur, pour être acceptable, doit donner de bons résultats quelque soit la graine choisie. Celle-ci étant, soit fixée « aléatoirement » avec par exemple

1. Même un jeu de pile ou face ne se comporte pas aussi « aléatoirement » que nous le voudrions.

l'heure machine au lancement du programme, soit fixée directement par l'utilisateur qui peut alors reproduire plusieurs fois la même séquence pseudo aléatoire en reprenant la même graine.

Un des problèmes de ces générateurs est qu'ils sont nécessairement de période inférieure à m , même si ce nombre est élevé, il peut poser problème dans certaines applications. C'est une des raisons qui a poussé à introduire une variante de ces modèles qui est le type de générateurs pseudo-aléatoires le plus utilisé actuellement. Elle consiste à prendre un modulo m très faible, typiquement 2, mais à travailler en dimension plus grande que 1. Ainsi, on pose :

$$X_{n+1} = AX_n \quad [2], \quad Y_n = BX_n \quad [2], \quad u_n = \sum_{i=1}^{\ell} y_{n,i} 2^{-i}$$

où $X_n \in \{0;1\}^k$, $Y_n = (y_{n,1}, \dots, y_{n,\ell}) \in \{0;1\}^\ell$, $A \in \mathcal{M}(k \times k)$ and $B \in \mathcal{M}(\ell \times k)$. Un des générateurs de ce type les plus utilisés actuellement est le « Mersenne Twister » qui a une période de $2^{19937} - 1$.

Nous n'étudierons pas plus en détails la construction de ces générateurs

1.2. Fonction de répartition et fonction quantile

Soit X une variable aléatoire réelle sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On appelle *fonction de répartition* de la v.a. X , la fonction

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto \mathbb{P}(X \leq t) \end{aligned}$$

PROPOSITION 1.1. Soit X une v.a.r. de fonction de répartition F_X . Alors :

- (1) F_X est croissante
- (2) $\lim_{t \rightarrow +\infty} F(t) = 1$ et $\lim_{t \rightarrow -\infty} F(t) = 0$.
- (3) Pour tout réel t , F est continue à droite en t et sa limite à gauche $F(t^-)$ existe et est donnée par $F(t^-) = \mathbb{P}(X < t)$.

De plus, si une fonction F vérifie les 3 points précédents alors il existe une variable aléatoire X dont F est la fonction de répartition.

DÉMONSTRATION. Cf cours de L3. □

Remarque : F_X est continue en $t \in \mathbb{R}$ si et seulement si $\mathbb{P}(X = t) = 0$.

PROPOSITION 1.2. La fonction de répartition caractérise la loi de la v.a.r. i.e. $\mathbb{P}_X = \mathbb{P}_Y$ ssi $F_X = F_Y$.

Soit F une fonction de répartition, on appelle *inverse généralisée* de F ou *fonction quantile* la fonction $Q : [0; 1] \rightarrow [-\infty; \infty]$ définie par

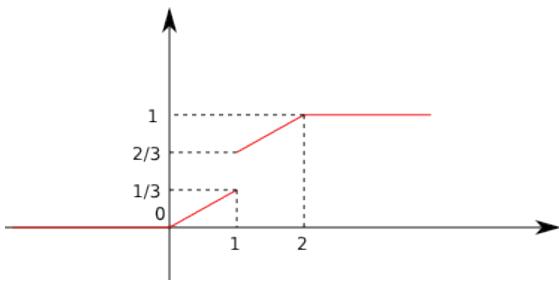
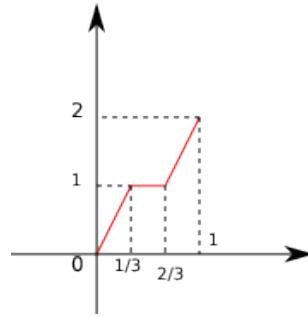
$$\forall t \in [0; 1], \quad Q(t) = \inf\{x \in \mathbb{R}, \quad F(x) \geq t\}.$$

Si F est inversible, la fonction quantile est l'inverse de F , $Q = F^{-1}$, mais la fonction quantile reste bien définie même lorsque F n'est pas inversible. pas d'inverse). Voir les figures 1.2.1 et 1.2.2 pour un exemple (pour $t \in [1/3; 2/3]$, $\{x, F(x) \geq t\} = [1; +\infty[$ et donc $Q(t) = 1$).

PROPOSITION 1.3. Soit X une v.a.r. de fonction de répartition F_X et de fonction quantile Q_X . Alors

- (1) La fonction quantile Q_X est croissante sur $[0; 1]$ et continue à gauche sur $]0; 1[$.
- (2) Si la fonction de répartition F_X est continue, alors la fonction quantile Q_X est strictement croissante.
- (3) Pour tout $x \in \mathbb{R}$, pour tout $t \in]0; 1[$,

$$Q_X(t) \leq x \text{ ssi } t \leq F_X(x)$$

FIGURE 1.2.1. Fonction de répartition F .FIGURE 1.2.2. Fonction Q .

PROPOSITION 1.4. Soit X une v. a. r. de fonction de répartition F et de fonction quantile Q et $U \sim \mathcal{U}([0; 1])$. Alors,

- (1) $Q(U)$ a même loi que X
- (2) si F est continue, alors $F(X)$ a même loi que U .

DÉMONSTRATION.

(1) Soit $u \in [0; 1]$ et $t \in \mathbb{R}$.

- Si $u \leq F(t)$ alors $t \in \{x, F(x) \geq u\}$ donc $Q(u) \leq t$.
- Supposons que $Q(u) \leq t$. Soit $\epsilon > 0$. Il existe $x \leq t + \epsilon$ tel que $F(x) \geq u$. La fonction F est croissante donc $F(x) \leq F(t + \epsilon)$, donc $u \leq F(t + \epsilon)$. Ceci est vrai pour tout $\epsilon > 0$ et la fonction F est continue à droite (parce que c'est une fonction de répartition) donc $u \leq F(t)$.

Nous avons donc

$$[u \leq F(t)] \Leftrightarrow [Q(u) \leq t].$$

ce qui se traduit par, pour tout $t \in \mathbb{R}$,

$$\mathbb{P}(Q(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t).$$

Donc $Q(U)$ a la même fonction de répartition que X (et donc ils ont la même loi).

(2) Cf TD. □

EXEMPLE 1.5. Dans le cas de la fonction F donnée dans les figures 1.2.1, 1.2.2, nous avons

$$F(x) = \begin{cases} \frac{x}{3} & \text{si } x \in [0; 1[\\ \frac{x}{3} + \frac{1}{3} & \text{si } x \in [1; 2] \\ 1 & \text{si } x \geq 2 \\ 0 & \text{si } x \leq 0 \end{cases} \quad \text{et} \quad Q(t) = \begin{cases} 3t & \text{si } t \in [0; \frac{1}{3}] \\ 1 & \text{si } t \in [\frac{1}{3}; \frac{2}{3}] \\ 3t - 1 & \text{si } t \in [\frac{2}{3}; 1] \end{cases}.$$

La variable `finv(u)` renvoyé par le programme 1.1 renvoie une variable de fonction de répartition F .

Simulation d'une loi discrète. Soit $n \in \mathbb{N}^*$. Soient $x_1 < x_2 < \dots < x_n$ dans \mathbb{R} . Soient $p_1, p_2, \dots, p_n \in [0; 1]$ tels que $p_1 + p_2 + \dots + p_n = 1$. Soit X une variable aléatoire telle que $\mathbb{P}(X_i = x_i)$ pour tout $i \in \{1, 2, \dots, n\}$. La fonction de répartition de X s'écrit

$$\forall t \in \mathbb{R}, \quad F(t) = \sum_{i=1}^n p_i \mathbb{1}_{t \geq x_i}$$

(voir un exemple de dessin avec $n = 3$ dans la Figure 1.2.3). La fonction quantile s'écrit

$$\forall u \in [0; 1], \quad Q(u) = \sum_{j=1}^n x_j \mathbb{1}_{\sum_{k=1}^{j-1} p_k \leq u < \sum_{k=1}^j p_k},$$

Programme 1.1 Simulation par inversion de la fonction de répartition

```
finv<-function(u)
{
  if (u<1/3)
  { z=3*u }
  else if (u<2/3)
  { z=1 }
  else
  { z=3*u-1 }
  return(z)
}
u=runif(1,0,1) #simulation d'une variable uniforme (dans [0;1])
print(finv(u))
```

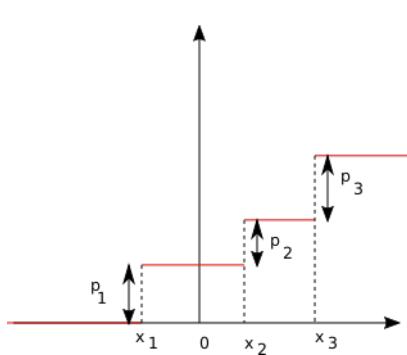


FIGURE 1.2.3. Fonction de répartition F .

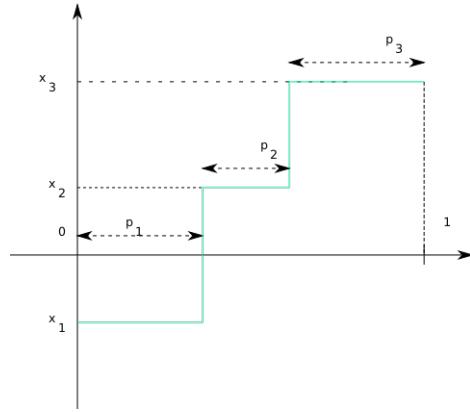


FIGURE 1.2.4. Fonction Q .

(voir un exemple de dessin avec $n = 3$ dans la Figure 1.2.4). Le programme 1.2 renvoie une variable de fonction de répartition F (le tableau x contient les valeurs x_i et le tableau p contient les valeurs p_i).

Programme 1.2 Simulation par inversion de la fonction de répartition.

```
i=0
t=0
b=0
u=runif(1,0,1)
while (b==0)
{
  i=i+1
  t=t+p[i]
  if (t>u)
  { b=1 }
}
print(x[i])
```

Simulation d'une loi exponentielle. On rappelle que X suit une loi exponentielle de paramètre $\lambda > 0$, on note $X \sim \mathcal{E}(\lambda)$ si pour tout $t \in \mathbb{R}$, sa fonction de répartition s'écrit :

$$F_X(t) = (1 - \exp(-\lambda t)) \mathbb{1}_{t>0}.$$

Ainsi, sa fonction quantile s'écrit

$$\forall x \in]0; 1[, \quad Q(x) = -\frac{\log(1-x)}{\lambda}.$$

Si $U \sim \mathcal{U}([0; 1])$, nous avons donc, d'après le résultat ci-dessus, $-\log(1-U)/\lambda \sim \mathcal{E}(\lambda)$. Remarquons que $1-U$ a même loi que U et donc

$$-\frac{\log(U)}{\lambda} \sim \mathcal{E}(\lambda).$$

Attention, nous écrivons log pour le logarithme népérien.

1.3. Simulation de variables gaussiennes (algorithme de Box-Müller)

LEMME 1.6.

(1) Si R et Θ sont deux variables indépendantes de loi respective $\mathcal{E}(1/2)$ et $\mathcal{U}([0; 2\pi])$ alors

$$(X, Y) = (\sqrt{R} \cos \Theta, \sqrt{R} \sin \Theta)$$

est un couple de variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$.

(2) Si U, V sont de loi $\mathcal{U}([0; 1])$ et indépendantes alors

$$(X, Y) = (\sqrt{-2 \log(U)} \cos(2\pi V), \sqrt{-2 \log(U)} \sin(2\pi V))$$

est un couple de variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$.

Pour simuler une variable X de loi $\mathcal{N}(0, 1)$, il suffit donc de prendre $U, V \sim \mathcal{U}([0; 1])$ indépendantes et poser $X = \sqrt{-2 \log(U)} \cos(2\pi V)$. Et pour simuler une variable X de loi $\mathcal{N}(\mu, \sigma^2)$, il suffit de prendre $X = \mu + \sigma \sqrt{-2 \log(U)} \cos(2\pi V)$.

DÉMONSTRATION. Au vu de ce qui précéde, le point (2) est une conséquence directe du point (1). On ne démontre donc que ce résultat.

On se rappelle que pour identifier la loi d'une variable aléatoire, il suffit de savoir calculer les espérances de fonctions de cette variables aléatoires (pour toute fonction dans un ensemble de fonctions test comme les fonctions continues bornées positives par exemple, cf cours de L3). Nous prenons donc $f \in \mathcal{C}_b^+(\mathbb{R}^2)$ ($\mathcal{C}_b^+(\mathbb{R}^2)$ est l'ensemble des fonctions positives, continues, bornées de \mathbb{R}^2 dans \mathbb{R}). Notons X et Y deux variables indépendantes de loi $\mathcal{N}(0, 1)$. Nous avons :

$$\begin{aligned} \mathbb{E}[f(X, Y)] &= \int_{\mathbb{R}^2} f(x, y) \frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy \\ &= \int_{\mathbb{R}^2 \setminus \{(x, 0), x \geq 0\}} f(x, y) \frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy \end{aligned}$$

car l'ensemble $\{(x, 0), x \geq 0\}$ est de mesure nulle. Ainsi, le changement de variables en coordonnées polaires donné par le difféomorphisme

$$(r, \theta) \in \mathbb{R}_+^* \times]0; 2\pi[\rightarrow (x, y) = (r \cos \theta, r \sin \theta) \in \mathbb{R}^2 \setminus \{(x, 0), x \geq 0\}.$$

admet pour matrice jacobienne

$$J_1(x, y) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -r \sin(\theta) & r \cos(\theta) \end{bmatrix},$$

de déterminant $\det(J_1) = r \cos^2(\theta) + r \sin^2(\theta) = r$. Donc

$$\mathbb{E}[f(X, Y)] = \int_{(r, \theta) \in \mathbb{R}_+^* \times]0; 2\pi[} f(r \cos \theta, r \sin \theta) \frac{e^{-r^2/2}}{2\pi} r dr d\theta$$

Nous effectuons maintenant un nouveau changement de variable dans l'intégrale en r : $\rho = r^2$, $d\rho = 2r dr$, ce qui donne :

$$\mathbb{E}[f(X, Y)] = \int_0^\infty \int_0^{2\pi} f(\sqrt{\rho} \cos \theta, \sqrt{\rho} \sin \theta) \frac{1}{2\pi} 2e^{-\rho/2} d\theta d\rho$$

Ceci est vrai pour toute fonction $f \in \mathcal{C}_b^+(\mathbb{R}^2)$ donc le couple (X, Y) a même loi que le couple $(\sqrt{R} \cos \Theta, \sqrt{R} \sin \Theta)$ où (R, Θ) admet pour densité la fonction

$$p_{(R,\Theta)}(\rho, \theta) = \frac{1}{2\pi} \mathbb{1}_{[0;2\pi]}(\theta) \times 2e^{-\rho/2} \mathbb{1}_{]0;\infty[}(\rho,$$

ce qui prouve le point (1) du lemme. \square

1.4. Simulation d'une variable aléatoire poissonienne

Une variable de loi de Poisson de paramètre $\lambda > 0$ est une variable X à valeurs dans \mathbb{N} telle que

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(X = n) = \frac{e^{-\lambda} \lambda^n}{n!}.$$

PROPOSITION 1.7. *Soit $(T_i)_{i \geq 1}$ une suite de variables i.i.d. de loi $\mathcal{E}(\lambda)$, $\lambda > 0$. On considère pour tout $n \in \mathbb{N}^*$, la somme des n premières variables aléatoires*

$$S_n = \sum_{i=1}^n T_i.$$

Alors

(1) pour tout $n \in \mathbb{N}^*$, S_n suit la loi $\Gamma(n, \lambda)$ de densité $p_{n,\lambda}$ définie par :

$$\forall x \in \mathbb{R}, \quad p_{n,\lambda}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbb{1}_{x>0}.$$

(2) pour tout $t > 0$,

$$N_t = \sum_{n \geq 1} n \mathbb{1}_{\{S_n \leq t < S_{n+1}\}} = \sum_{n \geq 1} \mathbb{1}_{S_n \leq t}$$

est une variable de Poisson de paramètre λt .

DÉMONSTRATION. Pour démontrer le résultatat (1), on procède par récurrence sur n . La densité $p_{1,\lambda}$ est simplement la densité exponentielle.

Supposons maintenant que S_n soit de densité $p_{n,\lambda}$. Comme $S_{n+1} = S_n + T_{n+1}$, par indépendance des variables S_n et T_{n+1} , on a

$$\begin{aligned} \forall x \in \mathbb{R}, \quad p_{n+1,\lambda}(x) &= p_{n,\lambda} * p_{1,\lambda}(x) \\ &= \int_{\mathbb{R}} p_{n,\lambda}(y) p_{1,\lambda}(x-y) dy \\ &= \int_{\mathbb{R}} \frac{\lambda^n}{(n-1)!} (y)^{n-1} e^{-\lambda y} \mathbb{1}_{y>0} \lambda e^{-\lambda(x-y)} \mathbb{1}_{x-y>0} dy \\ &= e^{-\lambda x} \int_0^x \frac{\lambda^{n+1}}{(n-1)!} y^{n-1} dy \\ &= e^{-\lambda x} \frac{\lambda^{n+1}}{n!} x^n. \end{aligned}$$

Le principe de récurrence nous dit alors que le résultatat est vrai pour tout $n \in \mathbb{N}^*$.

Continuons avec le résultatat (2). Commençons par vérifier que la variable N_t est bien définie. Pour tout $\omega \in \Omega$, il existe un et un seul $n(\omega) \in \mathbb{N}^*$ tel que $S_n(\omega) \leq t < S_{n+1}(\omega)$ et donc $N_t(\omega) = n(\omega)$ (tous les autres termes de la somme sont nuls). On peut également remarquer que pour tout $t > 0$, $n \in \mathbb{N}$,

$$\mathbb{P}(N_t \leq n) = \mathbb{P}(S_{n+1} > t)$$

En utilisant le résultatat du point (1), une IPP montre alors que

$$\mathbb{P}(S_{n+1} > t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} + \mathbb{P}(S_n > t).$$

Le résultatat est obtenu directement par récurrence. \square

Nous pouvons donc simuler une variable de Poisson à l'aide d'une boucle « tant que » (et en utilisant le résultatat de la partie 1.2) (voir le programme 1.3)

Programme 1.3 Simulation d'une variable aléatoire poissonienne.

```

lambda=2 #on fixe une valeur arbitraire de lambda
t=1 #on fixe une valeur arbitraire de t
b=0
s=0
n=-1
while (b==0)
{
  n=n+1
  s=s-log(runif(1,0,1))/lambda
  if (s>t)
  {
    b=1
  }
}
print(n)

```

1.5. Méthode de rejet

Simulation d'une loi uniforme sur un borélien de \mathbb{R}^d . On veut ici simuler une variable X de loi uniforme sur un ensemble (borélien) $D \subset \mathbb{R}^d$, de mesure de Lebesgue positive ie X doit avoir pour densité la fonction :

$$\forall x \in \mathbb{R}^d, p_X(x) = \frac{1}{|D|} \mathbb{1}_{x \in D}.$$

Le cas où D est un pavé $\prod_{i=1}^d [a_i; b_i]$ est simple à traiter car on peut remarquer que les marginales X_1, \dots, X_d sont des variables indépendantes de loi respective $\mathcal{U}([a_i; b_i])$.

Pour le cas général où il n'y a pas nécessairement de décomposition simple de ce type, on peut utiliser une méthode appelée algorithme de rejet basée sur le résultat suivant.

PROPOSITION 1.8. Soit D et R deux sous-ensembles (boréliens) de \mathbb{R}^d tels que $D \subset R$. Soit $(X_n)_{n \geq 1}$ une suite de variable i.i.d. de loi uniforme sur R . Soit $T = \inf\{n \in \mathbb{N}^*, X_n \in D\}$. Alors

- $T \sim \mathcal{G}(|D|/|R|)$, en particulier T est finie presque sûrement
- $X_T \sim \mathcal{U}(D)$
- les variables T et X_T sont indépendantes

Pour simuler des variables de loi uniforme sur D , l'idée de l'algorithme du rejet est donc de construire un ensemble R contenant D et sur lequel il est facile de simuler des variables de loi uniforme, par exemple un pavé, puis de tirer une suite de réalisations indépendantes de variables de loi uniforme sur R jusqu'à en obtenir une qui est dans D . Cette dernière valeur est donc d'après la proposition ci-dessus une réalisation d'une variable de loi uniforme sur D . Le temps d'exécution de l'algorithme est aléatoire. Néanmoins la loi de T étant la loi géométrique $\mathcal{G}(|D|/|R|)$, le nombre moyen de boucles effectuées est $\frac{|R|}{|D|}$ et le temps moyen d'exécution est proportionnel à ce nombre. Pour minimiser le temps moyen d'exécution, il faut donc prendre un ensemble R de taille la plus petite possible.

La proposition 1.8 précédente est un corollaire de la proposition 1.9 ci-dessous. Elle sera donc démontrée après.

Simulation d'une loi quelconque. On veut simuler une variable de densité f (dans \mathbb{R}^d) et on suppose qu'il existe une loi de densité g facile à simuler telle que

$$(1.1) \quad \forall x \in \mathbb{R}^d, f(x) \leq kg(x).$$

avec k une constante. On pose $\alpha(x) = \frac{f(x)}{kg(x)} \mathbb{1}_{\{g(x)>0\}}$.

PROPOSITION 1.9. Soit $(X_n, U_n)_{n \geq 1}$ une suite de variable i.i.d. telle que, pour tout n , X_n est de loi de densité g , U_n suit la loi uniforme sur $[0; 1]$ et toutes les v.a. $(X_n)_{n \geq 1}$ et $(U_n)_{n \geq 1}$ sont indépendantes. Soit $T = \inf\{n \geq 1, U_n \leq \alpha(X_n)\}$. Alors

- la v.a. T suit la loi $\mathcal{G}(1/k)$, en particulier T est finie presque sûrement
- la v.a. X_T est de loi de densité f .
- les v.a. T et X_T sont indépendantes

DÉMONSTRATION. Remarquons tout d'abord que comme f et g sont des densités, en intégrant la relation (1.1), on voit que nécessairement $k \geq 1$. On s'intéresse alors à la loi de T . Par définition,

$$T = \inf\{n \geq 1, \mathbb{1}_{U_n \leq \alpha(X_n)}\}$$

et les variables $(\mathbb{1}_{U_n \leq \alpha(X_n)})_{n \geq 1}$ sont des v.a.i.i.d. de loi de Bernoulli de paramètre $\mathbb{P}U_1 \leq \alpha(X_1)$. Il nous suffit donc de montrer que cette probabilité vaut $1/k$ pour obtenir $T \sim \mathcal{G}(1/k)$. Et, en effet, par indépendance de U_1 et de X_1 , en utilisant le théorème de Fubini,

$$\begin{aligned} \mathbb{P}(U_1 \leq \alpha(X_1)) &= \int_{\mathbb{R}^d} \int_0^1 \mathbb{1}_{u \leq \alpha(x)} g(x) du dx \\ &= \int_{\mathbb{R}^d} g(x) \alpha(x) dx = \frac{1}{k} \int_{\mathbb{R}^d} f(x) dx = \frac{1}{k}. \end{aligned}$$

Ainsi, on a bien $T \sim \mathcal{G}(1/k)$. En particulier, la v.a. T est p.s. finie et X_T est bien définie.

Nous calculons maintenant la loi du couple (T, X_T) par la méthode de la fonction muette. Soit $\varphi \in C_b^+(\mathbb{R}^d)$. Nous avons $\varphi(X_T) = \sum_{n=1}^{+\infty} \varphi(X_n) \mathbb{1}_{T=n}$ (pour tout ω , cette somme contient au plus un terme non nul). Ainsi,

$$\begin{aligned} \mathbb{E}[\varphi(T, X_T)] &= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(n, X_n) \mathbb{1}_{T=n}] \\ &= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(n, X_n) \prod_{i=1}^{n-1} \mathbb{1}_{U_i > \alpha(X_i)} \times \mathbb{1}_{U_n \leq \alpha(X_n)}] \\ (\text{par indépendance des } (X_i, U_i)) &= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(n, X_n) \mathbb{1}_{U_n \leq \alpha(X_n)}] \prod_{i=1}^{n-1} \mathbb{E}[\mathbb{1}_{U_i > \alpha(X_i)}]. \end{aligned}$$

On a déjà vu que, pour tout $i \geq 1$,

$$\mathbb{E}[\mathbb{1}_{U_i > \alpha(X_i)}] = 1 - \mathbb{P}(U_1 \leq \alpha(X_1)) = 1 - \frac{1}{k}$$

Un calcul analogue montre que, pour tout $n \geq 1$,

$$\mathbb{E}[\varphi(n, X_n) \mathbb{1}_{U_n \leq \alpha(X_n)}] = \frac{1}{k} \int_{\mathbb{R}^d} \varphi(n, x) f(x) dx.$$

Nous avons donc

$$\mathbb{E}[\varphi(T, X_T)] = \sum_{n=1}^{+\infty} \int_{\mathbb{R}^d} \varphi(n, x) f(x) \frac{1}{k} \left(1 - \frac{1}{k}\right)^{n-1} dx.$$

Cela montre que les v.a. T et X_T sont indépendantes et que leur loi respective est bien celle indiquée dans la proposition. \square

DÉMONSTRATION DE LA PROPOSITION 1.8. On applique le résultat de la proposition 1.9 avec

$$\forall x \in \mathbb{R}^d, \quad f(x) = \frac{1}{|D|} \mathbb{1}_{x \in D}, \quad g(x) = \frac{1}{|R|} \mathbb{1}_{x \in R} \quad \text{et} \quad k = \frac{|R|}{|D|}.$$

La fonction α s'écrit alors :

$$\forall x \in \mathbb{R}^d, \quad \alpha(x) = \mathbb{1}_{x \in D}.$$

Elle ne prend donc que les valeurs 0 et 1 ainsi, presque sûrement, l'événement $U_n \leq \alpha(X_n)$ ne dépend pas de la valeur de U_n et est équivalent à $X_n \in D$. On peut donc simplifier l'expression

de la variable T en $T = \min\{n \geq 1, X_n \in D\}$, ce qui nous permet d'obtenir le résultat de la proposition 1.8. \square

EXEMPLE 1.10. Soit $f(x) = \frac{e^{-x^3}}{Z} \mathbb{1}_{x \geq 1}$ avec $Z = \int_1^{+\infty} e^{-x^3} dx$ (la constante Z n'est pas connue mais nous n'en aurons pas besoin pour appliquer l'algorithme). On introduit la fonction g définie pour tout $x \in \mathbb{R}$ par $g(x) = \frac{\mathbb{1}_{x \geq 1}}{x^2}$. Les fonctions f et g sont des densités de probabilité. Il est facile de simuler suivant la loi de densité g en utilisant la technique d'inversion de la fonction de répartition. Pour appliquer l'algorithme du rejet, nous voulons alors dominer la fonction f par kg avec k la plus petite valeur possible. Nous commençons par étudier la fonction $h : x \mapsto x^2 e^{-x^3}$ sur $[1; \infty[$. Nous avons $h'(x) = (2x - 3x^4)e^{-x^3} \leq 0$ pour tout $x \geq 1$. Donc, pour tout $x \geq 1$, $h(x) \leq h(1) = e^{-1}$. Nous en déduisons, pour $x \geq 1$,

$$f(x) \leq \frac{1}{eZ} g(x).$$

Nous posons $k = 1/(eZ)$, soit

$$\alpha(x) = \frac{f(x)}{kg(x)} = x^2 e^{-x^3 + 1}.$$

Nous ne connaissons pas Z mais nous pouvons calculer $\alpha(x)$ pour tout $x \geq 1$ donc nous pouvons utiliser l'algorithme de simulation par rejet pour simuler suivant la densité f (voir programme 1.4).

Programme 1.4 Simulation par la méthode de rejet.

```
b=0
while (b==0)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  y=1/(1-v)
  if (u<y^2*exp(-y^3)*exp(1))
    { b=1 }
}
print(y)
```

Dans ce programme, nous utilisons les résultats suivants pour simuler suivant la densité g par inversion de la fonction de répartition :

$$\text{pour } x \geq 1, G(x) = \int_1^x g(u) du = \left[-\frac{1}{u} \right]_1^x = 1 - \frac{1}{x},$$

$$\text{pour } u \in [0; 1], G^{-1}(u) = \frac{1}{1-u}.$$

EXEMPLE 1.11 (Loi conditionnelle). Soit X un vecteur aléatoire à valeur dans \mathbb{R}^d , de densité g . Soit A un sous-ensemble (borélien) de \mathbb{R}^d tel que $\mathbb{P}(X \in A) > 0$. La loi de X conditionnée à tomber dans A a pour densité

$$x \mapsto f(x) = \frac{g(x) \mathbb{1}_A(x)}{\mathbb{P}(X \in A)} = \frac{g(x) \mathbb{1}_A(x)}{\int_{u \in A} g(u) du}.$$

Nous remarquons que pour tout $x \in \mathbb{R}^d$, $f(x) \leq Cg(x)$ avec $C = (\int_{u \in A} g(u) du)^{-1}$. Donc, si nous tirons $(X_n, U_n)_{n \geq 1}$ i.i.d. avec $X_n \perp\!\!\!\perp U_n$, X_n de loi de densité g , $U_n \sim \mathcal{U}([0; 1])$, et si nous posons $T = \inf\{n \geq 1 : U_n \leq f(X_n)/(Cg(X_n))\}$, alors X_T est de loi de densité f . Ici, le ratio $f/(Cg)$ se simplifie en

$$\frac{f(x)}{Cg(x)} = \mathbb{1}_A(x).$$

Donc, comme dans le cas de la loi uniforme sur des ensembles, l'algorithme de rejet consiste simplement à tirer des X_n jusqu'à ce qu'il y en ait un dans A .

REMARQUE 1.12 (Temps de calcul). Le nombre de boucles effectuées dans la méthode du rejet est aléatoire. Sous les hypothèses de la proposition ci-dessus, ce nombre de boucles est T de loi $\mathcal{G}(1/k)$ et de moyenne k . Donc, le temps de calcul moyen est proportionnel à k . C'est pourquoi on cherche une constante k satisfaisant (1.1) qui soit la plus petite possible.

REMARQUE 1.13. L'algorithme ci-dessus est encore valable si la v.a. X à simuler a une densité f par rapport à une mesure μ quelconque et que cette densité est majorée par kg où g est la densité par rapport à μ d'une variable Y facile à simuler. Ceci se traduit par, pour tout ensemble (borélien) A ,

$$\mathbb{P}(X \in A) = \int_A f(x)\mu(dx) \leq \int_A kg(x)\mu(dx) = k\mathbb{P}(Y \in A).$$

Si la v.a. X a une loi portée par un ensemble discret E , on peut choisir pour μ la mesure de comptage sur E et la méthode de rejet fonctionne pour les lois discrètes. On cherche alors à obtenir une majoration du type

$$\forall x \in E, \quad \mathbb{P}(X = x) \leq k\mathbb{P}(Y = x),$$

où Y est une v.a. facile à simuler et k une constante supérieure ou égale à 1.

1.6. Vérification par histogramme

Quand on simule une variable aléatoire réelle à l'aide d'une des méthodes ci-dessus, on peut vérifier empiriquement que la loi simulée est bien celle que l'on voulait. Si la représentation est directe pour les variables aléatoires discrètes, la logique derrière la vérification par histogramme pour les variables continues réelles peut paraître moins naturelle. Nous expliquons le principe général dans cette section.

Si X suit une loi de densité f sur \mathbb{R} , l'idée est d'approcher la densité de X sur un segment $[a; b]$ en considérant une subdivision de pas constant $h = (b - a)/N$

$$a = x_0 < x_1 < \cdots < x_k < \cdots < x_N = b$$

Pour tout $k \leq N - 1$, on approche alors $f(x)$ sur l'intervalle $[x_k; x_{k+1}]$ par

$$\forall x \in [x_k; x_{k+1}], \quad \hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \mathbb{1}_{[x_k; x_{k+1}]}(X_i)$$

où les $(X_n)_{n \in \mathbb{N}}$ sont des v.a.i.i.d. de loi de densité f .

Supposons f dérivable sur $[a; b]$ et de dérivée f' bornée $\sup_{x \in [a; b]} |f'(x)| < C$ pour une certaine constante C . Pour tout $k \leq N - 1$, $h > 0$, la LGN implique

$$\frac{1}{nh} \sum_{i=1}^n \mathbb{1}_{[x_k; x_{k+1}]}(X_i) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \frac{\mathbb{P}(X_1 \in [x_k; x_{k+1}])}{h}.$$

Or,

$$\frac{\mathbb{P}(X_1 \in [x_k; x_{k+1}])}{h} = \frac{1}{h} \int_{x_k}^{x_{k+1}} f(x) dx$$

donc

$$\begin{aligned} \left| \frac{\mathbb{P}(X_1 \in [x_k; x_{k+1}])}{h} - f(x) \right| &= \left| \frac{1}{h} \int_{x_k}^{x_{k+1}} f(t) - f(x) dt \right| \leq \frac{1}{h} \int_{x_k}^{x_{k+1}} |f(t) - f(x)| dt \\ &\leq \frac{1}{h} \int_{x_k}^{x_{k+1}} C|t - x| dt = \frac{C}{2h} ((x_{k+1} - x)^2 + (x - x_k)^2) \\ &\leq \frac{Ch}{2} \xrightarrow[h \searrow 0]{} 0 \end{aligned}$$

car l'avant dernière ligne est maximale en $x = x_k$ (ou $x = x_{k+1}$). Attention toutefois, ce résultat n'est pas suffisant pour montrer la convergence de l'histogramme dans son intégralité car on regarde

successivement une limite en n , puis une limite en h . On peut néanmoins montrer moyennant quelques calculs supplémentaires que pour une constante K dépendant de f , a et b :

$$\mathbb{E} \left[\int_a^b |\hat{f}_n(x) - f(x)|^2 dx \right] \leq K(h^2 + \frac{1}{nh}).$$

Ainsi, en prenant par exemple $h \sim n^{-1/3}$, on obtient la convergence globale de l'estimateur. Dans le programme 1.5, nous simulons des variables de densité $f : x \in \mathbb{R} \mapsto \mathbb{1}_{x \geq 1} e^{-x^3}/Z$ (comme dans l'exemple 1.10) et nous traçons l'histogramme empirique des variables et le graphe de f dans le même repère (voir figure 1.6.1, dans laquelle il y a aussi le graphe de g en bleu).

Programme 1.5 Tracé d'un histogramme

```
#Nous déclarons une fonction qui simule une variable de densité f.
#Remarque : l'argument t n'est pas utilisé.
simu<-function(t)
{
  b=0
  while (b==0)
  {
    u=runif(1,0,1)
    v=runif(1,0,1)
    y=1/(1-v)
    if (u<y^2*exp(-y^3)*exp(1))
      { b=1 }
  }
  return(y)
}
f1<-function(x)
{
  if (x>1)
    { z=exp(-x^3) }
  else
    { z=0 }
  return(z)
}
Zi=integrate(f1,1,5)
Z=Zi$value #calcule la constante Z
n=5000
tab=c()
for (i in 1:n)
{
  x=simu(1)
  tab=c(tab,x)
}
hist(tab,freq=FALSE)
abs=seq(1.01,5,0.01)
y=exp(-abs*abs*abs)/Z
lines(abs,y,type='l',col='red')
```

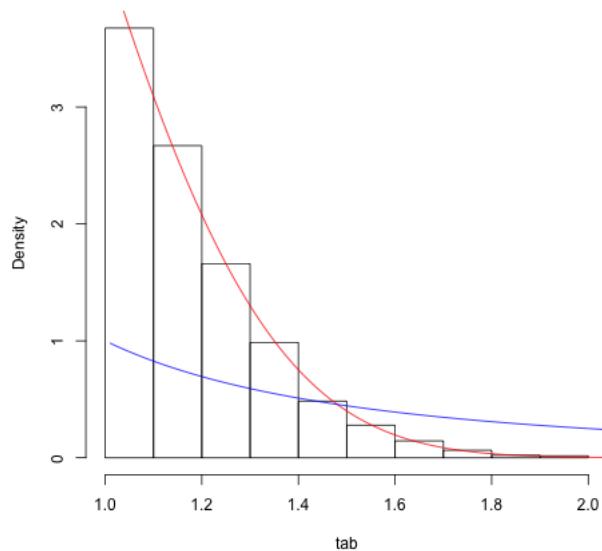


FIGURE 1.6.1. Histogramme et densité

Chapitre 2

Méthodes de Monte Carlo

2.1. Description de la méthode

Supposons que l'on veuille calculer une valeur approchée d'une quantité E . On va voir ici une méthode stochastique, basée sur la simulation de v.a.i.i.d., d'approximation de E . La première étape est de la mettre sous forme d'une espérance $E = \mathbb{E}[X]$ avec X une variable aléatoire \mathbb{L}^1 . Si on sait simuler des variables X_1, X_2, \dots i.i.d. de même loi que X , alors, la Loi des Grands Nombres nous dit que nous pouvons approcher E par

$$(2.1) \quad E \approx \frac{X_1 + X_2 + \dots + X_N}{N}$$

avec N « grand ». C'est ce type d'approximation que l'on appelle « méthode de Monte-Carlo ».

EXEMPLE 2.1. Supposons que l'on cherche à calculer

$$E = \int_{[0;1]^d} f(u_1, \dots, u_d) du_1 \dots du_d,$$

avec f bornée. Nous posons $X = f(U_1, \dots, U_d)$ où les U_1, \dots, U_d sont des variables aléatoires indépendantes suivant toutes une loi uniforme sur $[0; 1]$. Alors :

$$E = \mathbb{E}[f(U_1, \dots, U_d)] = \mathbb{E}[X].$$

Nous pouvons donc facilement simuler $U_1^{(1)}, \dots, U_d^{(1)}, U_1^{(2)}, \dots, U_d^{(2)}, \dots$ des variables i.i.d. de loi uniforme sur $[0; 1]$ (nous noterons cette loi $\mathcal{U}([0; 1])$) et donc $X_1 = f(U_1^{(1)}, \dots, U_d^{(1)}), X_2 = f(U_1^{(2)}, \dots, U_d^{(2)})$, ... sont i.i.d. Le programme 2.1 fournit un exemple de programme approxinant E par une méthode de Monte-Carlo (avec $d = 3$, $n = 1000$ itérations, $f(u_1, u_2, u_3) = \sin(u_1) \sin(u_2) \sin(u_3)$).

Si nous cherchons à évaluer une intégrale de la forme

$$E = \int_{\mathbb{R}^d} g(x) f(x) dx$$

avec f une densité de probabilité (c'est à dire f positive et $\int_{\mathbb{R}^d} f(x) dx = 1$), alors nous pouvons écrire $E = \mathbb{E}[g(X)]$ avec X de loi de densité f . Nous sommes encore dans la situation où nous voulons calculer l'espérance d'une variable aléatoire (si X est une variable aléatoire, alors $g(X)$ est une variable aléatoire, à condition que g soit borélienne).

Programme 2.1 approximation de Monte-Carlo

```

d=3
n=1000
s=0
for (i in 1:n)
{
  u=runif(d,0,1) #simulation de d variables uniformes indépendantes
  s=s+sin(u[1])*sin(u[2])*sin(u[3])
}
print(s/n)

```

2.2. Théorèmes de convergence

2.2.1. Convergence.

THÉORÈME 2.2 (Loi forte des grands nombres.). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d., à valeurs dans \mathbb{R}^d ($d \in \mathbb{N}^*$). On suppose que les X_i sont dans \mathbb{L}^1 . Alors

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X].$$

Ce théorème nous montre que l'approximation de Monte-Carlo (2.1) est valide pour des v.a.i.i.d. sous la simple hypothèse d'existence de $E = \mathbb{E}[f(X)]$.

2.2.2. Vitesse de convergence.

THÉORÈME 2.3 (Théorème de la limite centrale.). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.i.i.d. \mathbb{L}^2 , à valeurs dans \mathbb{R} . On note $\sigma^2 = \mathbb{V}(X)$, la variance de X . Alors

$$\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X] \right) \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0, 1).$$

Comme la loi limite est continue, ceci implique en particulier, sous les hypothèses du théorème ci-dessus, que, pour tout $a < b$,

$$\mathbb{P} \left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X] \right) \leq b \right) \xrightarrow[n \rightarrow +\infty]{} \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Si nous cherchons à approcher $\mathbb{E}[X]$ par une méthode de Monte-Carlo, nous pouvons donc dire que l'erreur $\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X]$ est d'ordre $\frac{\sigma}{\sqrt{n}}$. Mais cette erreur est aléatoire : le théorème ci-dessus nous dit que $\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X]$ est à peu près de loi $\mathcal{N}(0, \sigma^2/n)$. Nous ne pouvons donc pas la borner. En revanche, nous pouvons donner un intervalle de confiance pour le résultat.

Intervalle de confiance.

- Un *intervalle de confiance de niveau de confiance* $1 - \alpha$ (ou de niveau d'erreur α), $\alpha \in]0; 1[$ pour une quantité E est un intervalle aléatoire I , tel que

$$\mathbb{P}(E \in I) \geq 1 - \alpha.$$

- Un *intervalle de confiance asymptotique* de niveau de confiance $1 - \alpha$ (ou de niveau d'erreur α), $\alpha \in]0; 1[$ pour une quantité E est une suite d'intervalles aléatoires $(I_n)_{n \geq 1}$, telles que

$$\liminf_{n \rightarrow \infty} \mathbb{P}(E \in I_n) \geq 1 - \alpha.$$

Toujours sous les hypothèses du théorème ci-dessus, notons

$$\epsilon_n = \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X].$$

On cherche à approcher $\mathbb{E}[X]$ à 0,01 près avec une confiance de 95% (on cherche à construire un intervalle de confiance asymptotique). C'est à dire que l'on veut

$$\mathbb{P}(|\epsilon_n| \geq 0,01) \leq 0,05, \text{ ce qui est équivalent à } \mathbb{P}(|\epsilon_n| \leq 0,01) \geq 0,95.$$

Calculons :

$$\begin{aligned} \mathbb{P}(|\epsilon_n| \leq 0,01) &= \mathbb{P} \left(-0,01 \leq \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X) \leq 0,01 \right) \\ &= \mathbb{P} \left(-0,01 \frac{\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X) \right) \leq 0,01 \frac{\sqrt{n}}{\sigma} \right) \\ (\text{pour } n \text{"assez grand"}) &\approx \int_{-0,01 \frac{\sqrt{n}}{\sigma}}^{+0,01 \frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \\ (\text{par symétrie de } \mathcal{N}(0, 1)) &= 2 \int_{-\infty}^{+0,01 \frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt - 1. \end{aligned}$$

On veut donc n tel que $\int_{-\infty}^{+0,01\frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \geq 0,975$. On regarde dans une table de la loi normale et on voit qu'il suffit pour cela d'avoir $0,01\sqrt{n}/\sigma \geq 1,96$, c'est à dire

$$n \geq \frac{(1,96 \times \sigma)^2}{0,01^2}.$$

On remarque que le nombre de tirages nécessaire n pour atteindre un certain niveau d'erreur avec une certaine confiance est une fonction linéaire de σ^2 .

On dit que σ^2 est la variance de la méthode. Sous les hypothèses du théorème 2.3, la variance de l'erreur $\frac{\bar{X}_1 + \dots + \bar{X}_n}{n} - \mathbb{E}(X_1)$ est asymptotiquement σ^2/n . La variance σ^2 est aussi la variance des variables i.i.d. dont on fait la moyenne empirique.

Le calcul précédent se généralise. Pour cela, on rappelle la définition de la fonction quantile de la loi normale $\mathcal{N}(0, 1)$:

$$\forall t \in [0; 1], Q(t) = \inf \left\{ x \in \mathbb{R}, \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \geq t \right\}$$

On a alors :

PROPOSITION 2.4. *Soit $m \in \mathbb{R}$ une quantité à estimer et X une variable aléatoire \mathbb{L}^2 de moyenne m et de variance $\sigma^2 > 0$. Si $(X_n)_{n \in \mathbb{N}}$ est une suite de v.a.i.i.d. de même loi que X , alors pour tout $\alpha > 0$, l'intervalle*

$$I_{\alpha,n} := \left[\bar{X}_n - \sqrt{\frac{\sigma^2}{n}} Q\left(1 - \frac{\alpha}{2}\right); \bar{X}_n + \sqrt{\frac{\sigma^2}{n}} Q\left(1 - \frac{\alpha}{2}\right) \right]$$

est un intervalle de confiance asymptotique pour m .

Dans la pratique, la variance σ^2 peut ne pas être connue, mais on peut l'estimer par une méthode de Monte-Carlo.

LEMME 2.5 (Estimateur de la variance). *Si X_1, X_2, \dots sont i.i.d. avec $\mathbb{E}[X_1^2] < +\infty$ et $\sigma^2 = \mathbb{V}[X_1]$, alors*

$$\tilde{\sigma}_n^2 = \frac{(X_1 - X_2)^2 + (X_3 - X_4)^2 + \dots + (X_{2n-1} - X_{2n})^2}{2n} \xrightarrow[n \rightarrow +\infty]{p.s.} \sigma^2.$$

DÉMONSTRATION. Les variables $(X_1 - X_2)^2, (X_3 - X_4)^2, \dots$ sont i.i.d. avec

$$\mathbb{E}[(X_1 - X_2)^2] = \mathbb{E}[X_1^2 + X_2^2 - 2X_1 X_2] = 2\mathbb{E}[X_1^2] - 2\mathbb{E}[X_1]^2 < +\infty.$$

Donc le résultat découle de la loi des grands nombres. \square

Exemples.

Cas facile. Soit U une variables aléatoire dans \mathbb{R}^d ($d \in \mathbb{N}^*$) et $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable. On cherche à calculer $p = \mathbb{P}(f(U) \leq \lambda)$ pour un certain λ (p est donc dans $[0; 1]$). Soit $X = \mathbb{1}_{f(U) \leq \lambda}$. Alors $p = \mathbb{E}[X]$. En faisant des tirages i.i.d. X_1, X_2, \dots de même loi que X , on peut donc approcher p par

$$\forall n \in \mathbb{N}^*, \quad p_n = \frac{X_1 + \dots + X_n}{n} \approx p.$$

Nous calculons $\sigma^2 = \mathbb{V}(X) = p(1-p)$. Si nous voulons une erreur $|p - p_n|$ inférieure à 0,01 avec une confiance à 0,95% (comme dans l'exemple ci-dessus), il suffit de prendre

$$n \geq \frac{(1,96 \times \sigma)^2}{0,01^2}.$$

Pour tout $p \in [0; 1]$, $p(1-p) \leq 1/4$, donc il suffit de prendre

$$n \geq \frac{1,96^2}{0,01^2} \times \frac{1}{4}.$$

Cas difficile. On cherche à calculer $\mathbb{E}[e^{\beta Z}]$ avec $Z \sim \mathcal{N}(0, 1)$. Nous savons que

$$\begin{aligned}\mathbb{E}[e^{\beta Z}] &= \int_{-\infty}^{+\infty} e^{\beta x} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\beta)^2 + \frac{\beta^2}{2}} dx \\ &= e^{\beta^2/2}.\end{aligned}$$

Si nous tirons Z_1, Z_2, \dots i.i.d. de même loi que Z , nous aurons au bout de n tirages, l'intervalle de confiance asymptotique :

$$\forall \alpha \in]0; 1[, \quad I_{\alpha, n} := \left[\bar{Z}_n - \sqrt{\frac{\sigma^2}{n}} Q(1 - \frac{\alpha}{2}); \bar{Z}_n + \sqrt{\frac{\sigma^2}{n}} Q(1 - \frac{\alpha}{2}) \right]$$

avec $\sigma^2 = \mathbb{V}[e^{\beta Z}]$ et Q la fonction quantile de la loi normale centrée réduite. Un calcul similaire à celui que nous venons de faire nous donne : $\sigma^2 = e^{2\beta^2} - e^{\beta^2}$. Cela donne une erreur relative largeur de l'intervalle de confiance sur moyenne estimée :

$$\frac{\sigma}{\mathbb{E}[e^{\beta Z}] \sqrt{n}} 2Q(1 - \frac{\alpha}{2}).$$

Nous avons $\frac{\sigma}{\mathbb{E}[e^{\beta Z}]} = \sqrt{e^{\beta^2} - 1}$. Par exemple, si $\beta = 5$ et $\alpha = 0.05$, soit $Q(1 - \alpha/2) \approx 1.96$, et si nous voulons une erreur relative de l'ordre de 1, il nous faut prendre n tel que

$$\frac{\sqrt{e^{\beta^2} - 1}}{\sqrt{n}} \times 2 \times 1.96 \approx 1,$$

c'est à dire n de l'ordre de 2.8×10^{11} , ce qui est difficilement réalisable dans la pratique. C'est pourquoi il est important de réduire la variance.

2.2.3. Comparaison avec les méthodes déterministes. Replaçons-nous dans le cadre de l'exemple 2.1. Nous pouvons également approcher l'intégrale en question par des sommes de Riemann :

$$\int_{[0;1]^d} f(x) dx = \lim_{n \rightarrow +\infty} \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right).$$

Pour comparer cette approximation avec l'approximation de Monte-Carlo, il convient de comparer les temps de calcul nécessaires pour atteindre une précision fixée, ou de comparer les précisions atteintes pour des temps de calcul égaux.

LEMME 2.6. *Si f est Lipschitz sur $[0; 1]^d$, alors*

$$\left| \int_{[0;1]^d} f(x) dx - \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| \leq \frac{C}{n}$$

pour une certaine constante C .

DÉMONSTRATION. La propriété lipschitz veut dire : il existe une constante K telle que, pour tout $x, y \in [0; 1]^d$, $|f(x) - f(y)| \leq K\|x - y\|$. Nous avons donc

$$\begin{aligned}&\left| \int_{[0;1]^d} f(x) dx - \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| \\ &= \left| \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \cdots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} f(x) - f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) du_1 \dots du_d \right| \\ &\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \cdots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} \left| f(x) - f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| du_1 \dots du_d\end{aligned}$$

$$\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \dots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} K \frac{\sqrt{d}}{n} du_1 \dots du_d = K \frac{\sqrt{d}}{n}.$$

□

Pour calculer la somme $\sum_{1 \leq i_1, \dots, i_d \leq n} \dots$, nous devons exécuter d boucles emboîtées de longueur n (et nous sommes donc n^d termes). Nous dirons donc que le coût du calcul est d'ordre n^d . La précision est elle d'ordre n^{-1} . Si nous notons $N = n^d$ le coût du calcul, nous avons donc une précision en $N^{-1/d}$.

Dans le programme 2.1, le coût du calcul est d'ordre n (il y a n boucles) pour une précision d'ordre $n^{-1/2}$ (d'après le théorème 2.3, si $\mathbb{E}[X^2] < \infty$). Si nous notons $N = n$ le coût du calcul, nous avons donc une précision en $N^{-1/2}$.

REMARQUE 2.7.

- On voit donc que la méthode de Monte-Carlo n'est avantageuse numériquement qu'à partir de la dimension $d = 3$. Même si la plupart des exercices et exemples de ce livre sont donnés en dimension 1 par souci de simplicité, il convient de ne pas oublier ce fait dans la pratique
- Les hypothèses permettant d'appliquer l'une ou l'autre des méthodes sont différentes et moins contraignantes dans le cas Monte-Carlo.
- La méthode de Monte-Carlo est facile à implémenter, c'est ce qui fait son succès.
- Il existe des méthodes déterministes dites « quasi Monte-Carlo » qui sont plus rapides que les méthodes de Monte-Carlo simples (mais qui demandent plus d'hypothèses et qui sont plus difficiles à implémenter). Le lecteur intéressé pourra se reporter aux références pour en savoir plus.

2.3. Réduction de variance

Nous avons vu précédemment que si nous approchons une quantité $\mathbb{E}[X]$ par la méthode de Monte-Carlo, c'est à dire par une moyenne empirique $(X_1 + \dots + X_n)/n$, X_1, X_2, \dots i.i.d. de même loi que X , alors l'erreur $|\mathbb{E}[X] - \frac{X_1 + \dots + X_n}{n}|$ est d'ordre σ/\sqrt{n} , où $\sigma^2 = \mathbb{V}[X]$. Nous allons présenter ici des méthodes qui permettent d'écrire $\mathbb{E}[X] = \mathbb{E}[Y]$ avec $\mathbb{V}[Y] \leq \mathbb{V}[X]$. Ces méthodes sont dites « de réduction de variance ».

Échantillonage préférentiel (« importance sampling » en anglais).

Nous cherchons à calculer $\mathbb{E}[g(X)]$ avec X variable à valeurs dans \mathbb{R}^d , de densité f .

Pour toute densité \tilde{f} strictement positive sur $\{x \in \mathbb{R}^d / g(x)f(x) > 0\}$, nous pouvons écrire

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x)f(x) dx = \int_{\mathbb{R}^d} \frac{g(x)f(x)}{\tilde{f}(x)} \tilde{f}(x) dx = \mathbb{E}\left[\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right],$$

avec Y de densité \tilde{f} . Nous avons donc deux méthodes de Monte-Carlo pour approcher $\mathbb{E}[g(X)]$:

$$\mathbb{E}[g(X)] \approx \frac{g(X_1) + \dots + g(X_n)}{n} \quad \text{et} \quad \mathbb{E}[g(X)] \approx \frac{1}{n} \left(\frac{g(Y_1)f(Y_1)}{\tilde{f}(Y_1)} + \dots + \frac{g(Y_n)f(Y_n)}{\tilde{f}(Y_n)} \right),$$

avec des X_1, X_2, \dots i.i.d. de même loi que X , des Y_1, Y_2, \dots i.i.d. de même loi que Y . La deuxième méthode est plus intéressante que la première si

$$\mathbb{V}\left[\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right] \leq \mathbb{V}[g(X)].$$

Supposons $g \geq 0$ et choisissons $\tilde{f} : x \mapsto \frac{g(x)f(x)}{\mathbb{E}[g(X)]}$ (c'est bien une densité de probabilité), nous avons alors

$$\frac{g(Y)f(Y)}{\tilde{f}(Y)} = \mathbb{E}[g(X)].$$

Cette variable est donc constante égale à $\mathbb{E}[g(X)]$. Et, dans ce cas extrême, la variance obtenue avec la nouvelle loi est donc optimale car nulle. Mais choisir une telle densité n'est pas possible en

pratique En effet, il faudrait pour cela savoir simuler suivant la densité \tilde{f} , qui est définie à partir de la constante $\mathbb{E}[g(X)]$, inconnue par hypothèse.

La discussion ci-dessus nous donne néanmoins une idée d'une démarche à suivre pour réduire la variance :

- (1) Chercher une fonction \tilde{f}_1 proche de $|f \times g|$ et telle que l'on sache simuler suivant la densité $f = \tilde{f}_1 / \int_{\mathbb{R}^d} \tilde{f}_1(x) dx$.
- (2) Soit Y variable aléatoire de densité \tilde{f} . Comparer $\mathbb{V}[g(X)]$ et $\mathbb{V}\left[g(Y)f(Y)/\tilde{f}(Y)\right]$ (il faut en général approcher ces variables par un estimateur, voir par exemple le lemme 2.5 à ce sujet).

La fonction \tilde{f} s'appelle la fonction d'importance.

EXEMPLE 2.8. On cherche à calculer $I = \int_0^1 \cos(\pi x/2) dx$ par une méthode de Monte-Carlo. En interprétant I comme $\mathbb{E}[\cos(\pi U/2)]$ où $U \sim \mathcal{U}([0; 1])$, on peut approcher l'intégrale par

$$(2.2) \quad I \approx \frac{\cos(\pi U_1/2) + \cdots + \cos(\pi U_n/2)}{n},$$

avec U_1, U_2, \dots i.i.d. de loi $\mathcal{U}([0; 1])$.

Si nous reprenons les notations ci-dessus, nous avons $g(x) = \cos(\pi x/2)$ et $f(x) = 1$ (pour $x \in [0; 1]$). Soit $\tilde{f}_1(x) = 1 - x$, alors

$$\tilde{f}(x) = \frac{1-x}{(1/2)} = 2(1-x).$$

La fonction \tilde{f}_1 est, grossièrement, proche de $g \times f$. De plus, si $U \sim \mathcal{U}([0; 1])$ alors $1 - \sqrt{U}$ est de loi de densité \tilde{f} (Ce résultat peut être démontré par la méthode d'inversion de la fonction de répartition, voir chapitre 1). Donc nous savons simuler suivant la loi de densité \tilde{f} et pouvons approcher I par

$$I \approx \frac{\cos(\pi X_1/2)/(2 - 2X_1) + \cdots + \cos(\pi X_n/2)/(2 - 2X_n)}{n}$$

où X_1, \dots, X_n suivent la loi de densité \tilde{f} .

Dans le programme 2.2, nous estimons les variances des deux méthodes. Nous trouvons environ 1.10^{-1} pour la méthode de l'équation (2.2) ci-dessus et 7.10^{-3} pour la méthode d'échantillonage d'importance par \tilde{f} . La variance est donc réduite.

Variable de contrôle.

Là encore, on cherche à calculer $\mathbb{E}[f(X)]$ avec X une variable aléatoire en améliorant l'approximation de Monte-Carlo simple :

$$\mathbb{E}[f(X)] \approx \frac{f(X_1) + \cdots + f(X_n)}{n}$$

avec X_1, X_2, \dots i.i.d. de même loi que X (toujours pour « n grand »). Supposons que l'on sache calculer, efficacement, $\mathbb{E}[h(X)]$ pour une certaine fonction h , alors on peut aussi faire l'approximation

$$\mathbb{E}[f(X)] \approx \frac{(f(X_1) - h(X_1)) + \cdots + (f(X_n) - h(X_n))}{n} + \mathbb{E}[h(X)].$$

Il faut ensuite comparer les variances pour savoir quelle est la méthode la plus avantageuse. Dans le cas de la deuxième méthode, l'erreur est

$$\begin{aligned} \mathbb{E}[f(X)] - & \left(\frac{(f(X_1) - h(X_1)) + \cdots + (f(X_n) - h(X_n))}{n} + \mathbb{E}[h(X)] \right) \\ & = \mathbb{E}[f(X) - h(X)] - \frac{(f(X_1) - h(X_1)) + \cdots + (f(X_n) - h(X_n))}{n}. \end{aligned}$$

Elle est d'ordre de grandeur $\sqrt{\mathbb{V}[f(X) - h(X)]/n}$.

La méthode est donc la suivante :

Programme 2.2 Comparaison de variances

```

simu<-function(t)
{
  u=runif(1,0,1)
  return(1-sqrt(u))
}
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  u1=cos(pi*u/2)
  v1=cos(pi*v/2)
  s=s+(u1-v1)^2
}
s=s/(2*n)
cat("Variance de la première méthode : ",s)
s=0
for (i in 1:n)
{
  v=simu(1)
  v1=cos(pi*v/2)/(2*(1-v))
  w=simu(1)
  w1=cos(pi*w/2)/(2*(1-w))
  s=s+(v1-w1)^2
}
s=s/(2*n)
cat("Variance de la méthode d'échantillonage préférentiel : ",s)

```

- (1) Trouver une fonction h proche de f et telle que l'on sache calculer $\mathbb{E}[h(X)]$ (le fait que h soit proche de f devrait faire en sorte que $\mathbb{V}[f(X) - h(X)]$ est petite).
- (2) Estimer les variances $\mathbb{V}[f(X)]$ et $\mathbb{V}[f(X) - h(X)]$ et les comparer.

EXEMPLE 2.9. On chercher à calculer $I = \int_0^1 \exp(x^2)dx$ par une méthode de Monte-Carlo. Nous avons $I = \mathbb{E}\left[e^{U^2}\right]$ avec $U \sim \mathcal{U}([0; 1])$. Donc nous pouvons faire l'approximation

$$I \approx \frac{e^{U_1^2} + \cdots + e^{U_n^2}}{n},$$

avec U_1, U_2, \dots i.i.d., de loi U .

Avec les notations ci-dessus, nous avons $f(x) = \exp(x^2)$. Remarquons que $h : x \mapsto 1 + x^2$ est proche de f sur $[0; 1]$ (c'est le début du développement limité de f en 0). Nous savons calculer

$$\mathbb{E}[h(U)] = \int_0^1 1 + x^2 dx = 1 + \frac{1}{3} = \frac{4}{3}.$$

Nous pouvons faire l'approximation

$$I \approx \frac{\left(e^{U_1^2} - 1 - U_1^2\right) + \cdots + \left(e^{U_n^2} - 1 - U_n^2\right)}{n} + \mathbb{E}[h(U)].$$

Nous estimons les deux variances dans le programme 2.3. La variance est réduite d'un facteur 10 grâce à cette méthode.

Programme 2.3 Variable de contrôle.

```

n=5000
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  s=s+(exp(u*u)-exp(v*v))^2
}
s=s/(2*n)
cat("Variance de la première méthode : ",s)
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  s=s+(exp(u*u)-1-u*u-exp(v*v)+1+v*v)^2
}
s=s/(2*n)
cat("Variance de la méthode de variable de contrôle : ",s)

```

2.3.1. Variables antithétiques.

Supposons que l'on cherche à calculer une intégrale sur l'hypercube de dimension d : $I = \int_{[0;1]^d} f(u_1, \dots, u_d) du_1 \dots du_d$. Nous avons $I = \mathbb{E}[f(U)]$ avec $U = (U_1, \dots, U_d)$ vecteur de v.a.i.i.d. de loi $\mathcal{U}([0;1])$. Nous pouvons utiliser l'approximation

$$I \approx \frac{f(X^1) + \dots + f(X^n)}{n},$$

avec X^1, X^2, \dots i.i.d. de même loi que U .

De plus, on peut remarquer que $(1, \dots, 1) - U = (1 - U_1, \dots, 1 - U_d) \stackrel{\mathcal{L}}{=} U$. Ainsi,

$$I = \mathbb{E}\left[\frac{f(1 - U_1, \dots, 1 - U_d) + f(U)}{2}\right] = \mathbb{E}\left[\frac{f((1, \dots, 1) - U) + f(U)}{2}\right],$$

ce qui nous donne l'idée d'une deuxième approximation

$$I \approx \frac{1}{n} \sum_{i=1}^n \frac{f((1, \dots, 1) - X^i) + f(X^i)}{2}.$$

LEMME 2.10. *Sous les hypothèses ci-dessus et si $f(U)$ admet une variance finie, on a*

$$\mathbb{V}\left[\frac{f((1, \dots, 1) - U) + f(U)}{2}\right] \leq \mathbb{V}[f(U)].$$

(La variance est donc toujours réduite.)

DÉMONSTRATION. Pour simplifier les notations, on introduit $X = f((1, \dots, 1) - U)$ et $Y = f(U)$, ces deux variables ont donc même loi de moyenne $\mathbb{E}[f(U)]$ et de variance $\mathbb{V}[f(U)]$. On peut alors écrire en utilisant l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} \mathbb{V}\left[\frac{f((1, \dots, 1) - U) + f(U)}{2}\right] &= \frac{1}{4}\mathbb{V}[X + Y] = \frac{1}{4}(\mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}(X, Y)) \\ &\leq \frac{1}{4}(\mathbb{V}[X] + \mathbb{V}[Y] + 2\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}) \end{aligned}$$

Comme X et Y sont deux variables de même variance $\mathbb{V}[f(U)]$, on obtient bien le résultat. \square

2.3.2. Méthode de stratification.

On veut calculer une intégrale de la forme $I = \mathbb{E}[g(X)]$ avec X variable de densité f sur un ensemble D (par exemple par rapport à la mesure de Lebesgue). Cette écriture nous donne directement une première méthode de Monte-Carlo :

$$I \approx \frac{g(X_1) + \cdots + g(X_n)}{n},$$

avec X_1, X_2, \dots i.i.d. de même loi que X . L'erreur commise $I - \frac{g(X_1) + \cdots + g(X_n)}{n}$ est approximativement de loi $\mathcal{N}(0, \sigma^2/n)$ avec $\sigma = \mathbb{V}[g(X)]$.

Supposons que D est partitionné en D_1, D_2, \dots, D_m (ce que nous noterons $D = \bigsqcup_{1 \leq i \leq m} D_i$). Nous décomposons I en :

$$I = \sum_{i=1}^m p_i I_i$$

avec pour tout i , $I_i = \mathbb{E}[g(X)|X \in D_i]$, $p_i = \mathbb{P}(X \in D_i)$ (nous avons $\sum_{i=1}^m p_i = 1$). Nous supposons que nous savons simuler suivant la loi $\mathcal{L}(X|X \in D_i)$ pour tout i et que tous les p_i sont connus. Pour chaque i , nous pouvons approcher I_i par une méthode de Monte-Carlo à n_i tirages :

$$I_i \approx \widehat{I}_{i,n_i} = \frac{g(X_1^{(i)}) + \cdots + g(X_{n_i}^{(i)})}{n_i},$$

avec des variables $X_1^{(i)}, X_2^{(i)}, \dots$ i.i.d. $\sim \mathcal{L}(X|X \in D_i)$. Chacune de ces approximations a un coût numérique n_i (puisque'on fait n_i boucles pour calculer la somme ci-dessus). Nous en déduisons une deuxième approximation :

$$I \approx p_1 \widehat{I}_{1,n_1} + \cdots + p_m \widehat{I}_{m,n_m}.$$

L'erreur commise se décompose dans ce cas en une somme d'erreur :

$$I - (p_1 \widehat{I}_{1,n_1} + \cdots + p_m \widehat{I}_{m,n_m}) = \sum_{i=1}^m p_i (I_i - \widehat{I}_{i,n_i}).$$

Dans cette somme, chacun des termes est (approximativement) de loi $\mathcal{N}(0, p_i^2 \sigma_i^2 / n_i)$ avec

$$\begin{aligned} \sigma_i^2 &= \mathbb{V}[g(X)|X \in D_i] \\ &= \mathbb{E}[(g(X) - \mathbb{E}[g(X)|X \in D_i])^2 | X \in D_i] \\ &= p_i^{-1} \mathbb{E}[g(X)^2 \mathbb{1}_{\{X \in D_i\}}] - p_i^{-2} \mathbb{E}[g(X) \mathbb{1}_{\{X \in D_i\}}]^2. \end{aligned}$$

Donc l'erreur est (approximativement) de loi $\mathcal{N}(0, \sum_{i=1}^m \sigma_i^2 / n_i)$. Nous voulons que cette erreur soit la plus petite possible donc nous voulons minimiser la variance $\sum_{i=1}^m p_i^2 \sigma_i^2 / n_i$. Comme nous voulons pouvoir faire une comparaison avec la première méthode, nous ajoutons la contrainte $\sum_{i=1}^m n_i = n$.

LEMME 2.11. Soit $S = \{(x_1, \dots, x_m) \in \mathbb{R}_+^m : x_1 + \cdots + x_m = n\}$. La fonction $f : S \rightarrow \mathbb{R}_+$ définie par

$$\forall x = (x_1, \dots, x_m) \in S, \quad f(x) = \sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{x_i},$$

atteint son minimum en $\frac{n}{\sum_{i=1}^m p_i \sigma_i}$ ($p_1 \sigma_1, \dots, p_m \sigma_m$).

DÉMONSTRATION. Posons $g : x = (x_1, \dots, x_m) \in \mathbb{R}^m \mapsto x_1 + \cdots + x_m$. L'ensemble S est la surface $\{x \in \mathbb{R}^m / g(x) = n\}$ avec (plus précisément, l'intersection de cette surface avec \mathbb{R}_+^m). Nous commençons par chercher les points critiques de f (voir un cours d'optimisation pour les détails de la démonstration qui va suivre). Ce sont les points x de S tels que $\nabla f(x)$ est colinéaire à $\nabla g(x)$ (nous notons ∇ pour le gradient). Nous avons

$$\begin{aligned} \nabla f(x) &= \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_m}(x) \right) = \left(\frac{-p_1^2 \sigma_1^2}{x_1^2}, \dots, \frac{-p_m^2 \sigma_m^2}{x_m^2} \right), \\ \nabla g(x) &= (1, 1, \dots, 1). \end{aligned}$$

Donc nous cherchons $x \in S$ et $\lambda \in \mathbb{R}$ tels que, pour tout $i \in \{1, \dots, m\}$, $\frac{-p_i^2 \sigma_i^2}{x_i^2} = \lambda$. L'unique solution est $x_i = n \frac{p_i \sigma_i}{p_1 \sigma_1 + \dots + p_m \sigma_m}$, pour tout i . La fonction f est convexe (sur \mathbb{R}_+^m) donc ce point critique est un minimum. Comme il est l'unique point critique, c'est un minium absolu. \square

Comme nous voulons des n_i entiers, nous prenons $n_i = \left\lfloor n \frac{p_i \sigma_i}{p_1 \sigma_1 + \dots + p_m \sigma_m} \right\rfloor$, pour tout i (rappel : $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$). Dans la pratique, il faudra donc estimer les σ_i et les p_i (par une méthode de Monte-Carlo). Si on oublie les parties entières, la deuxième méthode est de variance

$$\sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{n_i} = \sum_{i=1}^m p_i^2 \sigma_i^2 \frac{(p_1 \sigma_1 + \dots + p_m \sigma_m)}{np_i \sigma_i} = \frac{(p_1 \sigma_1 + \dots + p_m \sigma_m)^2}{n}.$$

LEMME 2.12. *Avec les notations précédentes, nous avons*

$$\sigma^2 \geq (p_1 \sigma_1 + \dots + p_m \sigma_m)^2.$$

La variance est donc bien réduite par la stratification.

DÉMONSTRATION. Comme les D_i forme une partition de D ,

$$\begin{aligned} \mathbb{V}[g(X)] &= \mathbb{E}[g(X)^2] - \mathbb{E}[g(X)]^2 \\ &= \mathbb{E}\left[\sum_{i=1}^m g(X)^2 \mathbb{1}_{\{X \in D_i\}}\right] - \left(\sum_{i=1}^m p_i \mathbb{E}[g(X)|X \in D_i]\right)^2 \\ &= \sum_{i=1}^m \left\{ p_i \mathbb{E}[g(X)^2|X \in D_i] - p_i \mathbb{E}[g(X)|X \in D_i]^2 \right\} \\ &\quad + \sum_{i=1}^m \left\{ p_i \mathbb{E}[g(X)|X \in D_i]^2 \right\} - \left(\sum_{i=1}^m p_i \mathbb{E}[g(X)|X \in D_i]\right)^2 \end{aligned}$$

La convexité de la fonction $x \rightarrow x^2$ implique que

$$\sum_{i=1}^m p_i \mathbb{E}[g(X)|X \in D_i]^2 \geq \left(\sum_{i=1}^m p_i \mathbb{E}[g(X)|X \in D_i]\right)^2$$

Donc

$$\mathbb{V}(g(X)) \geq \sum_{i=1}^m p_i (\mathbb{E}[g(X)^2|X \in D_i] - \mathbb{E}[g(X)|X \in D_i]^2) = \sum_{i=1}^m p_i \sigma_i^2$$

Et, en utilisant encore la convexité de la fonction carré, on obtient $\sum_{i=1}^m p_i \sigma_i^2 \geq \sum_{i=1}^m (p_i \sigma_i)^2$, ce qui conclut la démonstration. \square

Exemple : Réseau de ponts. On s'intéresse au temps minimal moyen pour aller de A à D dans

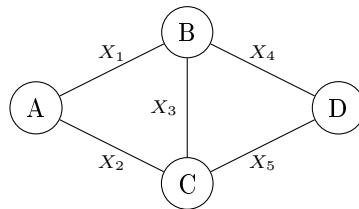


FIGURE 2.3.1. Réseau de ponts

le réseau indiqué sur la figure 2.3.1 lorsque le temps pour passer par une arête i est une variable aléatoire X_i . Ainsi, en posant

$$L = F(X_1, \dots, X_5) = \min(X_1 + X_4; X_1 + X_3 + X_5; X_2 + X_5; X_2 + X_3 + X_4),$$

on cherche à calculer $\mathbb{E}[L]$. On suppose, de plus, les X_i indépendantes et de loi respective $\mathcal{U}([0; a_i])$ avec $a = (1, 2, 3, 1, 2)$. Des calculs (fastidieux !) montrent que $\mathbb{E}[L] = \frac{1339}{1440} \approx 0.92986 \dots$

On va essayer de retrouver ce résultat en approchant $\mathbb{E}[L]$ par une méthode de Monte Carlo par partitionnement. Le domaine considéré est $D = \prod_{i=1}^5 [0; a_i]$; nous allons le partitionner en hyperrectangle de même dimension de côté de longueur a_i/K où $k = 4$ (les frontières étant de probabilité nulle, elles n'ont pas d'importance). Cela nous donne donc $m = K^5 = 4^5$ domaines D_i dans la partition. Chacun de ces domaines étant équiprobable, on a pour tout $i \leq m$, $p_i = 1/m$. On ne peut calculer directement les variances pour optimiser le choix des paramètres n_i . On peut alors soit faire le choix de n_i uniformes tous égaux à n/m comme dans l'algorithme 2.4, soit faire un premier Monte-Carlo uniforme avec un n relativement faible pour estimer les σ_i et utiliser ensuite les valeurs estimée $\hat{\sigma}_i$ pour définir les n_i :

$$n_i = n \frac{p_i \hat{\sigma}_i}{\sum_{j=1}^m p_j \hat{\sigma}_j} = n \frac{\hat{\sigma}_i}{\sum_{j=1}^m \hat{\sigma}_j},$$

le dernier résultat provenant de l'égalité de tous les p_i .

Programme 2.4 Méthode de stratification

```
# Nombre de strates et taille total de l'échantillon
K <- 4
m <- K^5
N <- 10^4

# Calcul de la taille de l'échantillon par strate en arrondissant vers le haut
n <- ceiling(N / m)

# Initialisation des vecteurs pour stocker les estimations
est <- numeric(m)
sdest<- numeric(m)
R <- 1:m

# Initialisation de la matrice W associée aux différents domaines
W <- matrix(0, nrow = m, ncol = 5)
W[, 1] <- R %% K
for (i in 2:5) {
  W[, i] <- (R %% K^i - (R %% K^(i - 1))) / K^(i - 1)
}

# Calcul des estimations par domaine
for (i in 1:m){
  V=c()
  for (j in 1:n) {
    # Générer une matrice V avec des valeurs aléatoires
    V <- c(V,F((W[i,] + runif(5)) / K))
  }
  est[i] <- mean(V)
  sdest[i]<- sd(V)
}

# Calcul de la moyenne des estimations
mest <- mean(est)
# Calculer le pourcentage de la Réduction de l'Erreur Relative
percRE <- sd(est) / (sqrt(n)*mest) * 100
```

2.3.3. Valeur moyenne ou conditionnement.

On cherche à calculer une intégrale de la forme

$$I = \mathbb{E}[g(X, Y)] = \int g(x, y) f(x, y) dx dy$$

où f est la densité du couple (X, Y) . Cette écriture nous donne l'idée d'une première approximation

$$I \approx \frac{1}{n} \sum_{i=1}^n g(X_i, Y_i)$$

avec des (X_i, Y_i) i.i.d. de même loi que (X, Y) . Soit $h(x) = \frac{1}{m(x)} \int g(x, y) f(x, y) dy$ avec $m(x) = \int f(x, y) dy$. Nous avons

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{x,y} h(x) f(x, y) dx dy \\ &= \int_{x,y} \frac{1}{\int_u f(x, u) du} \left(\int_v g(x, v) f(x, v) dv \right) f(x, y) dy \\ (\text{Fubini}) &= \int_{v,x} g(x, v) f(x, v) \frac{\int_y f(x, y) dy}{\int_u f(x, u) du} dx dv \\ &= \int_{v,x} g(x, v) f(x, v) dx dv \\ &= \mathbb{E}[g(X, Y)]. \end{aligned}$$

La fonction h est en fait l'espérance de $g(X, Y)$ conditionnée à la valeur de X :

$$\forall x, \quad h(x) = \mathbb{E}[g(X, Y) | X = x].$$

Supposons que nous savons calculer $h(x)$ pour tout x . Nous avons alors l'idée d'une deuxième approximation Monte-Carlo :

$$I \approx \frac{h(X_1) + \dots + h(X_n)}{n}.$$

LEMME 2.13. *Avec les notations précédentes, nous avons*

$$\mathbb{V}[h(X)] \leq \mathbb{V}[g(X, Y)].$$

Donc la méthode de la valeur moyenne réduit la variance.

DÉMONSTRATION. Nous savons que $\mathbb{E}[h(X)] = \mathbb{E}[g(X, Y)]$ donc pour comparer $\mathbb{V}[h(X)]$ et $\mathbb{V}[g(X, Y)]$, il suffit de comparer $\mathbb{E}[h(X)^2]$ et $\mathbb{E}[g(X, Y)^2]$. Nous avons :

$$\begin{aligned} \mathbb{E}[h(X)^2] &= \int_{x,y} h(x)^2 f(x, y) dx dy \\ &= \int_{x,y} \left(\frac{\int_v g(x, v) f(x, v) dv}{\int_u f(x, u) du} \right)^2 f(x, y) dx dy \\ (\text{Cauchy-Schwarz}) &\leq \int_{x,y} \frac{\int_v g(x, v)^2 f(x, v) dv}{\int_u f(x, u) du} f(x, y) dx dy \\ (\text{Fubini}) &= \int_{x,v} \frac{g(x, v)^2 f(x, v)}{\int_u f(x, u) du} \left(\int_y f(x, y) dy \right) dx dv \\ &= \int_{x,v} g(x, v)^2 f(x, v) dx dv \\ &= \mathbb{E}[g(X, Y)^2]. \end{aligned}$$

Donc $\mathbb{V}[h(X)] \leq \mathbb{V}[g(X, Y)]$. □

Exemple : On reprend l'exemple du réseau de ponts du paragraphe précédent. On rappelle qu'on cherche à calculer $\mathbb{E}[L]$ où

$$L = F(X_1, \dots, X_5) = \min(X_1 + X_4; X_1 + X_3 + X_5; X_2 + X_5; X_2 + X_3 + X_4),$$

les X_i étant indépendantes et de loi respective $\mathcal{U}([0; a_i])$ avec $a = (1, 2, 3, 1, 2)$.

On pose $Z_1 = \min(X_4; X_3 + X_5)$ et $Z_2 = \min(X_5; X_3 + X_4)$, alors $L = \min(Z_1, Z_2)$. Et on peut vérifier que

$$\forall (z_1, z_2) \in [0; 1] \times [0; 2], \mathbb{E}[L | Z_1 = z_1, Z_2 = z_2] = h(z_1, z_2)$$

avec

$$h(z_1, z_2) = \begin{cases} \frac{1}{2} + z_1 & \text{si } z_1 + 1 \leq z_2 \\ \frac{5}{12} + \frac{3z_1}{4} - \frac{z_1^2}{4} - \frac{z_1^3}{12} + \frac{z_2}{4} + \frac{z_1 z_2}{2} + \frac{z_1^2 z_2}{4} - \frac{z_2^2}{4} - \frac{z_1 z_2^2}{4} + \frac{z_2^3}{12} & \text{si } z_1 \leq z_2 \leq z_1 + 1 \\ \frac{1}{12} (5 - 3z_1^2 + 3z_2 - 3z_2^2 + z_1(9 + 6z_2)) & \text{si } z_2 \leq z_1 \end{cases}$$

L'algorithme 2.5 permet alors d'obtenir une valeur approchée de $\mathbb{E}[L]$.

Programme 2.5 Par conditionnement

```
# Fonction Zc
Zc <- function(u) {
  a <- c(1, 2, 3, 1, 2)
  X <- a*u
  Z <- c(min(c(X[4], X[3] + X[5])), min(c(X[5], X[3] + X[4])))
  return(Z)
}

# Taille de l'\`echantillon
N <- 10^4
# Initialisation d'un vecteur S pour stocker les résultats
S <- numeric(N)

for (i in 1:N) {
  # Calcul de la r\`ealisation de Z=(Z_1,Z_2)
  u <- runif(5)
  Z <- Zc(u)
  # Calcul de S_i=h(Z_1,Z_2)
  if (Z[2] > Z[1] + 1) {
    S[i] <- 1/2 + Z[1]
  } else if (Z[2] > Z[1]) {
    S[i] <- 5/12 + (3 * Z[1]) / 4 - Z[1]^2 / 4 - Z[1]^3 / 12 + Z[2] / 4 + (Z[1] * Z[2]) / 2
    + (Z[1]^2 * Z[2]) / 4 - Z[2]^2 / 4 - (Z[1] * Z[2]^2) / 4 + Z[2]^3 / 12
  } else {
    S[i] <- (5 - 3 * Z[1]^2 + 3 * Z[2] - 3 * Z[2]^2 + Z[1] * (9 + 6 * Z[2])) / 12
  }
}

# Calcul de la moyenne des résultats et de l'Erreur Relative
est <- mean(S)
RE <- 100*sd(S) / (sqrt(N)*est)
```

Chapitre 3

Simulation d'événements rares

On cherche à estimer $p = \mathbb{P}(A)$ dans le cas où p est très petit. Dans des domaines tels que l'aéronautique, l'énergie nucléaire, la finance, la santé, etc., il est en effet crucial de comprendre la fréquence d'événements rares mais potentiellement catastrophiques. L'estimation de telles probabilités permet d'évaluer la sécurité et la fiabilité des systèmes dans des conditions extrêmes.

Dans ce cas, la méthode de Monte Carlo classique fonctionne mal. En effet, si on dispose d'une suite de v.a.i.i.d. $(X_n)_{n \in \mathbb{N}}$ de même loi que $\mathbb{1}_A$, l'estimateur Monte-Carlo s'écrit :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Ainsi, $n\hat{p}_n \sim \mathcal{B}(n, p)$ et l'écart-type relatif vaut :

$$\frac{\sqrt{p(1-p)}}{\sqrt{np}} = \sqrt{\frac{1-p}{np}}$$

Pour $p \ll 1$, on obtient

$$\sqrt{\frac{1-p}{np}} \sim \frac{1}{\sqrt{np}}.$$

Pour avoir un ordre de grandeur de 1, il faut donc que n soit de l'ordre de $1/p$. Si on veut un intervalle de confiance de niveau $\alpha = 0.05$ avec une erreur relative inférieure à 0.1 pour un $p \sim 10^{-6}$, on doit par exemple choisir $n \sim 3.84 \times 10^8$. Il est donc utile d'utiliser des méthodes de réduction de variance pour obtenir une meilleure estimation de p .

Pour fixer un cadre mathématique, nous allons considérons des quantités sous la forme $p(x) = \mathbb{P}(A_x)$, $x > 0$, où, lorsque x tend vers l'infini, $p(x)$ tend vers 0. La variable x doit être pensée « grande ». On cherche donc à approcher $p(x)$ avec un estimateur $\tilde{p}(x)$ qu'on suppose non biaisé, ie : $\mathbb{E}[\tilde{p}(x)] = p(x)$, comme l'est l'estimateur obtenu par la méthode de Monte-Carlo classique.

Pour comparer la qualité d'estimation, nous introduisons deux types de contrôle de performance :

DÉFINITION 3.1.

— L'estimateur $\tilde{p}(x)$ est *d'erreur relative bornée* si

$$\limsup_{x \rightarrow \infty} \frac{\mathbb{V}[\tilde{p}(x)]}{p(x)^2} < \infty.$$

— L'estimateur $\tilde{p}(x)$ est *logarithmiquement efficace* si

$$\forall \epsilon > 0, \quad \lim_{x \rightarrow \infty} \frac{\mathbb{V}[\tilde{p}(x)]}{p(x)^{2-\epsilon}} = 0.$$

La proposition suivante explique la dénomination « logarithmiquement efficace » :

PROPOSITION 3.2. *Un estimateur $\tilde{p}(x)$ de $p(x)$ est logarithmiquement efficacessi*

$$\liminf_{x \rightarrow \infty} \frac{|\log \mathbb{V}[\tilde{p}(x)]|}{|\log p(x)^2|} \geq 1.$$

DÉMONSTRATION. Supposons $\tilde{p}(x)$ logarithmiquement efficace :

$$\forall \epsilon > 0, \quad \lim_{x \rightarrow \infty} \frac{\mathbb{V}[\tilde{p}(x)]}{p(x)^{2-\epsilon}} = 0.$$

Cela est équivalent à

$$\forall \epsilon > 0, \forall \delta > 0, \exists M > 0, \forall x \geq M, \mathbb{V}[\hat{p}(x)] \leq p(x)^{2-\epsilon}\delta$$

En prenant $\delta = 1$ et en passant au logarithme, on obtient

$$\forall \epsilon > 0, \exists M > 0, \forall x \geq M, \log \mathbb{V}[\hat{p}(x)] \leq (2 - \epsilon) \log p(x).$$

En divisant par $\log p^2(x) = 2 \log p(x)$, on a alors, comme $\log p(x) < 0$,

$$\forall \epsilon > 0, \exists M > 0, \forall x \geq M, \frac{|\log \mathbb{V}[\hat{p}(x)]|}{|\log p^2(x)|} \geq 1 - \epsilon/2.$$

Ce qui montre la première implication.

Supposons maintenant que $\hat{p}(x)$ vérifie :

$$\liminf_{x \rightarrow \infty} \frac{|\log \mathbb{V}[\hat{p}(x)]|}{|\log p(x)^2|} \geq 1$$

soit

$$\forall \epsilon > 0, \exists M > 0, \forall x \geq M, \frac{|\log \mathbb{V}[\hat{p}(x)]|}{|\log p^2(x)|} \geq 1 - \epsilon.$$

ou en multipliant par $\log p^2(x) < 0$ et en appliquant la fonction exponentielle, croissante,

$$\forall \epsilon > 0, \exists M > 0, \forall x \geq M, \mathbb{V}[\hat{p}(x)] \leq p^{2(1-\epsilon)}(x).$$

On obtient donc

$$\forall \epsilon > 0, \exists M > 0, \forall x \geq M, \frac{\mathbb{V}[\hat{p}(x)]}{p^{2-3\epsilon}(x)} \leq p^\epsilon(x).$$

Comme $\lim_{x \rightarrow \infty} p(x) = 0$ et que ϵ est un réel positif quelconque, cela nous montre bien la seconde implication. \square

On voit immédiatement qu'un estimateur à erreur relative bornée est logarithmiquement efficace. De plus, on utilisera parfois le majorant $\mathbb{E}[\hat{p}(x)^2]$ plutôt que $\mathbb{V}[\hat{p}(x)^2]$ pour simplifier les calculs.

Remarquons que pour $\hat{p}(x)$ l'estimateur donné par la méthode de Monte-Carlo classique, on a pour tout $\epsilon < 1$,

$$\frac{\mathbb{V}[\hat{p}(x)]}{p(x)^{2-\epsilon}} = \frac{1 - p(x)}{np(x)^{1-\epsilon}}$$

qui tend vers l'infini lorsque x tend vers 0. Ainsi l'estimateur n'est pas logarithmiquement efficace et donc n'est pas non plus d'erreur relative bornée.

Le problème de la simulation d'événements rares est complexe et nous allons nous intéresser ici à deux cas particuliers qui présentent déjà un nombre important d'application.

3.1. Simulation d'événements rares pour les sommes de vaid à queue légère.

On dit qu'une va X est à queue légère (à droite) si il existe $\theta > 0$ tel que

$$M_\theta := \mathbb{E}[e^{\theta X}] < \infty.$$

Si $\mathbb{E}[e^{\theta X}] = \infty$ pour tout $\theta > 0$, on dit que la variable est à queue lourde.

Les variables à queue légère ont ainsi une fonction de queue $x \rightarrow \mathbb{P}(X \geq x)$ qui idéocroît suffisamment vite lorsque x tend vers l'infini pour rendre intégrable des variables $e^{\theta X}$. Pour une variable à queue légère X , on pose

$$\theta_{\max} = \sup \{\theta > 0, M_\theta < \infty\}.$$

PROPOSITION 3.3. Soit X une variable à queue légère. Alors pour tout $\theta \in [0; \theta_{\max}[$, $M_\theta < \infty$.

DÉMONSTRATION. Soit $\theta < \theta_{\max}$. Par définition de θ_{\max} , il existe $\theta' \geq \theta$ tel que $M_{\theta'} < \infty$. On a alors

$$M_\theta = \mathbb{E}[e^{\theta X} \mathbb{1}_{X \geq 0}] + \mathbb{E}[e^{\theta X} \mathbb{1}_{X < 0}] \leq \mathbb{E}[e^{\theta' X} \mathbb{1}_{X \geq 0}] + 1 \leq M_{\theta'} + 1.$$

\square

Exemples :

- Les variables gaussiennes, quelles que soient la moyenne et la variance considérées, les variables exponentielles et les variables de loi uniforme sur un ensemble de mesure finie, sont à queue légère.
- Les variables suivant une loi de Pareto $\mathcal{P}(x_0, \alpha)$, $\alpha > 0, x_0 > 0$, de densité

$$f_{x_0, \alpha}(x) = \alpha \frac{x_0^\alpha}{x^{\alpha+1}} \mathbb{1}_{x \geq x_0}$$

sont à queue lourde.

Dans cette section, on considère des vajid X_1, \dots, X_n à queue légère de densité f . On note μ leur moyenne. Le but est d'estimer la probabilité

$$p_n = \mathbb{P}(S_n \geq n(\mu + \epsilon))$$

où $S_n = \sum_{i=1}^n X_i$ et ϵ . La Loi Forte des Grands Nombres montre que S_n/n converge p.s. et donc en probabilité vers μ . Ainsi, $\lim_{n \rightarrow \infty} p_n = 0$ et nous étudions ici la probabilité que la moyenne empirique s'écarte de la moyenne théorique. On pourrait adapter la méthode présentée, en considérant les variables $-X_i$, pour analyser la probabilité que la moyenne empirique soit notablement plus petite que l'espérance.

On va utiliser l'échantillonage préférentiel, introduit dans la section précédente, avec la famille de loi de densité $f_\theta(x) = \frac{e^{\theta x}}{M_\theta} f(x)$ où on rappelle que $M_\theta = \mathbb{E}[e^{\theta x}]$. Ces fonctions sont bien des densités de probabilité étant donné que la quantité M_θ est finie pour $\theta \in [0; \theta_{max}[$ avec $\theta_{max} > 0$. On pose également

$$\kappa_\theta = \log M_\theta.$$

PROPOSITION 3.4. Pour tout $\theta \in [0; \theta_{max}[$, $p_n = \mathbb{E}[Z_n^\theta]$ où

$$Z_n^\theta = e^{-\theta S_{\theta,n} + n\kappa_\theta} \mathbb{1}_{S_{\theta,n} \geq n(\mu + \epsilon)} \quad \text{et} \quad S_{\theta,n} = \sum_{i=1}^n X_{\theta,i},$$

les $X_{\theta,i}$ étant une suite de vajid de densité f_θ .

DÉMONSTRATION. On a pour tout $\theta \in [0; \theta_{max}[$,

$$\begin{aligned} p_n &= \mathbb{P}(S_n \geq n(\mu + \epsilon)) = \int_{\mathbb{R}^n} \mathbb{1}_{\sum_{i=1}^n x_i \geq n(\mu + \epsilon)} \prod_{i=1}^n f(x_i) dx_1 \dots dx_n \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{\sum_{i=1}^n x_i \geq n(\mu + \epsilon)} \prod_{i=1}^n \frac{f(x_i)}{f_\theta(x_i)} \prod_{i=1}^n f_\theta(x_i) dx_1 \dots dx_n \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{\sum_{i=1}^n x_i \geq n(\mu + \epsilon)} e^{-\theta \sum_{i=1}^n x_i + n\kappa_\theta} \prod_{i=1}^n f_\theta(x_i) dx_1 \dots dx_n \\ &= \mathbb{E}[Z_n^\theta] \end{aligned}$$

□

L'idée est de choisir θ de manière à avoir $\mathbb{E}[X_{\theta,1}] = \mu + \epsilon$. Ainsi l'événement $\{S_{\theta,n} \geq n(\mu + \epsilon)\}$ ne sera plus de probabilité faible mais de probabilité $1/2$ d'après le théorème central limite. Le fait que $p_n = \mathbb{E}[Z_n^\theta]$ est petit vient alors de la présence de l'exponentielle qui est une fonction beaucoup plus régulière que l'indicatrice.

On peut vérifier, grâce au théorème de dérivation sous l'intégrale, que

$$\mathbb{E}[X_{\theta,1}] = \frac{M'_\theta}{M_\theta} = \kappa'_\theta \quad \text{et} \quad \mathbb{V}[X_{\theta,1}] = \kappa''_\theta.$$

La fonction κ''_θ est donc strictement positive ce qui implique la convexité de κ et la stricte croissance de κ' . De plus, on a $\kappa'_0 = \mathbb{E}[X_1] = \mu$. Cela montre qu'il est toujours possible de trouver une solution à l'équation $\kappa'_\theta = \mu + \epsilon$ pour ϵ assez petit.

Exemple : Si X suit une loi $\mathcal{N}(\mu, 1)$, on peut calculer que $\theta_{max} = \infty$, $M(\theta) = e^{\mu\theta + \theta^2/2}$ et $\kappa(\theta) = \mu\theta + \frac{\theta^2}{2}$. Donc $\kappa'(\theta) = \mu + \theta$ et dans ce cas, $\theta^* = \epsilon$ et $f_{\theta^*} \sim \mathcal{N}(\mu + \epsilon, 1)$. Nous avons alors

$$Z_{n,\theta^*} = \mathbb{1}_{S_{\theta^*,n} \geq n(\mu+\epsilon)} e^{-\epsilon S_{\theta^*,n} + n(\mu\epsilon + \epsilon^2/2)}.$$

L'algorithme se résume ainsi :

- (1) Trouver θ^* tel que $\kappa'(\theta^*) = \mu + \epsilon$
- (2) Engendrer n réalisations indépendantes de loi $f_{\theta^*} : X_{\theta^*,1}, \dots, X_{\theta^*,n}$ et calculer $Z_n^{\theta^*}$ associé.
- (3) Répéter l'item précédent N fois et moyenner les $Z_{n,1}^{\theta^*}, \dots, Z_{n,N}^{\theta^*}$ pour obtenir une approximation de p_n :

$$\widehat{p}_{n,N} = \frac{1}{N} \sum_{i=1}^N Z_{n,i}^{\theta^*}.$$

THÉORÈME 3.5. *Si il existe $\theta^* \in]0; \theta_{\max}[$ tel que $\kappa'(\theta^*) = \mu + \epsilon$, la méthode de MC donnée par cet échantillonnage préférentiel est logarithmiquement efficace.*

DÉMONSTRATION. Posons $I = \theta^*(\mu + \epsilon) - \kappa_{\theta^*} = \theta^* \kappa'_{\theta^*} - \kappa_{\theta^*}$. Cette quantité est strictement positive par stricte convexité de κ .

- (1) *Montrons que $p_n \leq e^{-nI}$:*

$$\begin{aligned} p_n &= \mathbb{E} \left[\mathbb{1}_{\sum_{i=1}^n X_{i,\theta^*} \geq n(\mu+\epsilon)} e^{-\theta^* \sum_{i=1}^n X_{i,\theta^*} + n\kappa_{\theta^*}} \right] \\ &= e^{-nI} \mathbb{E} \left[\mathbb{1}_{\sum_{i=1}^n X_{i,\theta^*} \geq n(\mu+\epsilon)} e^{-\theta^* (\sum_{i=1}^n X_{i,\theta^*} - n(\mu+\epsilon))} \right] \\ &\leq e^{-nI} \end{aligned}$$

- (2) *Montrons que $\mathbb{E}[(Z_n^{\theta^*})^2] \leq e^{-2nI}$:*

$$\mathbb{E} \left[(Z_n^{\theta^*})^2 \right] = \mathbb{E} \left[\mathbb{1}_{\sum_{i=1}^n X_{i,\theta^*} \geq n(\mu+\epsilon)} e^{-2\theta^* \sum_{i=1}^n X_{i,\theta^*} + 2n\kappa_{\theta^*}} \right] \leq e^{-2nI}$$

- (3) *Montrons que pour tout $\delta > 0$, $\lim e^{(1+\delta)nI} p_n = \infty$: le TCL montre que*

$$Y_n = \frac{S_{n,\theta^*} - n(\mu + \epsilon)}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, \kappa''_{\theta^*})$$

Ainsi,

$$\begin{aligned} e^{(1+\delta)nI} p_n &= e^{\delta nI} \mathbb{E} \left[\mathbb{1}_{\sum_{i=1}^n X_{i,\theta^*} \geq n(\mu+\epsilon)} e^{-\theta^* (\sum_{i=1}^n X_{i,\theta^*} - n(\mu+\epsilon))} \right] \\ &= e^{\delta nI} \mathbb{E} \left[\mathbb{1}_{Y_n > 0} e^{-\sqrt{n} Y_n} \right] \geq e^{\delta nI - \theta^* \sqrt{n}} \mathbb{E} \left[\mathbb{1}_{Y_n \in]0;1]} \right] \end{aligned}$$

D'où le résultat car $\mathbb{E}[\mathbb{1}_{Y_n \in]0;1]}]$ converge vers une constante strictement positive d'après le TCL.

Nous pouvons maintenant démontrer le résultat du théorème : pour tout $\delta < 2$,

$$\frac{\mathbb{E}[(Z_n^{\theta^*})^2]}{p_n^{2-\delta}} \leq \frac{e^{-2nI}}{p_n^{2-\delta}} = \left(\frac{1}{e^{2/(2-\delta)nI} p_n} \right)^{2-\delta} \rightarrow 0$$

□

3.2. Simulation d'événements rares pour les sommes de vaid à queue lourde

On étudie maintenant le cas de somme de vaid X_1, \dots, X_n à queue lourde et on se restreint au cas où la loi des X_i est *continue et à variation régulière* i.e. en notant F la fonction de répartition commune des va, on suppose qu'il existe $\alpha > 0$ tel que

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

où pour tout $t > 0$, $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$. (typiquement un logarithme) Et on suppose de plus les variables p.s. positives. (par exemple : un coût, une durée, un niveau de radiation). Les lois de Pareto vues en section précédentes rentrent dans ce cadre.

On s'intéresse toujours à la probabilité que la somme $S_n = \sum_{i=1}^n X_i$ soit grande, $p_n(x) = \mathbb{P}(\sum_{i=1}^n X_i \geq x)$. Néanmoins, on ne prend plus maintenant comme valeur de référence la moyenne des X_i , celle-ci n'est même pas nécessairement définie pour les variables à queue lourde. Ici, c'est le paramètre x qui va tendre vers l'infini. L'idée est que, dans ce cas, contrairement au cas des v.a. à queue légère, la somme est grande essentiellement suite à l'a contribution d'une seule variable.

On va utiliser ici une approche Monte-Carlo par conditionnement. Une première idée pourrait être de conditionner par les $n - 1$ premières valeurs :

$$\tilde{Z}(x) = \mathbb{P}(S_n \geq x | X_1, \dots, X_{n-1}) = 1 - F(x - S_{n-1})$$

Mais cela fonctionne mal car

$$\mathbb{E}[\tilde{Z}^2(x)] \geq \mathbb{E}[(1 - F(x - S_{n-1}))^2 \mathbb{1}_{X_1 > x}] \geq 1 - F(x)$$

(car $F(y) = 0$ si $y < 0$). L'idée est alors de partitionner l'espace selon la plus grande valeur des X_1, \dots, X_n . On pose donc $M_n = \max_{1 \leq i \leq n} X_i$. on peut remarquer que par symétrie :

$$p_n(x) = n\mathbb{P}(S_n > x, M_n = X_n).$$

On pose alors l'estimateur

$$\begin{aligned} Z(x) &= n\mathbb{P}(S_n > x, M_n = X_n | X_1, \dots, X_{n-1}) \\ &= n\mathbb{P}(X_n > \max(x - S_{n-1}, M_{n-1}) | X_1, \dots, X_{n-1}) \\ &= n(1 - F(\max(x - S_{n-1}, M_{n-1})) \end{aligned}$$

On obtient alors l'algorithme suivant :

- (1) Produire $n - 1$ réalisations X_1, \dots, X_{n-1} de vaid de fonction de répartition F .
- (2) Calculer $Z(x) = n(1 - F(\max(x - S_{n-1}, M_{n-1})))$
- (3) Répéter N fois l'opération précédente indépendamment pour obtenir un échantillon de vaid $Z_1(x), \dots, Z_N(x)$ et approcher $p_n(x)$ par

$$\hat{p}_{N,n}(x) = \frac{1}{N} \sum_{j=1}^N Z_j(x).$$

THÉORÈME 3.6. *L'estimateur $Z(x)$ est à erreur relative bornée.*

DÉMONSTRATION. Si $M_{n-1} \leq x/n$, alors $x - S_{n-1} \geq x/n$. Ainsi, dans tous les cas,

$$\max(x - S_{n-1}, M_{n-1}) \geq x/n.$$

Cela montre que

$$\mathbb{E}[Z^2(x)] \leq n^2(1 - F(x/n))^2 = n^{2+2\alpha} \frac{L_{x/n}^2}{x^{2\alpha}}.$$

Par ailleurs,

$$p_n(x) = \mathbb{P}(S_n \geq x) \geq \mathbb{P}(\exists i \leq n, X_i \geq x) \geq \mathbb{P}(X_1 \geq x) = 1 - F(x) = \frac{L_x}{x^\alpha}.$$

On a donc

$$\frac{\mathbb{V}[Z(x)]}{p_x^2} \leq \frac{\mathbb{E}[Z^2(x)]}{p_x^2} \leq n^{2+2\alpha} \frac{L_x^2}{L_x^2} \xrightarrow{x \rightarrow \infty} n^{2+2\alpha}.$$

□

On pourrait également montrer avec quelques calculs supplémentaires que

THÉORÈME 3.7. *L'erreur relative tend vers 0 lorsque $x \rightarrow \infty$:*

$$\lim_{x \rightarrow \infty} \frac{\mathbb{V}[Z(x)]}{p_x^2} = 0.$$

Chapitre 4

Méthodes de Monte-Carlo par chaînes de Markov

On considère ici des chaînes de Markov sur un espace d'état E fini ou dénombrable. On parlera de chaîne de Markov (μ, P) pour indiquer que la chaîne a μ pour loi initiale et P pour matrice de transition, que nous nommerons également parfois *noyau markovien* ou *noyau de Markov*.

4.1. Simulation d'une chaîne de Markov

On souhaite simuler une réalisation de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de loi (μ, P) jusqu'à un temps $n \in \mathbb{N}$. Nous présentons deux méthodes mais l'idée est toujours de travailler récursivement pour utiliser la structure markovienne. Ainsi dans les deux cas, l'algorithme se présente ainsi :

- (1) Simuler une réalisation x_0 de X_0 selon la loi μ . (On suppose que cette étape n'est pas complexe).
- (2) Pour i allant de 0 à $n - 1$:
 - Simuler une réalisation x_{i+1} de X_{i+1} selon la loi $\mathbb{P}(X_{i+1} = \cdot | X_i = x_i) = P(x_i, \cdot)$
- (3) Renvoyer (x_0, \dots, x_n) .

Le problème est de savoir simuler selon la loi $\mathbb{P}(X_{i+1} = \cdot | X_n = x_i)$. La première méthode, dans le cas où l'espace d'état E est fini, est d'utiliser la matrice de transition P et de simuler à chaque étape une nouvelle réalisation x_{i+1} selon la loi $(p(x_i, y))_{y \in E}$. On peut pour cela utiliser la méthode de la fonction inverse et on peut faire de même pour la loi initiale $\mu = (\mu_y)_{y \in E}$. L'algorithme précédent devient alors si $E = \{e_1, \dots, e_m\}$:

- (1) Simuler une réalisation u de la loi $\mathcal{U}([0; 1])$ et poser

$$x_0 = \sum_{k=1}^m e_k \mathbb{1}_{\sum_{\ell=1}^{k-1} \mu_{e_\ell} \leq u < \sum_{\ell=1}^k \mu_{e_\ell}}$$

- (2) Pour i allant de 0 à $n - 1$:

— Simuler une réalisation u de la loi $\mathcal{U}([0; 1])$ (indépendamment de tout le reste) et poser

$$x_{i+1} = \sum_{k=1}^m e_k \mathbb{1}_{\sum_{\ell=1}^{k-1} p(x_i, e_\ell) \leq u < \sum_{\ell=1}^k p(x_i, e_\ell)}$$

- (3) Renvoyer (x_0, \dots, x_n) .

Exercice : La chaîne de Markov représentée dans la figure 4.1.1 admet pour matrice de transition :

$$P = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0.4 & 0 & 0 & 0.4 \\ 0 & 0 & 0.4 & 0 & 0.2 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.2 & 0 & 0.4 \\ 0.2 & 0 & 0.4 & 0 & 0 & 0.4 & 0 \end{pmatrix}$$

Écrire un algorithme permettant de simuler cette chaîne en prenant un état initial selon la loi uniforme sur $E = \{1, \dots, 7\}$.

Cette méthode peut se généraliser au cas où E est infini dénombrable. Néanmoins, elle fonctionne bien surtout si l'espace d'état est petit ou a minima si les lois de probabilités sont concentrées sur un petit nombre de valeurs.

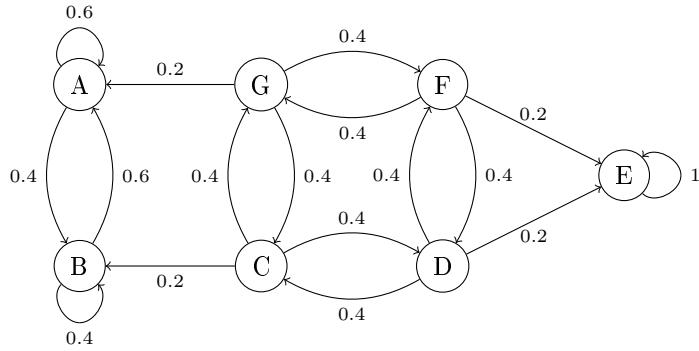


FIGURE 4.1.1. Graphe associée à la chaîne de Markov

Une autre méthode envisageable est d'utiliser la vision des chaînes de Markov comme des systèmes dynamiques aléatoires. On rappelle le théorème suivant :

THÉORÈME 4.1. *Soit $(X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans E et $(U_n)_{n \geq 1}$ une suite de v.a.i.i.d. à valeurs dans un espace F (non nécessairement dénombrable) et indépendantes de X_0 . On suppose de plus qu'il existe une fonction $f : E \times F \rightarrow E$ (mesurable) telle que*

$$\forall n \in \mathbb{N}, \quad X_{n+1} = f(X_n, U_{n+1}).$$

Alors le processus $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov.

Ainsi, si la chaîne de Markov est donnée sous cette forme, on peut utiliser l'algorithme suivant pour simuler une réalisation de (X_0, \dots, X_n) :

- (1) Simuler une réalisation u de la loi $\mathcal{U}([0; 1])$ et poser

$$x_0 = \sum_{k=1}^m e_k \mathbb{1}_{\sum_{\ell=1}^{k-1} \mu_{e_\ell} \leq u < \sum_{\ell=1}^k \mu_{e_\ell}}$$

- (2) Pour i allant de 0 à $n - 1$:

— Simuler une réalisation u de la loi $\mathcal{U}([0; 1])$ (indépendamment de tout le reste) et poser

$$x_{i+1} = f(x_i, u)$$

- (3) Renvoyer (x_0, \dots, x_n) .

Exemple : cette méthode peut par exemple être utilisée pour simuler des trajectoires de la marche aléatoire simple sur \mathbb{Z} .

Simulation d'une chaîne de Markov jusqu'à un temps d'arrêt T .

Il peut également être utile de simuler une chaîne jusqu'à une temps aléatoire T , cela est assez simple si l'événement $\{T \leq n\}$ dépend uniquement de X_0, \dots, X_n , c'est-à-dire si T est un temps d'arrêt pour la chaîne de Markov. On peut alors utiliser l'algorithme suivant :

- (1) Simuler une réalisation x_0 de X_0 selon la loi μ et initialiser $n = 0$.
- (2) Tant que $n < T$ et $n < n_{\max}$
 - (a) Simuler une réalisation x_{n+1} de X_{n+1} selon la loi $\mathbb{P}(X_{n+1} = \cdot | X_n = x_n)$
 - (b) Incrémenter n .
- (3) Renvoyer (x_0, \dots, x_n) et $T = n$.

Pour réaliser l'étape (2.a), on peut utiliser les méthodes proposées ci-dessus. On rajoute la condition $n < n_{\max}$ pour éviter de se retrouver dans une boucle infinie si jamais l'événement $\{T \leq n\}$ ne se réalisait jamais. Il faut alors bien faire attention que nous simulons en réalité une réalisation de la chaîne jusqu'au temps $\min(T, n_{\max})$. Par ailleurs, le temps T étant un temps d'arrêt, l'événement $\{T \leq n\}$ s'exprime toujours en fonction de X_0, \dots, X_n et le critère d'arrêt peut bien être implémenté.

Finalement, rappelons qu'il est également possible d'obtenir directement la loi de X_n en calculant la puissance n^e de la matrice P . Là encore, cette procédure fonctionne bien si l'espace d'état est petit ou si la matrice est creuse (avec un nombre important de 0 par ligne).

4.2. Rappels sur le comportement asymptotique des chaînes de Markov

On rappelle que pour une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de loi (μ, P) (ou pour une matrice de transition P , la notion ne dépend en fait pas du choix de la loi initiale), un état $x \in E$ est dit *récurrent positif* si l'espérance du temps de retour en x partant de x est finie :

$$\mathbb{E}_x [T_1^x] < \infty \text{ où } T_x^1 = \inf \{n > 0, X_n = x\}.$$

Cela implique pour ces états d'être récurrents.

PROPOSITION 4.2. *Le fait d'être récurrent positif est une propriété de classe : si x et y sont deux états appartenant à une même classe alors ils sont soit tous les deux récurrents positifs, soit aucun des deux n'est récurrent.*

THÉORÈME 4.3 (Théorème ergodique). *Soit P une matrice de transition irréductible (sur E dénombrable ou fini). Alors, il existe une probabilité P -invariante π ssi il existe un état récurrent positif. De plus, dans ce cas :*

- (1) π est l'unique probabilité P -invariante.
- (2) Tous les états sont récurrents positifs (pour P).
- (3) Si $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de loi initiale μ quelconque et de transition P alors

$$\frac{1}{n+1} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow +\infty]{p.s.} \int f(x) \pi(dx) = \sum_{x \in E} f(x) \pi(x),$$

pour toute fonction f qui est π -intégrable.

Ce théorème permet notamment d'expliquer le comportement asymptotique d'une chaîne de Markov irréductible fixée. Néanmoins, il peut être également utilisé dans l'autre sens. Supposons vouloir estimer une quantité I et que celle-ci puisse s'exprimer sous la forme d'une somme ou d'une série du type : $\sum_{x \in E} f(x) \pi(x)$ pour un certain ensemble E et une certaine mesure de probabilité π . Si on réussit à construire une chaîne de Markov irréductible $(X_n)_{n \in \mathbb{N}}$ sur E de probabilité invariante π , alors on peut approcher I par la somme

$$\frac{1}{n+1} \sum_{k=0}^n f(X_k).$$

Ici, la suite des $(X_n)_{n \in \mathbb{N}}$ n'est pas i.i.d. mais est une chaîne de Markov. On parle alors de méthode de Monte-Carlo par chaîne de Markov (« MCMC » pour « Monte Carlo Markov Chain » en anglais).

Nous verrons dans la prochaine section comment construire une telle chaîne.

DÉFINITION 4.4. Si π une probabilité sur E vérifie $\pi(x)P(x, y) = \pi(y)P(y, x)$ pour tous x, y , alors π est dite *symétrique* (par rapport à P) ou *P-symétrique*. On dit également que π est *réversible* (par rapport à P).

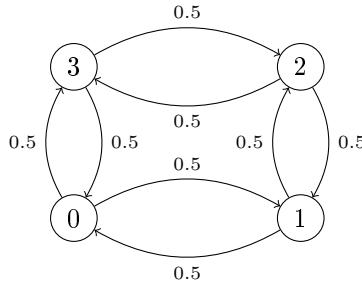
LEMME 4.5. *Si π est P -symétrique, alors elle est P -invariante.*

DÉMONSTRATION. La propriété de symétrie montre que, pour tout y ,

$$(\pi P)(y) = \sum_{x \in E} \pi(x)P(x, y) = \sum_{x \in E} \pi(y)P(y, x) = \pi(y).$$

□

Une question naturelle au vu du théorème ergodique est celle de la convergence en loi : sous les mêmes hypothèses, a-t-on convergence en loi de X_n vers la loi invariante π ? La réponse est non. On peut voir le problème qui apparaît avec la chaîne de Markov associée au graphe suivant :



On peut vérifier que cette chaîne admet pour unique probabilité invariante la loi uniforme sur $\{0, 1, 2, 3\}$: $\pi = \frac{1}{4}(1; 1; 1; 1)$. De plus, une récurrence simple montre que partant de l'état 0, on aura \mathbb{P}_0 -p.s. :

$$\forall n \in \mathbb{N}, \quad \begin{cases} X_{2n} \in \{0, 2\} \\ X_{2n+1} \in \{1, 3\} \end{cases}.$$

Ainsi, $\lim_{n \rightarrow \infty} \mathbb{P}_0(X_{2n} = 1) = 0 \neq \pi_1$ et il n'y a pas convergence en loi.

La convergence en loi est ici impossible de par le caractère périodique de la chaîne. Néanmoins, nous allons voir maintenant que si ce phénomène de périodicité n'est pas présent, il y a bien convergence en loi.

DÉFINITION 4.6. On dit qu'une chaîne de Markov irréductible $(X_n)_{n \in \mathbb{N}}$ de transition P est *apériodique* si

$$\forall x, y \in E, \exists n_{x,y} > 0, \forall n \geq n_{x,y}, P^n(x, y) > 0.$$

PROPOSITION 4.7. Une chaîne de Markov irréductible de transition P est apériodiquessi il existe un état $x \in E$ et un entier $n_x \in \mathbb{N}$ tels que $\forall n \geq n_x, P^n(x, x) > 0$.

DÉMONSTRATION. La première implication est évidente. Dans l'autre sens, supposons qu'il existe un état $x_0 \in E$ et un entier $n_{x_0} \in \mathbb{N}$ tels que :

$$\forall n \geq n_{x_0}, P^n(x_0, x_0) > 0.$$

Soit $x, y \in E$. La chaîne étant irréductible, il existe $p, q \in \mathbb{N}$, tels que $P^p(x, x_0) > 0$ et $P^q(x_0, y) > 0$. Alors, pour tout $n \geq n_{x_0}$,

$$P^{p+n+q}(x, y) \geq P^p(x, x_0)P^n(x_0, x_0)P^q(x_0, y) > 0.$$

Ainsi, on obtient le résultat attendu en prenant $n_{x,y} = n_{x_0} + p + q$. □

En particulier, une chaîne de Markov est apériodique si il existe un état $x \in E$ tel que $P_{x,x} > 0$, mais la condition n'est pas nécessaire.

EXEMPLE 4.8. la marche aléatoire simple symétrique sur \mathbb{Z} n'est pas apériodique. Mais si on ajoute une probabilité $\epsilon \in]0, 1[$ de rester en 0, elle le devient (cf Figure 4.2.1).

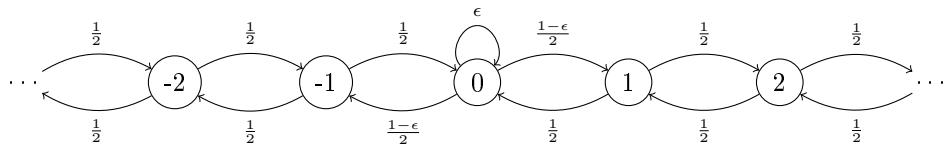


FIGURE 4.2.1. Modification apériodique de la marche aléatoire simple symétrique.

THÉORÈME 4.9. Soit $(X_n)_{n \in \mathbb{N}}$ une CM irréductible, récurrente positive de probabilité invariante π . Si la chaîne est apériodique alors pour toute loi initiale μ sur E ,

$$\forall x \in E, \lim_{n \rightarrow \infty} \mathbb{P}_\mu(X_n = x) = \lim_{n \rightarrow \infty} (\mu P^n)_x = \pi(x).$$

Ainsi, $X_n \xrightarrow[n \rightarrow +\infty]{loi} \pi$.

DÉMONSTRATION. Par couplage de $(X_n)_{n \in \mathbb{N}}$ avec une chaîne $(Y_n)_{n \in \mathbb{N}}$ de transition P mais de loi initiale π indépendante de $(X_n)_{n \in \mathbb{N}}$. Voir Théorème 1.8.3 de *Markov Chains*, Norris. \square

Le théorème suivant présente deux résultats qui ne seront au programme des examens. Ces résultats permettent de contrôler la vitesse de convergence vers la loi invariante dans le cas où l'espace d'états E est fini.

THÉORÈME 4.10. Soit P une matrice de transition irréductible, apériodique et admettant une probabilité invariante π sur un espace d'état E fini. Alors

- (1) Il existe deux réels $\alpha \in [0; 1[$ et $M > 0$ tels que

$$\forall (x, y) \in E^2, |P^n(x, y) - \pi(y)| \leq M\alpha^n.$$

- (2) Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de loi initiale μ quelconque et de transition P . Alors, pour toute fonction $f : E \rightarrow \mathbb{R}$ dans $\mathbb{L}^1(\pi)$,

$$\sqrt{n+1} \left(\frac{1}{n+1} \sum_{k=0}^n f(X_k) - \sum_{y \in E} \pi(y) f(y) \right) \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0, \sigma^2),$$

avec une variance $\sigma^2 \in [0; +\infty[$.

DÉMONSTRATION. Admis \square

REMARQUE 4.11. Le point 2 du théorème ci-dessus nous donne une vitesse de convergence pour le théorème ergodique (comme le théorème de la limite centrale nous donne la vitesse de convergence de la loi des grands nombres).

4.3. Algorithme de Hastings-Metropolis

On considère dans cette section une loi π sur un espace E , dénombrable ou fini, suivant laquelle on aimera simuler ou dont on voudrait calculer une intégrale $I = \int_E f(x)\pi(dx) = \sum_{x \in E} f(x)\pi(x)$. On suppose que pour tout $x \in E$, $\pi(x) > 0$, sinon il est toujours possible de réduire l'espace E . Nous allons voir ici une méthode pour construire une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ (ou plus exactement le noyau de transition de la chaîne) irréductible qui admet π pour loi invariante. De plus, la chaîne peut être choisie apériodique. Les théorèmes de la section précédente permettent alors, en simulant $(X_n)_{n \in \mathbb{N}}$, d'obtenir des approximations de réalisations de π ou de I . Nous appellerons π la *loi cible*.

On suppose disposé d'un noyau de transition *irréductible* Q sur E , qui a priori n'admet pas π pour loi invariante, et qui vérifie

$$\forall (x, y) \in E^2, Q(x, y) \neq 0 \Rightarrow Q(y, x) \neq 0.$$

On appelle Q le *noyau de proposition*. On se donne également une fonction $h :]0; \infty[\rightarrow]0; 1]$ vérifiant

$$\forall u > 0, h(u) = uh(1/u).$$

Par exemple, $h(u) = \min(u; 1)$ ou $h(u) = u/(1+u)$. La construction de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ admettant π pour probabilité invariante est basée sur un algorithme de rejet :

- (1) Nous prenons $X_0 = x_0 \in E$.
(2) Si $X_n = x$, nous simulons Y_{n+1} et U_{n+1} indépendants (et indépendants des simulations passées) avec

$$Y_{n+1} \sim Q(x, \cdot), U_{n+1} \sim \mathcal{U}([0; 1]).$$

La variable Y_{n+1} s'appelle une *proposition*. Posons,

$$\forall (x, y) \in E^2, \alpha(x, y) = h \left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right) \mathbb{1}_{Q(x, y) > 0}.$$

Alors :

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{si } U_{n+1} \leq \alpha(X_n, Y_{n+1}), \\ X_n & \text{sinon.} \end{cases}$$

Dans le cas où $U_{n+1} \leq \alpha(X_n, Y_{n+1})$, on dit qu'on accepte la proposition, et dans le cas contraire on dit qu'on refuse la proposition.

PROPOSITION 4.12. *La suite aléatoire $(X_n)_{n \in \mathbb{N}}$ construite ci-dessus est une chaîne de Markov irréductible sur E de transition P avec*

$$\begin{cases} P(x, y) = Q(x, y)\alpha(x, y) & \text{si } x \neq y, \\ P(x, x) = 1 - \sum_{y \neq x} P(x, y) & \text{si } y = x. \end{cases}$$

La loi π est P -réversible donc est l'unique loi invariante pour la chaîne $(X_n)_{n \in \mathbb{N}}$. Si, de plus, pour tout $u > 0$, $h(u) < 1$, elle est apériodique.

DÉMONSTRATION. Pour tout x , $\mathbb{P}(X_{n+1} = x | X_0, \dots, X_n) = \mathbb{P}(X_{n+1} = x | X_n)$ (d'après la construction ci-dessus). Donc $(X_n)_{n \in \mathbb{N}}$ est bien une chaîne de Markov. Calculons, pour tout $x \neq y$,

$$\begin{aligned} \mathbb{P}(X_{n+1} = y | X_n = x) &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(X_n, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, y) | X_n = x) \end{aligned}$$

Ainsi, comme U_{n+1} est indépendante de X_n et de Y_{n+1} , on a

$$P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x) = \mathbb{P}(Y_{n+1} = y | X_n = x) \mathbb{P}(U_{n+1} \leq \alpha(x, y)) = Q(x, y)\alpha(x, y).$$

Nous en déduisons que, pour tout $x \in E$,

$$P(x, x) = \mathbb{P}(X_{n+1} = x | X_n = x) = 1 - \sum_{y \neq x} P(x, y) = 1 - \sum_{y \neq x} Q(x, y)\alpha(x, y).$$

Montrons maintenant que P est irréductible. Soient x, y dans E , $x \neq y$. Par irréductibilité du noyau Q , il existe n dans \mathbb{N} et x_1, \dots, x_n dans E tels que

$$Q(x, x_1)Q(x_1, x_2) \dots Q(x_n, y) > 0.$$

On peut toujours supposer que les points x, x_1, \dots, x_n, y sont deux à deux distincts (quitte à raccourcir le chemin). On déduit alors de la définition de P que

$$P(x, x_1)P(x_1, x_2) \dots P(x_n, y) > 0$$

c'est-à-dire que P est irréductible.

La chaîne admet donc au plus une probabilité invariante. Montrons que π est P -réversible et est donc cette unique probabilité : pour tout $x \neq y$,

$$\begin{aligned} \pi(x)P(x, y) &= \pi(x)Q(x, y)\alpha(x, y) \\ &= \pi(x)Q(x, y)h\left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right) \\ &= \pi(y)Q(y, x)h\left(\frac{\pi(x)Q(x, y)}{\pi(y)Q(y, x)}\right), \end{aligned}$$

car la fonction h vérifie pour tout $u > 0$, $h(u) = uh(1/u)$. Cette dernière égalité nous dit donc bien que $\pi(x)P(x, y) = \pi(y)P(y, x)$. Donc π est symétrique par rapport à P . D'après le lemme 4.5, nous avons ainsi $\pi P = \pi$.

Il nous reste à montrer que P est apériodique si pour tout $u > 0$, $h(u) < 1$. Mais dans ce cas, par définition de α , on a pour tout x, y , $\alpha(x, y) < 1$. Ainsi,

$$\sum_{y \neq x} P(x, y) < \sum_{y \neq x} Q(x, y) = 1$$

et donc $P(x, x) > 0$ pour tout $x \in E$. Cela implique bien que la chaîne de Markov est apériodique. \square

EXEMPLE 4.13. Soit Q le noyau de la marche aléatoire simple symétrique :

$$\forall x \in \mathbb{Z}, Q(x, x+1) = Q(x, x-1) = \frac{1}{2}$$

et π une loi sur \mathbb{Z} telle que $\pi_x = C \exp(-\sqrt{|x|})$ pour tout x (C est alors la constante telle que $\sum_{x \in \mathbb{Z}} \pi_x = 1$). Le programme 4.1 fait une simulation d'une marche de Metropolis de noyau de proposition Q et de loi-cible π . On remarque qu'il n'est pas nécessaire de connaître C pour implémenter l'algorithme.

Programme 4.1 Chaîne de Metropolis

```
n=100
f<-function(k)
{
  return(exp(-sqrt(abs(k))))
}
liste=c()
x=0
for (k in 1:n)
{
  liste=c(liste,x)
  v=runif(1,0,1)
  if (v<0.5)
  { y=x+1 }
  else
  { y=x-1 }
  u=runif(1,0,1)
  alpha=min(1,f(y)*0.5/(f(x)*0.5))
  if (u<alpha)
  { x=y }
}
```

4.4. Algorithme de Metropolis simple

Si $Q(x, y) = Q(y, x)$ pour tout x, y (on dit que le noyau Q est *symétrique*). Dans la version originale de l'algorithme de Metropolis, le noyau Q est symétrique. Dans ce cas, α se simplifie en $\alpha(x, y) = h\left(\frac{\pi_y}{\pi_x}\right)$ pour tout $x, y \in E$.

PROPOSITION 4.14. Si Q est un noyau de Markov irréductible, symétrique et si π est une probabilité non constante, alors la chaîne de Markov de Metropolis de noyau de proposition Q et de loi cible π est apériodique quelle que soit la fonction h .

DÉMONSTRATION. Pour prouver l'apéridicité de P , il suffit de montrer qu'il existe x tel que $P(x, x) > 0$. Supposons que $P(x, x) = 0$ pour tout x . Or, pour tout x ,

$$\begin{aligned} P(x, x) &= 1 - \sum_{y:y \neq x} P(x, y) \\ &= \sum_y Q(x, y) - \sum_{y:y \neq x} Q(x, y)\alpha(x, y) \\ &= Q(x, x) + \sum_{y:y \neq x} Q(x, y)(1 - \alpha(x, y)) \\ &= Q(x, x) + \sum_{y:x \neq y} Q(x, y) \left(1 - h\left(\frac{\pi_y}{\pi_x}\right)\right). \end{aligned}$$

Donc, si $P(x, x) = 0$,

$$-Q(x, x) = \sum_{y:y \neq x} Q(x, y) \left(1 - h\left(\frac{\pi_y}{\pi_x}\right) \right).$$

Fixons $x \in E$. Comme h est à valeur dans $[0; 1]$, les deux termes de l'équation ci-dessus sont de signes opposés, ils valent donc 0 tous les deux. Nous avons ainsi, pour tout $y \neq x$ tel que $Q(x, y) \neq 0$ (il en existe car Q est irréductible),

$$h\left(\frac{\pi_y}{\pi_x}\right) = 1.$$

Comme Q est symétrique, nous avons $Q(y, x) = Q(x, y)$ et donc $h(\pi_x/\pi_y) = 1$ également. Comme $h(\pi_x/\pi_y) = h(\pi_y/\pi_x)\pi_x/\pi_y$, cela montre que pour tous les x, y tels que $Q(x, y) > 0$, $\pi_x = \pi_y$. Comme Q est irréductible, cela implique que π est constante, ce qui est en contradiction avec les hypothèses.

Conclusion : Il existe x tel que $P(x, x) > 0$ et le noyau P est apériodique. \square