

## Chapitre 3

# Approximations Stochastiques

### 1. Introduction

La possibilité de simulation des réalisations des variables aléatoires peut s'avérer être un puissant outil de calcul. L'exemple le plus naturel est celui de l'utilisation de la loi des grands nombres pour le calcul d'intégrales : la moyenne des réalisations d'une v. a. simulée converge vers l'espérance, qui est une intégrale. Dans certaines situations les méthodes stochastiques sont en concurrence avec les méthodes d'analyse numériques « déterministe », dans d'autres, elles constituent le seul outil adapté. L'objectif premier de ce chapitre est de présenter quelques méthodes de base de calcul de certaines approximations stochastiques parmi les plus répandues. Des techniques plus récentes, comme filtrage particulaire ou l'algorithme de Hasting-Metropolis, sont également brièvement décrits.

### 2. Générations des variables aléatoires

Nous considérons le problème de génération des réalisations d'une variable aléatoire réelle dont la loi admet une densité  $f$  par rapport à la mesure de Lebesgue. Pour simplifier, nous dirons que l'on "simule  $f$ ".

#### 2.1 Fonction de répartition inversible

Supposons que l'on sache simuler des réalisations d'une variable notée  $U$ , de la loi uniforme sur  $[0,1]$  (possibilité offerte par la plupart des ordinateurs). Considérons une densité de probabilité sur  $R$  notée  $f$ , supposons que la fonction de répartition  $F$  correspondant à  $f$  est inversible. Alors la variable aléatoire  $V = F^{-1}(U)$  suit la loi donnée par  $f$ . En effet, la fonction de répartition de  $V = F^{-1}(U)$  s'écrit :

$$P[V \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = \int_0^{F(x)} dt = F(x) \quad (2.1)$$

À titre d'exemple, il est ainsi possible de simuler les lois de densités exponentielles.

#### 2.2 Loi de Gauss et lois associées

La loi de Gauss intervient fréquemment en applications ayant trait aux traitements des signaux ou des images. En effet, les divers "bruitages" sont souvent modélisés par des réalisations des variables gaussiennes. Cette modélisation est souvent justifiée par le théorème central limite selon lequel, grosso modo, une somme « grande » des quantités aléatoires indépendantes tend en loi vers une distribution Gaussienne. Cependant, même lorsque ce type de justification ne peut être mis en avant, la modélisation Gaussienne est quand même souvent utilisée à cause des facilités de calcul qu'elle offre. La loi de Gauss ne fait pas partie de la famille du paragraphe précédent; en effet, sa fonction de répartition n'est pas exprimable analytiquement. Cependant,

on peut montrer que pour deux variables indépendantes  $U_1, U_2$ , chacune étant de la loi uniforme sur  $[0,1]$ , la loi de la variable

$$X = \sqrt{-2\text{Log}(U_1)} \cos(2\pi U_2) \quad (2.2)$$

est la loi normale de moyenne 0 et variance 1. En conséquence

$$\sigma \sqrt{-2\text{Log}(U_1)} \cos(2\pi U_2) + m \quad (2.3)$$

suit la loi normale de moyenne  $m$  et de variance  $\sigma^2$ .

Notons que la possibilité de simuler une variable aléatoire gaussienne réelle implique la possibilité de simuler tout vecteur gaussien. En effet, soit  $X = (X_1, \dots, X_n)$  un vecteur gaussien et  $f$  la densité de sa loi sur  $R^n$ . Nous pouvons écrire

$$f(x) = f(x_1, \dots, x_n) = f(x_1) f(x_2|x_1) f(x_3|x_1, x_2) \dots f(x_n|x_1, x_2, \dots, x_{n-1}) \quad (2.4)$$

et il est connu que chacune des densités conditionnelles figurant dans le produit à droite de l'égalité (2.4) est une densité gaussienne. On simule ainsi les composantes du vecteur de proche en proche : on commence par  $X_1 = x_1$  en utilisant (2.3), ensuite on calcule la moyenne et la variance de  $f(x_2|x_1)$  et on simule  $X_2 = x_2$  (toujours par (2.3)), et ainsi de suite ...

Un certain nombre de lois classiques sont des lois des variables aléatoires liées aux lois normales de manière déterministe :  $V = g(X_1, \dots, X_n)$ , avec  $X = (X_1, \dots, X_n)$  un vecteur gaussien et  $g$  une application de  $R^n$  dans  $R$ . De telles lois sont immédiatement simulables à partir des simulations des lois gaussiennes.

### 2.3 Méthode des lois marginales

Il est parfois commode - voir indispensable - de considérer la loi à simuler comme étant la loi marginale d'un couple de variables aléatoires.

Considérons l'exemple de la loi de probabilité de la variable « taille » d'un individu pris au hasard dans une foule. L'individu peut être un homme ou une femme, les deux populations présentant des tailles distribuées selon les Gaussiennes  $f_H$  et  $f_F$ . On sait donc simuler la taille d'un homme choisi au hasard, ainsi que la taille d'une femme choisie au hasard. Comment simuler la taille d'un individu choisi dans une foule comportant des hommes et des femmes ? On introduit une variable aléatoire  $X$  prenant ses valeurs dans l'ensemble  $\Omega = \{H, F\}$ , avec  $P[X = H] = \Pi(H)$  (proportion des hommes dans la foule), et  $P[X = F] = \Pi(F)$ . Si  $Y$  est la variable aléatoire modélisant la taille d'un individu pris au hasard,  $f_H$  est la loi de  $Y$  conditionnelle à  $X = H$ , et  $f_F$  est la loi de  $Y$  conditionnelle à  $X = F$ . Il en résulte que la loi de  $(X, Y)$  est donnée par  $\Pi(x)f_x(y)$ , et donc loi de  $Y$  est la loi marginale de celle de  $(X, Y)$ , obtenue en sommant sur les  $x$  :  $f(y) = \Pi(H)f_H(y) + \Pi(F)f_F(y)$ . On voit que si l'on sait simuler  $(X, Y) = (x, y)$  on saura simuler  $Y = y$  (il suffit de ne regarder que  $Y = y$ ). Comment

simuler  $(X, Y) = (x, y)$  ? On simule d'abord  $X = x$  ; et ensuite  $Y = y$  (cette dernière simulation est faite selon  $f_H$  si  $x = H$  , et elle est faite selon  $f_F$  si  $x = F$  ).

C'est une démarche générale ; on sait parfois simuler la réalisation d'une variable  $Y$  conditionnellement à une autre variable  $X$  . Si l'on sait simuler  $X$  , on saura simuler  $Y$  ; en effet, la loi de  $Y$  est la loi marginale de la loi de  $(X, Y)$  , cette dernière étant simulable.

## 2.4 Méthodes d'acceptation-rejet

Considérons une densité à simuler  $f$  , une densité simulable  $g$  , et une constante connue  $M$  pour laquelle  $f(x) \leq Mg(x)$  .

Alors  $f$  peut être simulée de la manière suivante :

- (i) Simuler  $Z = z$  selon  $g$  , et  $U = u$  selon la loi uniforme sur  $[0,1]$  ;
- (ii) Prendre  $x = z$  si  $u \leq \frac{f(z)}{Mg(z)}$  ; retourner en (i) sinon.

Ainsi, si  $z_1, \dots, z_n$  est une suite des valeurs simulées selon  $g$  , certaines seront "acceptées" et d'autres seront "rejetées", d'où l'appellation de la méthode.

Notons la grande généralité de cette méthode. En particulier, la densité  $f$  peut être connue à un facteur multiplicatif près. Cependant, la probabilité d'acceptation d'une réalisation selon  $g$  étant  $\frac{1}{M}$  , on recherchera une fonction  $g$  simulable pour laquelle la majoration  $f(x) \leq Mg(x)$  est vraie pour  $M$  aussi petit que possible.

## 3. Intégration par la méthode de Monte Carlo

L'intégration par la méthode de Monte Carlo est fondée sur la loi des grands nombres : pour une densité de probabilité  $f$  sur  $R^n$  et  $h$  une application quelconque sur  $R^n$  telle que  $fh$  soit intégrable sur  $R^n$  nous avons :

$$\frac{h(X_1) + \dots + h(X_k)}{k} \xrightarrow[k \rightarrow +\infty]{} \int_{R^n} h(x)f(x)dx \quad (3.1)$$

où  $X_1, \dots, X_k$  sont des variables aléatoires indépendantes dont la loi commune admet  $f$  pour densité. La convergence (3.1) a lieu presque sûrement; de plus, lorsque les variables  $h(X_1), \dots, h(X_k)$  sont de carré intégrable on peut appliquer le théorème central limité et avoir des précision sur sa vitesse. La convergence (3.1) peut déjà être utilisée dans un grand nombre de situations; cependant, elle ne donne pas nécessairement des résultats plus intéressants que des méthodes d'intégration numériques.

### Remarque 3.1

Pour  $X_1 = x_1, \dots, X_k = x_k$  , la quantité  $(h(x_1) + \dots + h(x_k))/k$  peut être vue comme une intégrale par rapport à la mesure « empirique » définie par cette réalisation, cette dernière consistant à associer à chaque ensemble  $B \subset R^n$  la proportion des  $x_1, \dots, x_k$  qu'il contient.

Une des possibilités d'amélioration de la vitesse de convergence de (3.1) est de noter que pour une densité de probabilité quelconque  $g$  l'intégrale d'intérêt  $\int_{R^n} h(x)f(x)dx$  peut aussi s'écrire

$\int_{R^n} h(x)f(x)dx = \int_{R^n} \frac{h(x)f(x)g(x)}{g(x)}dx$ , et donc pour une suite  $X_1, \dots, X_k$  des variables aléatoires indépendantes dont la loi commune admet  $g$  pour densité, nous avons

$$\frac{1}{k} \left( \frac{h(X_1)f(X_1)}{g(X_1)} + \dots + \frac{h(X_k)f(X_k)}{g(X_k)} \right) \xrightarrow{k \rightarrow +\infty} \int_{R^n} h(x)f(x)dx \quad (3.2)$$

Il est ainsi possible de rechercher des densités  $g$  pour lesquelles la vitesse de la convergence vers  $\int_{R^n} h(x)f(x)dx$  est la plus intéressante. On a alors le résultat suivant :

### Proposition 3.1

La densité  $g^*(x) = \frac{|h(x)|f(x)}{\int_{R^n} |h(t)|f(t)dt}$  procure la variance minimum pour les estimateurs de type (3.2).

### Exemple 3.1

Considérons une suite de variables aléatoires réelles  $X_1, \dots, X_n, \dots$  markovienne, dont la loi admet, pour chaque  $n \in N$ , pour densité

$$f(x_1, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_2) \dots f(x_n|x_{n-1}) \quad (3.3)$$

Une telle suite est également dite « sans mémoire » ; en effet, si l'on observe  $X_{n-1} = x_{n-1}$ , la loi de  $X_n$  ne dépend que de  $x_{n-1}$  et l'information contenue dans  $x_1, \dots, x_{n-2}$ , qui est la mémoire de ce qui s'est passé avant l'instant  $n-1$ , n'a aucune influence sur le comportement de  $X_n$ . Si l'on suppose que  $n-1$  est l'instant présent, les instants  $1, \dots, n-2$  sont le passé, et l'instant  $n$  est le futur, on dit aussi que « le futur et le passé sont indépendants conditionnellement au présent ». Attention, si  $x_{n-1}$  n'est pas observé,  $X_n$  n'est plus indépendant de  $X_1, \dots, X_{n-2}$ .

Supposons que  $X_1, \dots, X_n, \dots$  n'est pas observable et l'on observe une suite  $Y_1, \dots, Y_n, \dots$ , avec pour tout  $n \in N$ , la loi de  $(Y_1, \dots, Y_n)$  conditionnelle à  $(X_1, \dots, X_n) = (x_1, \dots, x_n)$  est donnée par

$$f(y_1, \dots, y_n | x_1, \dots, x_n) = f(y_1 | x_1)f(y_2 | x_2) \dots f(y_n | x_n) \quad (3.4)$$

Le problème, admettant de multiples applications, est alors d'estimer  $X_1, \dots, X_n, \dots$  à partir de  $Y_1, \dots, Y_n, \dots$ . Une approche particulière consiste en calcul de  $f(x_n | y_1, \dots, y_n)$  à partir de  $f(x_{n-1} | y_1, \dots, y_{n-1})$  et  $y_n$ . Dans une telle approche (appelée « filtrage adaptatif ») on considère généralement deux étapes :

$$f(x_n|y_1, \dots, y_{n-1}) = \int_R f(x_{n-1}, x_n|y_1, \dots, y_{n-1}) dx_{n-1} = \int_R f(x_n|x_{n-1}) f(x_{n-1}|y_1, \dots, y_{n-1}) dx_{n-1} \quad (3.5)$$

$$f(x_n|y_1, \dots, y_{n-1}, y_n) = \frac{f(x_n, y_n|y_1, \dots, y_{n-1})}{f(y_n|y_1, \dots, y_{n-1})} = \frac{f(x_n|y_1, \dots, y_{n-1}) f(y_n|x_n)}{f(y_n|y_1, \dots, y_{n-1})} \quad (3.6)$$

Dans des cas simples, typiquement « linéaires et Gaussiens », (3.6) admet une solution analytique (célèbre « filtre de Kalman »). Dans le cas général on peut effectuer l'intégration dans (3.5) en utilisant des tirages stochastiques. De telles méthodes, dites « filtrage particulaire » en références aux points, ou « particules », obtenues par les tirages, sont actuellement très à la mode et font objet de nombreuses recherches.

## 4. Méthodes de Monte Carlo par Chaînes de Markov (MCMC).

### 4.1 Cas discret

La problématique abordée dans ce paragraphe est la même que celle du paragraphe 2 : génération des réalisations d'une variable aléatoire dont la loi admet une densité  $f$  par rapport une certaine mesure. La différence est que les variables considérées prennent leurs valeurs dans des espaces trop compliquées (souvent, simplement trop vastes) pour que les méthodes du paragraphe 2 puissent être appliquées. L'idée de base est de construire une chaîne de Markov homogène telle que

- (i) on sait simuler ses réalisations ;
- (ii) la loi de probabilité donnée par  $f$  est la loi marginale limite de la chaîne.

Nous nous intéressons dans ce premier sous-paragraphe aux chaînes de Markov à états discrets. On considère donc une suite de variables aléatoires  $X_1, \dots, X_n, \dots$  prenant leurs valeurs dans un espace fini des états  $\Omega = \{e_1, \dots, e_k\}$ . Pour chaque  $n \in N$ , la loi de  $(X_1, \dots, X_n)$  s'écrit

$$p(x_1, \dots, x_n) = p(x_1) p(x_2|x_1) p(x_3|x_2) \dots p(x_n|x_{n-1}) \quad (4.1)$$

où  $p(x_1)$  est une probabilité sur  $\Omega = \{e_1, \dots, e_k\}$ , et  $p(x_2|x_1), \dots, p(x_n|x_{n-1})$  sont des probabilités conditionnelles (dites aussi des « transitions »), qui sont données par des « matrices de transition » carrées de dimension  $k$ . Notons bien que nous utilisons la même lettre  $p$  pour simplifier l'écriture, mais les  $k-1$  matrices de transition définissant  $p(x_2|x_1), \dots, p(x_n|x_{n-1})$  sont, dans le cas général, différentes.

Supposons que ces matrices sont égales (la chaîne est dite « homogène ») à une matrice  $M$  (sur la ligne  $i$  de la matrice on met les probabilités de  $e_1, \dots, e_k$  conditionnelles à  $e_i$ ). La probabilité  $p(x_1, \dots, x_n)$  donnée par (4.1) est ainsi déterminée par  $p(x_1)$ , que nous noterons  $\Pi$ , et par la matrice  $M$ . Considérons le problème de l'évolution des lois de  $X_n$  lorsque  $n \in N$  varie. La loi de  $X_1$  est  $\Pi = (\Pi_1, \dots, \Pi_k)$ , la loi de  $X_2$  est  $p(x_2) = \sum_{1 \leq i \leq k} p(x_1 = e_i, x_2) = \sum_{1 \leq i \leq k} p(x_1 = e_i) p(x_2|x_1 = e_i)$ . Si l'on considère  $p(x_2)$  comme un

vecteur ligne à  $k$  composantes, on note qu'il peut s'écrire  $\Pi M$  (produit « à gauche »,  $\Pi$  étant un vecteur « ligne »). En faisant le même raisonnement (la loi de  $X_3$  est obtenue à partir de la loi de  $X_2$  de la même manière que la loi de  $X_2$  est obtenue à partir de celle de  $X_1$ ), on note que la loi de  $X_3$  est  $(\Pi M)M = \Pi M^2$ . De proche en proche, on peut affirmer que la loi de  $X_n$  est  $\Pi M^{n-1}$ .

Une question intéressante est de savoir si  $\Pi M^{n-1}$  converge vers une limite indépendante de  $\Pi$  et il existe différents résultats précisant des conditions suffisantes pour qu'il en soit ainsi. Pour l'application qui nous intéresse ici mentionnons les conditions suivantes :

### Proposition 4.1

Sous les conditions

- (i) les termes diagonaux de  $M$  sont non nuls ;
- (ii) pour tous  $e_i, e_j$ , la probabilité de passer de  $e_i$  à  $e_j$  en un temps fini est non nulle,

$\Pi M^{n-1}$  converge vers une limite indépendante de  $\Pi$ .

Notons également

### Proposition 4.2

Si  $\Pi M^{n-1}$  converge vers une limite  $L$  indépendante de  $\Pi$ , alors  $LM = L$  (la limite est un vecteur propre à gauche de la matrice de transition).

En utilisant ces deux propositions, il suffit donc de vérifier (i), (ii), et trouver un vecteur propre à gauche de  $M$ .

En Méthodes de Monte Carlo par Chaînes de Markov (MCMC) considérées ici le problème est le suivant. On dispose d'une loi de probabilité  $L$  trop compliquée pour pouvoir être simulée directement. On cherche alors à construire une chaîne de Markov homogène (de matrice de transition  $M$ ) vérifiant deux conditions suivantes :

- (i) les transition sont faciles à simuler ;
- (ii)  $L$  est la limite des lois des  $X_n$ .

Afin d'illustrer l'intérêt des chaînes de Markov en approximation stochastique considérons l'exemple suivant. Soit un ensemble de points disposés en carré de côté  $m$  (un « réseau » simple). Chaque point peut être, de manière aléatoire, 0 ou 1. L'ensemble des réalisations possibles d'un tel réseaux est de  $\Omega = \{0,1\}^{m \times m}$ . Comment simuler les réalisations d'un tel réseau ? On voit que très rapidement le cardinal de  $\Omega$  est trop grand pour utiliser les méthodes du de la section 1. En considérant les couples  $(s,t)$  de points voisins (horizontalement ou verticalement), soit une probabilité  $L$  définie sur  $\Omega = \{0,1\}^{m \times m}$  par

$$L(x) = \gamma \exp\left[\sum_{(s,t)} \varphi_{(s,t)}(x_s, x_t)\right] \quad (4.2)$$

La constante  $\gamma$  est inconnue et ne peut pas être calculée ; cependant, pour chaque point  $s$  la loi de  $X_s$  conditionnelle aux autres  $X_t$  est calculable : si  $V_s$  désigne le voisinage de  $s$ , on a

$$L(x_s | (x_t)_{t \in V_s}) = \frac{\exp\left[\sum_{t \in V_s} \varphi_{(s,t)}(x_s, x_t)\right]}{\exp\left[\sum_{t \in V_s} \varphi_{(s,t)}(x_s = 0, x_t)\right] + \exp\left[\sum_{t \in V_s} \varphi_{(s,t)}(x_s = 1, x_t)\right]} \quad (4.3)$$

L'idée est alors de balayer le carré (« ligne par ligne » par exemple) et faire un tirage, sur chaque point, selon la très simple loi conditionnelle (4.3). Ce faisant, on utilise des transitions dans l'espace  $\Omega = \{0,1\}^{m \times m}$  et on peut montrer que  $L$  est bien la loi limite d'une chaîne de Markov ainsi construite.

A titre d'exemple, on présente sur la Figure 1 des tirages effectués avec  $m=128$  et quarante balayages du carré, soit  $40 \times 128 \times 128$  tirages élémentaires. Malgré ce nombre élevé le tirage final, dont trois versions obtenues pour des paramètres différents sont visualisées sur la Figure 1, apparaît comme quasi instantané et prend moins de temps que le lancer d'une pièce de monnaie.

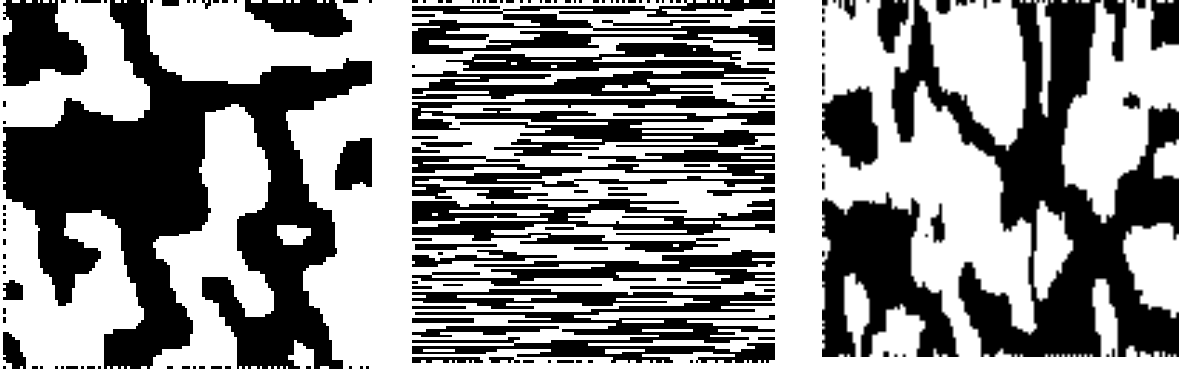


Figure 1

Réalisations de  $X = (X_s)_{s \in S}$ , avec  $S$  un carré de coté comportant  $m=128$  points, obtenues avec 40 balayages.

#### 4.2 Cas continu

Considérons le cas plus général où l'espace des états est  $R^m$ . Dans une chaîne de Markov vérifiant l'écriture (3.3) la matrice de transitions est remplacée par le noyau de transitions, qui est une application  $N: R^m \times B(R^m) \rightarrow [0,1]$  vérifiant

(N1) pour tout  $x \in R^m$ , l'application  $B \in B(R^m) \rightarrow N(x, B)$  est une probabilité sur  $B(R^m)$ ;

(N2) pour tout  $B \in B(R^m)$ , l'application  $x \in R^m \rightarrow N(x, B)$  est mesurable.

Cela veut dire, plus simplement, que chaque  $X_i = x_i \in R^m$  définit la loi de  $X_{i+1}$  sur  $R^m$ .

Comme précédemment, la loi d'une chaîne de Markov homogène est entièrement déterminée par la loi de la première variable  $X_1$  et le noyau de transitions.

Pour une densité de probabilité  $f$  trop compliquée pour pouvoir être simulée directement, le problème est donc de rechercher un noyau (qui devient une matrice de transition dans le cas discret) exploitable et tel que la loi limite soit la loi donnée par  $f$ . L'étude des chaînes de Markov où l'espace des états est  $R^m$  est notablement plus compliquée que celle des cas où cet espace est discret. Nous nous contentons d'énoncer deux manières de rechercher un noyau de transition convenable.

### - Algorithme de Hasting-Metropolis

Soit  $f$  la loi que l'on souhaite simuler. Soit une famille des densités conditionnelles  $q(y|x)$ , avec  $x$  et  $y$  dans  $R^m$ . On suppose que l'une de deux conditions suivantes est vérifiée :

- (i)  $q(\cdot|x)$  est disponible analytiquement (à une constante indépendante de  $x$  près);
- (ii)  $q$  est symétrique :  $q(y|x) = q(x|y)$ .

À  $X_n = x_n$  donné, on simule alors  $X_{n+1}$  de la manière suivante (ce qui donne le noyau recherché):

1. Générer  $Y_n = y_n$  selon  $q(y|x_n)$  ;

2. calculer

$$\rho(x_n, y_n) = \min\left[\frac{f(y_n)q(x_n|y_n)}{f(x_n)q(y_n|x_n)}, 1\right] \text{ et poser } x_{n+1} = \begin{cases} y_n & \text{avec la probabilité } \rho(x_n, y_n) \\ x_n & \text{avec la probabilité } 1 - \rho(x_n, y_n) \end{cases} \quad (4.4)$$

Un processus ainsi obtenu est bien une chaîne de Markov (la loi de  $X_{n+1}$  conditionnelle à  $X_1 = x_1, \dots, X_n = x_n$  ne dépend que de  $x_n$ ).

On montre alors que la loi marginale de la chaîne  $(X_n)$  converge bien vers la loi donnée par la densité  $f$ . Le principe de la démonstration est analogue aux démarches utilisées dans le cas discret ci-dessus : on montre d'abord qu'il existe une loi limite indépendante de la loi de  $X_1$  (régularité de la chaîne), et ensuite on montre que  $f$  est "invariante" (ce qui correspond aux vecteurs propres à gauche ci-dessus) au sens suivant :

$$\text{Pour tout } B \in \mathcal{B}(R^m), \text{ on a } \int_B f(t)dt = \int_{R^m} N(t, B)dt \quad (4.5)$$

### Remarque

Remarquons la très grande généralité de cet algorithme. Tout d'abord, la densité  $f$  peut n'être connue qu'à une constante multiplicative près (seul le rapport intervient dans (4.4)). Ensuite on a une grande latitude de choix dans les densités conditionnelles  $q$ .



**- Échantillonneur de Gibbs**

Supposons que l'on peut décomposer le vecteur  $x \in R^m$  en  $x = (x^1, \dots, x^k)$  tel que chaque  $X^i$  soit simulable selon sa distribution conditionnelle aux autres composantes :  $f_i(x^i | x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^k)$ . L'échantillonnage de Gibbs consiste alors en simulation de  $X_{n+1} = x_{n+1} = (x_{n+1}^1, \dots, x_{n+1}^k)$  à partir de  $X_n = x_n = (x_n^1, \dots, x_n^k)$  en  $k$  étapes :

1. Simuler  $x_{n+1}^1$  selon  $f_1(x^1 | x_n^2, \dots, x_n^k)$  ;
2. simuler  $x_{n+1}^2$  selon  $f_2(x^2 | x_{n+1}^1, x_n^3, \dots, x_n^k)$  ;
3. simuler  $x_{n+1}^3$  selon  $f_3(x^3 | x_{n+1}^1, x_{n+1}^2, x_n^4, \dots, x_n^k)$  ;
- .....
- k. simuler  $x_{n+1}^k$  selon  $f_k(x^k | x_{n+1}^1, x_{n+1}^2, \dots, x_{n+1}^{k-1})$

Remarquons qu'à l'étape  $j$  on doit prendre en compte les  $j-1$  nouvelles composantes simulées aux  $j-1$  étapes précédentes.

## Exercices

### Exercice 1

On considère la méthode d'acceptation-rejet, avec  $f$  la densité à simuler,  $g$  une densité simulable, et  $M$  une constante connue pour laquelle

$$f(x) \leq Mg(x) \quad (1)$$

La méthode consiste alors en:

- (i) Simuler  $Z = z$  selon  $g$ , et  $U = u$  selon la loi uniforme sur  $[0,1]$ ;
- (ii) Prendre  $x = z$  si  $u \leq \frac{f(z)}{Mg(z)}$ ; retourner en (i) sinon.

1. Remarquer que  $M \geq 1$  et montrer que  $P[U \leq \frac{f(z)}{Mg(z)}] = \frac{1}{M}$ ;
2. Montrer que la réalisation de  $X$  ainsi simulée suit bien la loi de densité  $f$ .
3. Commenter l'intérêt de choisir  $M$  aussi petit que possible.

### Exercice 2

On considère la densité de probabilité  $f$  définie sur  $R$  par

$$f(x) = \begin{cases} \frac{1}{2} \lambda \exp[-\lambda x] & \text{pour } x \geq 0 \\ \frac{1}{\sqrt{2\pi}\sigma^2} \exp[-\frac{x^2}{2\sigma^2}] & \text{pour } x \leq 0 \end{cases}$$

Proposer une méthode de simulation de  $f$ .

## Solutions des exercices

### Exercice 1

1. La quantité  $P[U \leq \frac{f(Z)}{Mg(Z)}]$  à calculer est de la forme  $E[\varphi(V, W)]$  (la probabilité d'un ensemble est l'espérance de sa fonction indicatrice).

De manière générale, pour  $\varphi$  à valeurs réelles  $E[\varphi(V, W)]$  est une intégrale d'une fonction de deux variables ; on effectue les calculs « en cascade ». Pour savoir par quelle variable commencer on se pose la question de savoir si le calcul était facile si l'une des variables était une constante. Par exemple, si  $E[\varphi(v, W)]$  est facile, on fera les calculs en utilisant la formule

$E[\varphi(V, W)] = E[E[\varphi(V, W)|V]]$ , où  $E[\varphi(V, W)|V = v]$  est l'espérance conditionnelle (qui peut être vue comme l'espérance selon la loi conditionnelle). Afin d'illustrer cette formule d'importance capitale en calcul des probabilités, supposons que  $(V, W)$  est à valeurs dans  $R^2$  et que sa loi admet une densité  $f_{(U, V)}$ . La densité  $f_V$  de la loi de  $V$  est alors

$f_V(v) = \int_R f_{(U, V)}(u, v) du$ , et la densité  $f_{U|V=v}$  de la loi de  $U$  conditionnelle à  $V = v$  est  $f_{U|V=v}(w) = \frac{f_{(U, V)}(u, v)}{f_V(v)}$ . La formule  $E[\varphi(V, W)] = E[E[\varphi(V, W)|V]]$  s'écrit alors

$$E[E[\varphi(V, W)|V]] = \int_R \int_R \varphi(v, w) f_{U|V=v}(w) dw f_V(v) dv = \int_R \int_R \varphi(v, w) \frac{f_{(U, V)}(u, v)}{f_V(v)} dw f_V(v) dv = \int_R \int_R \varphi(v, w) f_{(U, V)}(u, v) dw dv$$

ce qui le calcul classique de l'intégrale double  $E[\varphi(V, W)] = \int_{R^2} \varphi(v, w) f_{(U, V)}(u, v) dudv$

En revenant à l'exercice, on constate que pour  $Z = z$  fixé, le calcul de  $P[U \leq \frac{f(z)}{Mg(z)}]$  est

immédiat et donne  $\frac{f(z)}{Mg(z)}$ . On conditionnera donc par  $Z$ , ce qui donne

$$P[U \leq \frac{f(Z)}{Mg(Z)}] = E[P[U \leq \frac{f(Z)}{Mg(Z)}|Z]] = \int_R [\frac{f(z)}{Mg(z)}] f_Z(z) dz \quad (1)$$

Etant donné que  $f_Z = g$ , on a

$$\int_R [\frac{f(z)}{Mg(z)}] f_Z(z) dz = \int_R [\frac{f(z)}{Mg(z)}] g(z) dz = \int_R \frac{f(z)}{M} dz = \frac{1}{M}. \quad (2)$$

2. Cherchons la fonction de répartition  $P[X \leq t]$  de la loi de  $X$ . La variable  $X$  étant définie sur l'ensemble  $A = [U \leq \frac{f(Z)}{Mg(Z)}]$ , nous cherchons  $P[X \leq t|A] = \frac{P[(X \leq t) \cap A]}{P[A]}$ . Sachant que

$P[A] = \frac{1}{M}$  et  $X = Z$  sur  $A$ , nous avons  $P[X \leq t|A] = P[(Z \leq t) \cap (U \leq \frac{f(Z)}{Mg(Z)})|A]$ . Comme

précédemment, on conditionne par  $Z$  et on arrive à un calcul analogue à celui fait plus haut dans (1), (2) :

$$P[X \leq t|A] = P[(Z \leq t) \cap (U \leq \frac{f(Z)}{Mg(Z)})|A] = E[P[(Z \leq t) \cap (U \leq \frac{f(Z)}{Mg(Z)})|Z]] =$$

$$M \int_{-\infty}^t \frac{f(z)}{Mg(z)} f_z(z) dz = \int_{-\infty}^t \frac{f(z)}{g(z)} g(z) dz = \int_{-\infty}^t f(z) dz$$

qui est bien la fonction de répartition de la loi de  $X$ .

3. Un tirage a un coût en temps calcul ; dans la mesure où  $1 - \frac{1}{M}$  est la proportion des tirages rejetés, on a intérêt de choisir  $M$  aussi petit que possible (rappelons qu'il est minoré par 1).

### Exercice 3

Résolvons d'abord les deux questions générales suivantes : si  $g$  est une densité de la loi simulable d'une variable aléatoire  $X$  sur  $R$  et  $[a, b] \subset R$ , quelle est la loi de  $X$  conditionnelle à  $X \in [a, b]$  ? Comment la simuler ?

La fonction de répartition de cette loi est

$$P[X \leq x | X \in [a, b]] = \frac{P[X \leq x, X \in [a, b]]}{P[X \in [a, b]]} = \begin{cases} 0 & \text{pour } x \leq a \\ \frac{P[X \in [a, x]]}{P[X \in [a, b]]} & \text{pour } x \in [a, b] \\ 1 & \text{pour } x \geq b \end{cases}$$

La densité  $g_{ab}$  de cette loi est donc nulle en dehors de  $[a, b]$ , et elle vaut  $g_{ab}(x) = \frac{g(x)}{\int_a^b g(t) dt}$ .

Par ailleurs, on voit immédiatement que  $g_{ab} \leq Mg(x)$  (avec  $M = \frac{1}{\int_a^b g(t) dt}$ ) :  $g$  étant

simulable,  $g_{ab}$  l'est également par la méthode d'acceptation-rejet.

On répond alors à la question de l'exercice en utilisant la méthode des lois marginales. Puisque l'on sait, en vertu de ce qui précède, faire des tirages conditionnellement à  $X \leq 0$  et  $X \geq 0$ , on procède de la façon suivante :

(i) on fait le tirage dans  $X \leq 0$ ,  $X \geq 0$  selon  $\int_{-\infty}^0 f(x) dx = \frac{1}{2}$ ,  $\int_0^{+\infty} f(x) dx = \frac{1}{2}$  ;

(ii) si le tirage a donné  $X \leq 0$ , on fait le tirage dans  $] -\infty, 0]$  selon la méthode ci-dessus ( $g$  est une gaussienne sur  $R$ ,  $a = -\infty$ ,  $b = 0$ ). Si le tirage a donné  $X \geq 0$ , on fait le tirage dans  $[0, +\infty[$  selon la méthode ci-dessus ( $g$  est une exponentielle sur  $R^+$ ,  $a = 0$ ,  $b = +\infty$  : donc, ici,  $g_{ab} = g$ ).

#### Remarques

1. La méthode est généralisable aux cas des densités « simulables par morceaux » : on se donne une partition  $C_1, \dots, C_n$  de  $R$ , des densités  $f_1, \dots, f_n$  simulables, et on considère une densité  $f$  telle que sa restriction à chaque  $C_i$  soit, à une constante près,  $f_i$ .  $f$  est alors simulable.

2. Attention, le cas ci-dessus n'est pas le cas du « mélange » décrit dans le sous-paragraphe 2.3.