# Simulation de variables aléatoires

## 1 Inversion de la fonction de répartition

Soit $X$ une variable aléatoire réelle de fonction de répartition $F$ définie par $F(x) = \mathbb{P}(X \leq x)$ pour $x \in \mathbb{R}$.

**Exercice 1** (Fonction de répartition)**.** Montrer que la fonction $F$ est croissante, continue à droite, admet en tout point une limite à gauche et le nombre de points de discontinuité de $F$ est fini ou dénombrable. On pourra introduire les ensembles $A_n = \{x \in \mathbb{R},\ F(x) - F(x-) \geq 1/n\}$, pour $n \geq 1$.

La fonction $F$ n'est en général pas une bijection mais, en tant que fonction croissante, elle admet une *fonction inverse généralisée* $F^-$ définie par

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

Cette fonction coïncide avec $F^{-1}$ lorsque $F$ est bijective. La méthode de simulation, dite d'inversion (sous-entendu de la fonction de répartition) s'appuie sur le résultat suivant :

**Proposition 2.** *Si $U$ est une variable aléatoire de loi uniforme sur $[0,1]$, alors $F^-(U)$ admet $F$ pour fonction de répartition.*

**Exemple 3.** on peut ainsi simuler des v.a. de loi exponentielle (de densité $\lambda e^{-\lambda x}\mathbf{1}_{\{x>0\}}$), de loi de Cauchy (de densité $1/(\pi(1+x^2))$), de loi de Rayleigh (de densité $xe^{-x^2/2}\mathbf{1}_{\{x>0\}}$), de loi de densité $2x^{-3}\mathbf{1}_{\{x>1\}}\cdots$

Remarquons que la proposition 2 admet une réciproque, mais sous des hypothèses plus fortes, qui jouera un rôle essentiel dans l'étude de la fonction de répartition empirique et des théorèmes de Glivenko-Cantelli et Kolmogorov-Smirnov.

**Proposition 4.** *Si la fonction de répartition $F$ d'une v.a.r. $X$ est continue, alors $F(X)$ suit la loi uniforme sur $[0,1]$.*

**Exercice 5** (Absence de mémoire)**.** Proposer un algorithme qui illustre l'absence de mémoire de la loi exponentielle : si $X$ suit la loi $\mathcal{E}(\lambda)$ alors $\mathbb{P}(X > s+t | X > t) = \mathbb{P}(X > s)$. Montrer que les lois exponentielles sont les seules lois continues qui vérifient cette propriété. Quelles sont les lois sur $\mathbb{N}$ qui possèdent la propriété d'absence de mémoire ?

Pour la démonstration des propositions 2 et 4, on pourra consulter [CGCDM99, p. 57]. La description de la méthode d'inversion ainsi que de nombreux exemples sont présentés dans [Yca02].

## 2 Simulation de lois discrètes

La méthode d'inversion permet aussi de simuler les lois discrètes. Commençons par des mesures sur un ensemble fini $\{x_1, \dots, x_k\}$. En général (et c'est ce que fait Scilab), on se place sur $\{1, \dots, k\}$.

**Exercice 6.** Soit $\mu = p_1\delta_1 + \cdots + p_k\delta_k$ une loi de probabilité sur $\{1, \dots, k\}$. Montrer que la v.a. suivante a pour loi $\mu$ :

$$X = \mathbf{1}_{\{U<p_1\}} + 2\mathbf{1}_{\{p_1 \leq U < p_1+p_2\}} + \cdots + k\mathbf{1}_{\{p_1+\cdots+p_{k-1} \leq U < 1\}}.$$

Pour simuler une variable aléatoire chargeant un ensemble dénombrable, on peut utiliser l'exercice 6 à l'aide d'une boucle `while`. Cependant, pour certaines lois classiques, comme la loi géométrique ou la loi de Poisson, il existe des astuces plus efficaces.

**Exercice 7** (Loi géométrique)**.** Si $X$ suit une loi exponentielle de paramètre $\lambda$, quelle est la loi de $[X]$ (où $[x]$ désigne la partie entière de $x$) ? En déduire un algorithme de simulation de la loi géométrique de paramètre $p \in ]0,1[$ qui affecte à $k \in \mathbb{N}^*$ le poids $p(1-p)^{k-1}$.

**Exercice 8** (Loi de Poisson)**.** Soit $(T_k)_{k\geq 1}$ une suite de v.a.r. i.i.d. de loi $\mathcal{E}(\lambda)$. On note $S_0 = 0$ et, pour $n \geq 1$, $S_n = T_1 + \cdots + T_n$. De plus, on définit $N = \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k \leq 1\}}$. Montrer que $S_n$ suit la loi Gamma de densité $\lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} \mathbf{1}_{\{x>0\}}$ pour $n \geq 1$. Montrer que $\{S_n \leq 1\} = \{N \geq n\}$. Calculer $\mathbb{P}(N = n)$. En déduire que $N$ suit une loi de Poisson de paramètre $\lambda$.

Cette remarque simple est la partie émergée de l'iceberg *Processus de Poisson* : la loi au temps 1 d'un processus de Poisson d'intensité $\lambda$ suit la loi de Poisson de paramètre $\lambda$. On trouvera la solution de cet exercice et plein d'autres choses sur le processus de Poisson dans [FF02].

## 3 Méthode du rejet

très souvent se pose le problème de simuler des v.a. de loi uniforme sur un sous-ensemble de $\mathbb{R}^d$. Lorsqu'il est borné, il existe une façon très simple de procéder.

**Proposition 9** (Simulation de lois uniformes)**.** *Soit $\lambda_d$ la mesure de Lebesgue dur $\mathbb{R}^d$, $D$ et $D'$ deux boréliens de $R^d$ tels que*

$$D \subset D' \quad et \quad 0 < \lambda_d(D) \leq \lambda_d(D') < +\infty.$$

*Soit $X$ un point aléatoire de loi uniforme sur $D'$. Alors la loi conditionnelle de $X$ sachant que $X \in D$ est la loi uniforme sur $D$.*

**Exercice 10.** Décrire un algorithme qui fournit des réalisations i.i.d. de la loi uniforme sur le disque unité dans $\mathbb{R}^d$ à partir des v.a. de loi uniforme sur le pavé $[-1,1]^d$. Combien d'itérations en moyenne sont nécessaires ? Quelle est la loi du nombre d'itérations ?

**Théorème 11.** *Soit $\mu$ une loi de densité $f$ sur $\mathbb{R}$.*

1. *Supposons que $f$ soit continue et à support compact $[a, b]$ bornée par $M$ (le graphe de $f$ est donc contenu dans le pavé $[a, b] \times [0, M]$). Soit $((X_n, Y_n))_{n\geq 1}$ une suite de vecteurs aléatoires indépendants de même loi uniforme sur $[a, b] \times [0, M]$. On définit $T$ comme le plus petit entier $n \geq 1$ tel que $f(X_n) \geq Y_n$. Alors $T$ est fini presque sûrement, $X_T$ et $T$ sont indépendantes et suivent respectivement la loi $\mu$ et la loi géométrique de paramètre $1/(M(b - a))$.*
2. *Supposons qu'il existe une densité $g$ facile à simuler telle que $f \leq cg$ pour une constante $c > 0$. Soit alors $(W_n)_{n\geq 1}$ et $(U_n)_{n\geq 1}$ deux suites de v.a.r. i.i.d. indépendantes de lois respectives de densité $g$ et uniformes sur $[0, 1]$. On pose $Y_n = cU_n g(W_n)$ et $T$ le plus petit entier $n \geq 1$ tel que $f(W_n) \geq Y_n$. Alors la loi de $W_T$ a pour densité $f$.*

Un algorithme de rejet est d'autant meilleur que le rapport entre l'aire sous la densité $f$ et l'aire du rectangle (ou l'aire sous $cg$) est petit puisque c'est l'inverse du nombre moyen de couples qu'il faudra générer (l'espérance de la loi géométrique de paramètre $p$ vaut $1/p$). Il existe de nombreuses astuces pour construire des domaines efficaces. On pourra consulter [Yca02] pour la preuve de ces résultats et leur mise en pratique sur des exemples.

## 4 Quelques théorèmes limites

**Théorème 12** (Loi forte des grands nombres). *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. On note $\overline{X}_n = (X_1 + \cdot + X_n)/n$ la moyenne des $n$ premiers termes de la suite.*

1. *Si $X_1$ est intégrable alors $(\overline{X}_n)_{n\geq 1}$ converge presque sûrement vers $\mathbb{E}(X_1)$.*
2. *Si $(\overline{X}_n)_{n\geq 1}$ converge presque sûrement vers $a \in \mathbb{R}$ alors $X_1$ est intégrable et $\mathbb{E}(X_1) = a$.*

On peut illustrer le point 2 en simulant la suite des moyennes empiriques pour des v.a. de loi de Cauchy.

**Théorème 13** (Théorème limite central). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. de carré intégrable. On note $\sigma^2 = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$ la variance de $X_1$. Alors*

$$\sqrt{n}(\overline{X}_n - \mathbb{E}(X_1)) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Pour la présentation de ces deux théorèmes fondamentaux, leurs démonstrations et quelques applications, on pourra consulter [BL98].

## 5 Variables aléatoires gaussiennes

On dit que $X$ suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ si sa loi admet pour densité la fonction

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right).$$

Dans ce cas, $m$ et $\sigma^2$ sont respectivement la moyenne et la variance de $X$ et $(X - m)/\sigma$ suit la loi $\mathcal{N}(0, 1)$. Cette stabilité par translation et homothétie permet de ramener le problème à la simulation de v.a. de loi $\mathcal{N}(0, 1)$.

Pour générer des v.a. de loi gaussienne, on ne peut pas utiliser la méthode d'inversion (sauf si quelqu'un est très fort en calculs de primitives ;-) ). Il faut donc utiliser des moyens détournés.

**Exercice 14** (Conditionnement de v.a. exponentielles)**.** Soit $X$ et $Y$ deux v.a. de même loi exponentielle $\mathcal{E}(1)$. On note $E$ l'événement $\left\{Y > (1-X)^2/2\right\}$. Calculer $\mathbb{P}(E)$ et $\mathbb{P}(\{X \leq x\} \cap E)$. En déduire que la loi de $X$ sachant $E$ admet pour densité

$$\frac{2}{\sqrt{2\pi}} e^{-x^2/2} \mathbf{1}_{\{x>0\}}.$$

Utiliser ce résultat pour construire un algorithme pour simuler une v.a. de loi $\mathcal{N}(0,1)$. Combien de v.a. uniformes doit-on simuler en moyenne ?

Cet algorithme n'est pas le plus efficace, loin s'en faut. L'un des plus célèbre s'appelle l'algorithme de Box-Muller.

**Exercice 15** (Algorithme de Box-Muller)**.** Montrer que si $U$ et $V$ sont deux v.a.r. i.i.d. de loi uniforme sur $[0,1]$ alors $X$ et $Y$ définies par

$$\begin{cases} X = \sqrt{-2\ln U} \cos(2\pi V) \\ Y = \sqrt{-2\ln U} \sin(2\pi V) \end{cases}$$

sont indépendantes et de même loi $\mathcal{N}(0,1)$.

Citons encore un autre algorithme souvent présenté dans les ouvrages traitant de simulation. Il mêle rejet et changement de variables.

**Exercice 16** (Algorithme polaire-rejet)**.** Soit $((U_n, V_n))_{n\in\mathbb{N}^*}$ des v.a. i.i.d. de loi uniforme sur $[-1,1]$. On note $T = \inf\left\{n \geq 1,\ U_n^2 + V_n^2 \leq 1\right\}$ et $R^2 = U_T^2 + V_T^2$. Montrer que les variables aléatoires

$$\begin{cases} X = \sqrt{-2(\ln R^2)/R^2}\, U_T \\ Y = \sqrt{-2(\ln R^2)/R^2}\, V_T \end{cases}$$

sont indépendantes et de même loi $\mathcal{N}(0,1)$.

Un algorithme est d'autant plus efficace qu'il ne fait appel qu'à un petit nombre d'opérations compliquées au rang desquelles se trouvent les générations de v.a. et les évaluations de fonctions comme ln, cos sin... Pour une comparaison des temps de calculs des algorithmes, on pourra consulter [Yca02].

## 6 Vecteurs aléatoires gaussiens

Un vecteur (colonne) aléatoire $X = (X_1, \ldots, X_d)^T$ est dit gaussien si toute combinaison linéaire de ses coordonnées est une variable aléatoire (réelle) gaussienne. Il ne suffit pas que $X_1, \ldots, X_d$ le soient (contre-exemple ?). La loi d'un vecteur gaussien est caractérisé par son (vecteur-)espérance $m = (\mathbb{E}(X_1, \ldots, X_d)^T$ et sa matrice de covariance $\Gamma$ dont les coefficients sont définis par

$$\Gamma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = (\mathbb{E}(XX^T) - mm^T)_{ij}.$$

On note $\mathcal{N}(m, \Gamma)$ la loi de $X$.

**Exercice 17** (Décomposition de Cholesky et vecteurs gaussiens)**.** Montrer que la matrice $\Gamma$ est symétrique positive. Quelle contrainte sur les coefficients de $\Gamma$ l'inégalité de Cauchy-Schwarz impose-t-elle ? Montrer que toute matrice symétrique positive $\Gamma$ peut s'écrire sous la forme $\Gamma = AA^T$ avec A triangulaire inférieure et que, de plus, $A$ est inversible ssi $\Gamma$ est définie positive. Soit $Y = (Y_1, \ldots, Y_d)^T$ un vecteur aléatoire gaussien centrée de matrice de covariance $I_d$. Quelle est la loi de $AY + m$ ?

## 7 Mélange de lois

La notion de mélange est très importante en probabilités, statistique et modélisation. Imaginons que l'on veuille modéliser la taille des Français (hommes et femmes) adultes. Notons $m_f$ et $m_h$ les tailles moyennes respectives et $\sigma_f^2$ et $\sigma_h^2$ leurs variances. En vertu du théorème limite central, il semble naturel de modéliser la taille d'une française prise au hasard par une loi $\mathcal{N}(m_f, \sigma_f^2)$. Ainsi, si $p$ désigne la proportion de femmes dans la population totale, la loi de la taille d'un individu pris au hasard dans la population sera donnée par le mélange de paramètre $p$ des lois $\mathcal{N}(m_f, \sigma_f^2)$ et $\mathcal{N}(m_h, \sigma_h^2)$. On la note $p\mathcal{N}(m_f, \sigma_f^2) + (1-p)\mathcal{N}(m_h, \sigma_h^2)$. Sa densité est donnée par

$$\frac{p}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{(x-m_f)^2}{2\sigma_f^2}\right) + \frac{1-p}{\sqrt{2\pi\sigma_h^2}} \exp\left(-\frac{(x-m_h)^2}{2\sigma_h^2}\right).$$

**Exercice 18** (Estimation par la méthode des moments)**.** On suppose que l'on observe des v.a. i.d.d. $X_1, \ldots, X_n$ de loi $p\mathcal{N}(a,1)+(1-p)\mathcal{N}(-a,1)$ avec $a > 0$ et $p \in ]0,1[$ inconnus. Calculer $\mathbb{E}(X_1)$ et $\mathbb{E}(X_1^2)$ en fonction de $a$ et $p$. Comment estimer $\mathbb{E}(X_1)$ et $\mathbb{E}(X_1^2)$ en fonction de $X_1, \cdots, X_n$ ? En déduire une procédure d'estimation des paramètres $a$ et $p$.

On a ici mélangé deux lois de probabilité mais on peut aussi en mélanger une famille non dénombrable.

**Définition 19.** Soit $(\mu_\theta)_{\theta\in\Theta}$ une famille de probabilités sur $(\Omega, \mathcal{A})$ et $\nu$ une probabilité sur $\Theta$. On appelle mesure-mélange de $(\mu_\theta)_\theta$ de poids $\nu$ la mesure définie par

$$\forall A \in \mathcal{A}, \quad \mu(A) = \int_\Theta \mu_\theta(A)\nu(d\theta).$$

L'algorithme permettant de simuler une v.a. de loi $\mu$ est le suivant : on génère une v.a. $X$ de loi $\nu$ puis une v.a. $Y$ de loi $\mu_X$ indépendante de $X$ : $Y$ suit alors la loi $\nu$.

**Exercice 20.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. et $N$ une v.a. indépendante de la suite précédente à valeurs dans $\mathbb{N}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On définit la v.a. $Z$ par

$$\forall \omega \in \Omega, \quad Z(\omega) = \begin{cases} 0 & \text{si } N(\omega) = 0, \\ \displaystyle\sum_{n=1}^{N(\omega)} X_n(\omega) & \text{si } N(\omega) \geq 1. \end{cases}$$

Calculer la fonction caractéristique de $Z$. En déduire son espérance et sa variance lorsque ses quantités existent. Quelle est la loi de $Z$ lorsque $X_1$ suit la loi de Bernoulli $\mathcal{B}(p)$ et $N$ la loi de Poisson $\mathcal{P}(\lambda)$ ? Quel rapport avec le mélange de lois ?

## Références

[BL98] P. Barbe et M. Ledoux – *Probabilités*, De la licence à l'agrégation, Belin, 1998.

[CGCDM99] M. Cotrell, V. Genon-Catalot, C. Duhamel et T. Meyre – *Exercices de probabilités*, Cassini, 1999.

[FF02] D. Foata et A. Fuchs – *Processus stochastiques*, Dunod, 2002.

[Yca02] B. Ycart – *Modèles et algorithmes markoviens*, Mathématiques et Applications, vol. 39, Springer, 2002.