

Simulation de variables aléatoires

S. ROBIN
INA–PG, Biométrie

Décembre 1997

Table des matières

1	Introduction	2
2	Variables aléatoires discrètes	3
2.1	Pile ou face	3
2.2	Loi de Bernoulli	3
2.3	Loi binomiale	3
2.4	Loi de probabilité discrète sur un ensemble fini	3
2.5	Loi de Poisson	4
2.5.1	Généralisation des lois sur un ensemble fini	4
2.5.2	Cumul de durées exponentielles	5
3	Variables aléatoires continues	6
3.1	Méthode de la fonction inverse	6
3.2	Algorithme du rejet	7
3.3	Loi normale	8
3.3.1	Méthode de Box-Müller	8
3.3.2	Application du Théorème Central Limite	10
3.4	Extension de l'algorithme du rejet aux densités à support non compact . .	10

1 Introduction

On présente ici quelques méthodes de simulation de variables aléatoires de lois classiques.

On notera F_X la loi (*i.e.* la fonction de répartition) d'une variable aléatoire X et, si elle existe, f_X sa densité :

$$\begin{aligned}F_X(x) &= \Pr\{X \leq x\}, \\f_X(x) &= \frac{d}{dx}F_X(x).\end{aligned}$$

Hypothèse générale : On suppose qu'on dispose d'un générateur de variable aléatoire de loi uniforme sur $[0; 1]$ indépendantes :

$$U \sim \mathcal{U}_{[0;1]}$$

c'est à dire

$$\begin{aligned}F_U(u) &= 0 \quad \text{si } u < 0, \\&= u \quad \text{si } 0 \leq u \leq 1, \\&= 1 \quad \text{si } u > 1, \\f_U(u) &= \mathbf{1}\{u \in [0; 1]\}.\end{aligned}$$

L'indépendance des valeurs est une des conditions essentielles de la validité de la plupart des algorithmes présentés ici.

Exemples :

- Dans SAS: fonction "RANUNI",
- En Pascal: "RANDOMIZE" (initialisation du générateur), puis fonction "RANDOM".

On ne s'étend pas ici sur la conception des ces générateurs. Ils sont souvent fondés sur des calculs de congruence sur des nombres de grande dimension et initialisés à partir de l'horloge de la machine. Il faut cependant savoir qu'il ne s'agit que de générateurs de nombres pseudo-aléatoires, et que notamment, les valeurs obtenues ne sont qu'apparemment indépendantes.

2 Variables aléatoires discrètes

2.1 Pile ou face

On veut simuler une variable aléatoire de loi "Pile ou Face" *i.e.* $X \sim \mathcal{B}(\frac{1}{2})$.

On tire $U \rightarrow$ "Pile" si $U \leq \frac{1}{2}$,
"Face" sinon.

Formellement, cette procédure s'écrit

$$X = \mathbf{1}\{U \leq \frac{1}{2}\}.$$

2.2 Loi de Bernoulli

On veut simuler $X \sim \mathcal{B}(p)$.

On tire $U \rightarrow x = 1$ si $u \leq p$,
 $x = 0$ sinon

c'est à dire

$$X = \mathbf{1}\{U \leq p\}.$$

2.3 Loi binomiale

On veut simuler $X \sim \mathcal{B}(n, p)$.

On sait qu'une variable binomiale représente la somme de n variables indépendantes de Bernoulli de paramètre p :

$$\{Y_1, \dots, Y_n\} \text{ i.i.d., } Y_i \sim \mathcal{B}(p) \Rightarrow X = \sum_{i=1}^n Y_i \sim \mathcal{B}(n, p)$$

Il suffit donc de simuler n variables aléatoires indépendantes de loi $\mathcal{B}(p)$ et d'en faire la somme :

$$X = \sum_{i=1}^n \mathbf{1}\{U_i \leq p\}.$$

2.4 Loi de probabilité discrète sur un ensemble fini

Soit X une variable aléatoire à valeurs dans $\{1, \dots, K\}$, on note

$$p_k = \Pr\{X = k\} \quad (\text{avec } \sum_{k=1}^K p_k = 1)$$

et P_k le cumul des p_k :

$$P_k = \sum_{j \leq k} p_j, \quad P_0 = 0.$$

On a donc $P_1 = p_1$ et $P_K = 1$.

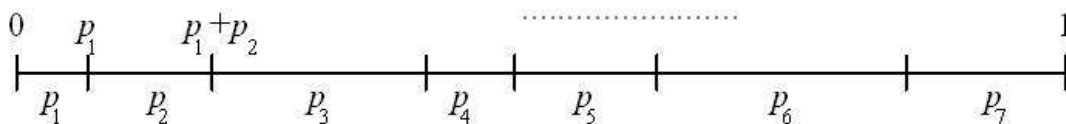
L'algorithme est donc

On tire $U \rightarrow$ si $P_{k-1} \leq u \leq P_k$

ce qui s'écrit également

$$X = \sum_{k=1}^{k=K} k \mathbf{1}\{P_{k-1} \leq U < P_k\}$$

Remarque : Cette méthode revient à découper l'intervalle $[0; 1]$ en K morceaux de longueurs respectives p_k :



2.5 Loi de Poisson

2.5.1 Généralisation des lois sur un ensemble fini

On veut simuler $X \sim \mathcal{P}(\lambda)$.

Une variable aléatoire poissonnienne ne prend pas ses valeurs dans un ensemble fini, mais on peut étendre la méthode précédente au cas où X prend ses valeurs dans \mathbb{N} . En fait, la méthode proposée ici annonce la méthode de la fonction inverse.

On sait que

$$X \sim \mathcal{P}(\lambda) \Leftrightarrow p_k = \Pr\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!} \quad (\text{pour } k \in \mathbb{N})$$

ce qui implique que

$$p_{k+1} = \frac{\lambda}{k+1} p_k$$

et, en notant P_k le cumul des p_k ($P_k = \sum_{j=0}^{j=k} p_j$), que

$$P_{k+1} = P_k + \frac{\lambda}{p+1} p_k.$$

On simule donc un variable de Poisson de paramètre λ en prenant

$$X = \sum_{k \geq 0} k \mathbf{1}\{P_{k-1} \leq U < P_k\}$$

avec la convention $P_{-1} = 0$.

Cet algorithme est assez simple à programmer grâce aux remarques sur les p_k et les P_k faites ci-dessus.

2.5.2 Cumul de durées exponentielles

Une autre méthode de simulation de variables aléatoires de Poisson est issue d'une des propriétés des processus de Poisson.

On sait que si des événements surviennent à des dates séparées par des durées exponentielles de paramètre λ , le nombre d'événements survenant en une unité de temps suit une loi de Poisson de même paramètre.

On simule des variables aléatoires $\{Y_1, Y_2, \dots\}$ i.i.d., $Y_i \sim \mathcal{E}(\lambda)$ (voir le paragraphe sur les variables aléatoires continues) et définit X par

$$X = \sum_{k \geq 0} k \mathbf{1}\{Z_k \leq 1 < Z_{k+1}\}$$

en notant Z_k le cumul des durées Y_i : $Z_k = \sum_{i=1}^k Y_i$.

Il faut donc simuler des variables aléatoires exponentielles de paramètre λ et compter le nombre de simulations nécessaires pour dépasser 1, ou bien simuler des variables aléatoires exponentielles de paramètre 1 et compter le nombre de simulations nécessaires pour dépasser λ .

3 Variables aléatoires continues

3.1 Méthode de la fonction inverse

On veut simuler une variable aléatoire continues X de fonction de répartition F .

Théorème : Soit X une variable aléatoire de fonction de répartition F strictement croissante, on a

$$F(X) \sim \mathcal{U}_{[0;1]}.$$

Démonstration : On pose

$$u = F(x) \quad \Leftrightarrow \quad x = F^{-1}(u),$$

par définition, on a $F(x) = \Pr\{X \leq x\}$ et donc

$$F(F^{-1}(u)) = \Pr\{X \leq F^{-1}(u)\}$$

or $F(F^{-1}(u)) = u$ par définition de la réciproque et $\Pr\{X \leq F^{-1}(u)\} = \Pr\{F(X) \leq u\}$ car F est strictement croissante. On a donc

$$u = \Pr\{F(X) \leq u\}$$

et on reconnaît la fonction de répartition de la loi uniforme.

Méthode : Si on connaît la fonction F^{-1} , réciproque de F , il suffit de tirer

$$X = F^{-1}(U).$$

Exemple : loi exponentielle

$$X \sim \mathcal{E}(\lambda) \quad \Leftrightarrow \quad F(x) = 1 - e^{-\lambda x} \quad \Leftrightarrow \quad F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u).$$

On pourrait donc poser $X = -\ln(1 - U)/\lambda$, mais on peut remarquer que si U suit une loi $\mathcal{U}_{[0;1]}$, $1 - U$ également. On pose donc :

$$X = -\frac{\ln U}{\lambda}.$$

Remarque : L'hypothèse de la connaissance de F^{-1} n'a de sens que si F est strictement croissante. Cependant, même dans ce cas, il se peut que F^{-1} n'ait pas d'expression analytique simple, c'est le cas par exemple pour la loi normale.

Si on note Φ la fonction de répartition de la loi normale centrée réduite,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

il n'existe pas de formulation simple de $\Phi(x)$ et encore moins de $\Phi^{-1}(x)$; la méthode de la fonction inverse ne peut donc pas s'appliquer directement à la loi normale.

Cependant il existe des polynômes donnant d'assez bonnes approximations de (x) et de $\Phi^{-1}(x)$ qui permettent donc d'appliquer la méthode la fonction inverse à la loi normale moyennant une excellente approximation.

3.2 Algorithme du rejet

On veut simuler une variable aléatoire X de densité f et de fonction de répartition F .

Hypothèses :

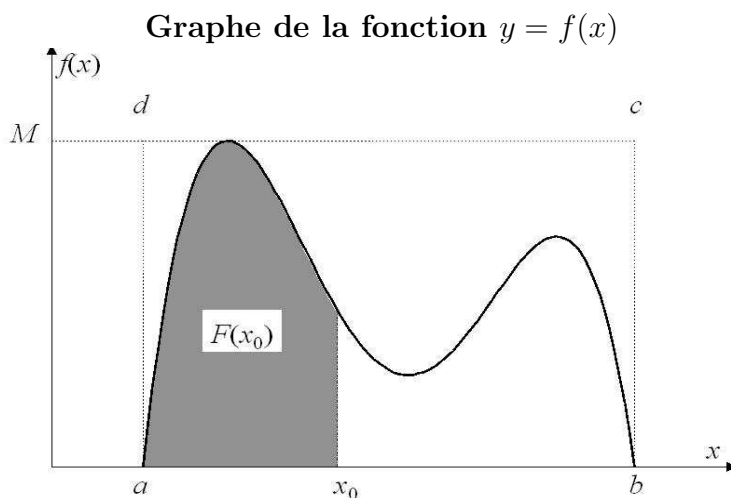
1. f est à support compact, *i.e* que. f est nulle en dehors d'un intervalle $[a; b]$.

Exemple :

$$f(x) = 6x(1 - x)\mathbf{1}\{0 \leq x \leq 1\}.$$

2. Il existe un majorant M de $f(x)$:

$$\forall x \in [a; b] : f(x) \leq M.$$



On a

- surface du rectangle $(abcd) = M(b - a)$,
- surface sous la courbe = 1 (puisque c'est une densité),
- surface de la zone hachurée = $F(x_0)$.

Méthode :

1. On tire un point $A = (X, Y)$ uniformément distribué sur le rectangle $(abcd)$:

$$\text{on tire } U_1 \quad \rightarrow \quad X = a + (b - a)U_1,$$

$$\text{on tire } U_2 \quad \rightarrow \quad Y = MU_2.$$

2. Si $Y \leq f(X)$, on garde X sinon on tire un autre point.
3. On réitère jusqu'à ce que la condition $Y \leq f(X)$ soit remplie.

Théorème : La variable X ainsi obtenue a bien une densité f et une fonction de répartition F .

Démonstration :

$$\begin{aligned} \Pr\{X \leq x_0\} &= \Pr\{X \leq x_0 \mid \text{on garde } A = (X, Y)\} \\ &= \Pr\{X \leq x_0 \mid Y \leq f(X)\} = \frac{\Pr\{X \leq x_0, Y \leq f(X)\}}{\Pr\{Y \leq f(X)\}} \end{aligned}$$

or

$$\begin{aligned} \Pr\{X \leq x_0, Y \leq f(X)\} &= \frac{\text{surface de la zone hachurée}}{\text{surface du rectangle } (abcd)} = \frac{F(x_0)}{M(b-a)} \\ \Pr\{Y \leq f(X)\} &= \frac{\text{surface sous la courbe}}{\text{surface du rectangle } (abcd)} = \frac{1}{M(b-a)} \end{aligned}$$

et donc

$$\Pr\{X \leq x_0\} = \frac{F(x_0)}{M(b-a)} \times \frac{M(b-a)}{1} = F(x_0).$$

Remarque : Il est de plus important de noter qu'on choisit la constante M la plus petite possible pour minimiser le nombre de rejets : plus la majoration est grossière, plus il faut de tirages pour obtenir une valeur acceptable.

3.3 Loi normale

3.3.1 Méthode de Box-Müller

La loi normale n'a pas une densité à support compact et on ne connaît pas d'expression simple de l'inverse de sa fonction de répartition. On ne peut donc, théoriquement, employer aucune des deux méthodes précédentes. On présente ici une méthode qui permet de simuler un couple des variables aléatoires normales, centrées, réduites et indépendantes.

On veut simuler $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$ indépendantes.

On connaît la densité jointe de X et Y :

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

On effectue le passage en coordonnées polaires

$$x = \rho \cos \theta, \quad y = \rho \sin \theta$$

et on obtient

$$\begin{aligned} f_{X,Y}(x,y)dx dy &= \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) dx dy \\ &= \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2}\right) \rho d\rho d\theta \\ &= f_{R,\Theta}(\rho,\theta) d\rho d\theta. \end{aligned}$$

Dans la densité jointe des variable R et Θ , on reconnaît

$$\begin{aligned} \frac{1}{2\pi} &= \text{densité de } \Theta \text{ qui suit une loi } \mathcal{U}_{[0;2\pi]}, \\ \rho \exp\left(-\frac{\rho^2}{2}\right) &= \text{densité de } R. \end{aligned}$$

On en déduit la fonction de répartition de R :

$$F_R(\rho) = \Pr\{R \leq \rho\} = \int_0^\rho \rho \exp\left(-\frac{t^2}{2}\right) dt = 1 - \exp\left(-\frac{t^2}{2}\right)$$

où on reconnaît une loi exponentielle $\mathcal{E}(\frac{1}{2})$ pour R^2 .

On a donc les lois de R et Θ :

$$R^2 \sim \mathcal{E}\left(\frac{1}{2}\right), \quad \Theta \sim \mathcal{U}_{[0;2\pi]}.$$

Méthode :

$$\begin{aligned} \text{On tire } U_1 &\rightarrow R = \sqrt{-2 \ln U_1}, \\ \text{on tire } U_2 &\rightarrow \Theta = 2\pi U_2. \end{aligned}$$

et on pose :

$$\begin{aligned} X &= R \cos \Theta, \\ Y &= R \sin \Theta. \end{aligned}$$

Ces deux variables aléatoires sont indépendantes par construction (leur densité jointe est définie comme le produit de leurs densités respectives).

Remarque : Pour simuler $Z \sim \mathcal{N}(\mu, \sigma^2)$, on simule $X \sim \mathcal{N}(0, 1)$ et on effectue la transformation:

$$Z = \mu + \sigma X.$$

3.3.2 Application du Théorème Central Limite

Une application immédiate du théorème central limite donne une méthode très simple de génération de variables aléatoires normales.

On sait que si on a n variable aléatoire i.i.d. d'espérance μ et de variance σ^2 , en notant S leur somme, on a :

$$\frac{S - n\mu}{\sqrt{n}\sigma} \xrightarrow{loi} \mathcal{N}(0, 1).$$

Dans de nombreux cas, cette convergence est assez rapide et permet d'assez bonnes approximations pour $n \simeq 10$.

On utilise ici les propriétés des variable aléatoire de loi $\mathcal{U}_{[0;1]}$:

$$\mathbb{E}(U) = \frac{1}{2}, \quad \mathbb{V}(U) = \frac{1}{12}$$

et on simule 12 variables aléatoires indépendantes uniformes sur $[0; 1]$ et on en fait la somme. La variance de cette somme vaut 1 et il faut ensuite la centrer en lui retranchant son espérance, c'est à dire 6 :

$$X = \sum_{i=1}^{12} U_i - 6$$

La loi de cette variable est très proche de la loi normale $\mathcal{N}(0, 1)$.

3.4 Extension de l'algorithme du rejet aux densités à support non compact

On veut simuler une variable aléatoire X de densité f telle qu'il existe une autre densité g pour laquelle on connaît un algorithme de génération et qui, à une constante a près, majore f partout :

$$X \sim F, f : \quad \exists g \text{ une densité}, \exists a \in \mathbb{R}^+, \forall x, \quad f(x) \leq a \times g(x).$$

Il est clair qu'on ne peut pas trouver une densité g qui majore f directement (*i.e.* prendre $a = 1$) puisque ces deux fonctions ont une intégrale sur \mathbb{R} égale à 1.

Algorithme du rejet généralisé : On appelle respectivement C_f , C_g et C_{ag} les graphes de $f(x)$, $g(x)$ et $a \times g(x)$.

1. On tire un point $A = (X, Y)$ sous C_{ag} distribué selon G en abscisse et uniformément sur $[0; ag(x)]$ en ordonnée :

$$\begin{aligned} \text{on tire } X &\sim G && (\text{par exemple } X = G^{-1}(U_1)), \\ \text{on tire } Y &\sim \mathcal{U}_{[0, ag(X)]} && (\text{par exemple } Y = a \times g(X) \times U_2). \end{aligned}$$

2. Si $Y \leq f(X)$, on garde X , sinon on tire un autre point.
3. On réitère jusqu'à ce que la condition $Y \leq f(X)$ soit remplie.

Démonstration : On a

$$\Pr\{X \leq x_0\} = \frac{\Pr\{X \leq x_0, Y \leq f(X)\}}{\Pr\{Y \leq f(X)\}}$$

or

$$\begin{aligned} \Pr\{X \leq x_0, Y \leq f(X)\} &= \int_{-\infty}^{x_0} g(x) \Pr\{ag(x)U \leq f(x)\} dx \\ &= \int_{-\infty}^{x_0} g(x) \Pr\left\{U \leq \frac{f(x)}{ag(x)}\right\} dx = \int_{-\infty}^{x_0} g(x) \frac{f(x)}{ag(x)} dx = \frac{F(x_0)}{a}, \\ \Pr\{Y \leq f(X)\} &= \int_{\mathbb{R}} g(x) \Pr\{ag(x)U \leq f(x)\} dx = \frac{1}{a} \end{aligned}$$

et donc

$$\Pr\{X \leq x_0\} = F(x_0).$$

Application à la loi normale : En notant φ la densité de la loi normale, on a

$$\forall x \quad \varphi(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}} \leq \frac{\exp\left(\frac{1}{2} - |x|\right)}{\sqrt{2\pi}}.$$

En effet

$$\exp\left(-\frac{x^2}{2}\right) \leq \exp\left(\frac{1}{2} - |x|\right) \Leftrightarrow \frac{x^2}{2} - |x| + \frac{1}{2} \geq 0 \Leftrightarrow \frac{(|x| - 1)^2}{2} \geq 0$$

ce qui est toujours vrai.

Or

$$\frac{\exp\left(\frac{1}{2} - |x|\right)}{\sqrt{2\pi}} = \sqrt{\frac{2e}{\pi}} \times \frac{1}{2} e^{-|x|} = \sqrt{\frac{2e}{\pi}} \times g(x)$$

où $g(x)$ est la densité de la loi "double exponentielle" qui correspond à une variable exponentielle $\mathcal{E}(1)$ affectée d'un signe est tiré à Pile ou Face.

Il est de plus important de noter qu'on choisit la constante a la plus petite possible pour minimiser le nombre de rejet : plus la majoration est grossière, plus il faut de tirages

pour obtenir une valeur acceptable.

