NOTE 24 - LA MÉTHODE DE VON NEUMANN (1951)

www.daniel-saada.eu

En 1951, John von Neumann publia une méthode de simulation d'une variable aléatoire à densité *bornée*. En voici le fondement mathématique :

TH. – Soit (U_n) une suite de variables uniformes, indépendantes et à valeurs dans $[0,1]$ et f une densité bornée sur $[0,1]$: f est mesurable, positive et $\int_0^1 f = 1$. Soit M un majorant de f, avec $M > 1$, et N la variable aléatoire entière définie par $N = \{2n : M.U_{2n-1} \leq f(U_{2n})\}$: alors $X = U_N$ est de densité f.

DÉMONSTRATION (GÉRARD LETAC)

On pose $g = \dfrac{f}{M}$, g est alors une application de $[0,1]$ dans $[0,1]$; de ce fait, $p\left(\mathrm{U}_{2n-1} \leq g(U_{2n})\right)$ vaut évidemment $\int_0^1 g = 1/M$ pour tout n.

Plus généralement, $p\left(U_{2n} \in B \text{ et } \mathrm{U}_{2n-1} \leq g(U_{2n})\right) = \int_B g$ pour tout borélien B de $[0,1]$.

Par indépendance, $p(N = 2n) = \left(1 - \dfrac{1}{M}\right)^{n-1} \dfrac{1}{M}$; en particulier, $p(N \geq 2n) = \left(1 - \dfrac{1}{M}\right)^{n-1}$.

On écrit ensuite

$$\begin{aligned} p\left(U_{2n} \in B \text{ et } N = 2n\right) &= p\left(U_{2n} \in B \text{ et } \mathrm{U}_{2n-1} \leq g(U_{2n}) \text{ et } N \geq 2n\right) \\ &= \left(1 - \frac{1}{M}\right)^{n-1} \int_B g \end{aligned}$$

la dernière égalité étant justifiée par le fait que l'événement $N \geq 2n$ ne dépend que de $U_1, \dots, U_{2n-2}$ et est donc indépendant des variables U_{2n-1} et U_{2n}.

Enfin, $p(U_N \in B) = \sum_{n=1}^{\infty} p\left(U_{2n} \in B \text{ et } N = 2n\right) = \int_B g \sum_1^{\infty} \left(1 - \dfrac{1}{M}\right)^{n-1} = \int_B f$, ce qui prouve que

$X = U_N$ est de densité f.

REMARQUES.

– Il est à peu près sûr que $\sup\limits_{[0,1]} f > 1$, sinon f serait égale à 1 presque partout.

– La simulation est d'autant plus rapide que M est proche de 1 car $E(N) = 2M$, on a donc intérêt à choisir $M = \sup\limits_{[0,1]} f$ (si c'est calculable).

EXEMPLE. Soit X de densité $f(x)=\frac{\pi}{2}\sin(\pi x)$ sur $[0,1]$, pour laquelle $M=\pi/2>1$.

On simule X par l'algorithme de von Neumann et on compte les réalisations de X qui sont dans l'intervalle $[i/10,(i+1)/10[$ pour i allant de 0 à 9, puis on compare avec

$$F(i)=P\left(i/10\le X<(i+1)/10\right)=\frac{\cos(\pi i/10)-\cos(\pi(i+1)/10)}{2}.$$

Algorithme :

1. mettre $F(i)=0$ pour i allant de 0 à 9

2. faire 5000 fois

3. $U\leftarrow random$ et $V\leftarrow random$

4. Si $U>\sin(\pi V)$ aller en 3.

5. $X\leftarrow V$

6. $i\leftarrow \text{int}(10X)$ et $F(i)\leftarrow 1+F(i)$.

Résultats obtenus par 5000 simulations :

i	0	1	2	3	4	5	6	6	8	9
fréquences observées	2,72	6,86	10,88	13,42	16,16	15,40	13,58	11,56	6,96	2,46
fréquences théoriques	2,45	7,10	11,06	13,94	15,45	15,45	13,94	11,06	7,10	2,45

Examinons pour finir les cas où X, toujours de densité f bornée, ne prend pas ses valeurs seulement dans $[0,1]$:

• Cas où X est bornée : $a\le X\le b$.

On pose $Y=\frac{X-a}{b-a}$: la variable aléatoire Y est à valeurs dans $[0,1]$ et est à densité bornée, on peut donc la simuler, puis on utilise $X=a+Y(b-a)$.

• Cas où X n'est pas bornée

La variable aléatoire $Y=\frac{1}{\pi}\arctan X+\frac{1}{2}$ est à valeurs dans $[0,1]$ et est à densité bornée ; on termine par $X=-1/\tan(\pi Z)$.

COMPLÉMENT. Pour simuler une variable aléatoire à densité non bornée, consulter :

http://fr.wikipedia.org/wiki/M%C3%A9thode_de_rejet

****