Simulation des variables aléatoires

Prof. Mohamed El Merouani
http://elmerouani.jimdo.com
e-mail: m_merouani@yahoo.fr

1

Simulation par la méthode de rejet

2

Introduction

- Pour la méthode d'inversion, il convient de connaître la fonction de répartition.
- Mais, parfois, on connait la fonction de densité et on ne connait pas la fonction de répartition, comme il arrive, par exemple, pour la loi normale.
- Dans ces cas, on peut appliquer la méthode de rejet, introduite par Von Neumann (1951).

3

Introduction

- Supposons que nous voulons échantillonner à partir d'une v.a. X de fonction de densité f.
- On ne sait pas le faire directement, mais nous disposons d'un procédé pour échantillonner à partir d'une autre fonction de densité g telle que $f(x) \leq cg(x)$ pour tout x (avec c une constante finie).

4

Simulation par la méthode de rejet

- La méthode ou l'algorithme de rejet est:

> Jusqu'à ce que $U \leq f(x)/cg(x)$
> Générer $X \to g$
> Générer $U \to \mathcal{U}(0,1)$
> Sortir X

- La méthode de rejet est équivalente à générer des variables $Y \to \mathcal{U}(0,cg(X))$ et les accepter si $Y \leq f(X)$.

5

Simulation par la méthode de rejet (Preuve):

- Pour démontrer que le procédé est correct, il faut vérifier que $P(X \leq x \,/\, X$ acceptée$)=F(x)$, où F est la fonction de répartition de X. On a

$$P(X \leq x \,/\, X \text{ acceptée})=\frac{P(X \leq x \,,\, X \text{ acceptée})}{P(X \text{ acceptée})}$$

En plus,

$$P(X \text{ acceptée})=\int_{-\infty}^{+\infty} P(Y \leq f(X) \,/X = z)g(z)dz$$

$$=\int_{-\infty}^{+\infty} \frac{f(z)}{cg(z)} g(z)dz = \frac{1}{c} \qquad \textbf{(3,1)}$$

6

Simulation par la méthode de rejet (Preuve):

$P(X \leq x \,, X$ acceptée$)=\int_{-\infty}^{+\infty} P(X \leq x, Y \leq f(X) \,/ X = z) g(z) dz$

$$=\int_{-\infty}^{x} \frac{f(z)}{cg(z)} g(z) dz \;=\frac{\mathrm{F(x)}}{\mathrm{c}}$$

Et on obtient le résultat.

(3.1) indique que en chaque itération de l'algorithme, on accepte une valeur d'une forme indépendante avec probabilité 1/c, ce qui donne l'efficacité du procédé. Le nombre des itérations ou tests indépendants avant l'acceptation d'une valeur suit la loi géométrique de paramètre 1/c, et c étant le nombre moyen de tests. Plus il est proche de 1 (toujours c≥1) plus la méthode est efficace, car les densités f et g seront plus proches.

Exemple:

- Génération de la loi beta B(3,4). Sa fonction de densité est $f(x)=60x^2(1-x)^3$, $0<x<1$
- On prend comme fonction g la densité de la loi $\mathcal{U}(0,1)$. Déterminons une constante c telle que $f \leq cg$. Le maximum de la fonction $f(x)/g(x)$ est atteint en $x=2/5$.
- Ainsi: $\frac{f(x)}{g(x)} \leq 60(\frac{2}{5})^2(1-\frac{2}{5})^3=\frac{1296}{625} \equiv c$
- Par conséquent, $\frac{f(x)}{cg(x)} = \frac{3125}{108} x^2(1-x)^3$

Exemple:

- L'algorithme de rejet pour cet exemple sera, donc:

> Jusqu'à ce que $U_2 \leq \frac{3125}{108} U_1{}^2(1 - U_1)^3$
>
> Générer U_1, $U_2 \rightarrow \mathcal{U}(0,1)$
>
> Sortir U_1

- Le nombre moyen des itérations jusqu'à l'acceptation est c=2,07, l'efficience est 1/c=0,48

9

Exercice:

- On souhaite échantillonner à partir d'une v.a. X de fonction de densité

$$f(x)=3x(2-x)/4, \quad 0 \leq x \leq 2$$

Appliquer la méthode de rejet pour le faire et donner l'algorithme correspondant.

10

Solution:

- On prend comme fonction g la densité de la loi U(0,2). Déterminons une constante c telle que f≤cg. Le maximum de la fonction $\frac{f(x)}{g(x)}$ est attaint en x=1.
- Ainsi $\frac{f(x)}{g(x)} \leq \frac{3}{2} \equiv c$
- Par conséquent, $\frac{f(x)}{cg(x)} = x(2-x)$

11

Solution:

- L'algorithme correspondant est donc:

> Jusqu'à ce que $U_2 \leq U_1(2- U_1)$
>
> Générer $U_1 \rightarrow \mathcal{U}(0,2)$
>
> Générer $U_2 \rightarrow \mathcal{U}(0,1)$
>
> Sortir U_1

- Le nombre moyen des itérations jusqu'à l'acceptation est c=3/2, l'efficience est 1/c=2/3

12

Remarques:

- La probabilité d'acceptation est $1/c$.
- Donc, la probabilité d'acceptation est élevée lorsque c est proche de 1, et cela arrive, bien sûr, lorsque $g(x)$ est pareille à $f(x)$.
- D'où, d'une part, on veut que $g(x)$ soit tellement proche de $f(x)$ que ce soit possible de façon que la probabilité d'acceptation soit élevée.
- D'autre part, on veut que $g(x)$ soit tellement facile à simuler que ce soit possible.

13

Efficience de l'algorithme de rejet:

- Lorsque c est grand, l'efficience de l'algorithme de rejet est basse et le pourcentage des valeurs rejetées est élevé.
- Donc, il faut générer un nombre très élevé de valeurs aléatoires de $g(y)$ pour obtenir un petit échantillon valable de $f(x)$.

14