

REVUE DE STATISTIQUE APPLIQUÉE

D. BOSQ

D. SMILI

Comparaison des méthodes de simulation d'un bruit blanc gaussien

Revue de statistique appliquée, tome 39, n° 3 (1991), p. 5-15

http://www.numdam.org/item?id=RSA_1991__39_3_5_0

© Société française de statistique, 1991, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

COMPARAISON DES MÉTHODES DE SIMULATION D'UN BRUIT BLANC GAUSSIEN

D. BOSQ, D. SMILI

LSTA, Université Pierre et Marie Curie (Paris VI),
CEREMADE, Université Dauphine (Paris IX)

RÉSUMÉ

Nous présentons une étude à la fois théorique et pratique de la structure d'auto-covariance d'un bruit blanc gaussien simulé par les méthodes d'inversion, de Box-Muller et par approximation normale. Nous montrons que la méthode d'inversion n'engendre pas un processus gaussien et que l'approximation normale est la plus performante.

Mots-clés : *Approximation normale, Auto-covariance, Inversion, Gaussien, Générateur, Régression, Simulation.*

ABSTRACT

In this paper, we present a theoretical and practical study of auto-covariance function of a gaussian white noise simulated by the inversion method, the Box-Muller method and normal approximation. We show that the inversion method does not generate a gaussian processes and that the normal approximation is the most efficient.

Key-words : *Normal approximation, Auto-covariance, Inversion, Gaussian, Generator, Regression, Simulation.*

1. Position du Problème

La simulation en statistique est basée sur la génération de variables aléatoires uniformes sur $(0,1)$. Ceci a fait l'objet de nombreux travaux de recherches depuis Lehmer (voir (7)). La méthode classique consiste à construire une suite d'entiers définie par la relation de récurrence :

$$\begin{cases} x_n = a x_{n-1} \pmod{M} & n \geq 1, \\ x_0, M, a \text{ entiers naturels et } a \geq 2. \end{cases}$$

d'où les variables pseudo-aléatoires "uniformes" $u_n = \frac{x_n}{M}$, $n \geq 0$.

Pour faciliter l'étude de ce générateur, nous lui associons le modèle théorique suivant :

$$(T) \begin{cases} U_n = a U_{n-1} \pmod{1}, n \geq 1 ; a \text{ entier} \geq 2 \\ U_0 \text{ est une variable uniforme sur } [0,1[\end{cases}$$

Nous utilisons ce générateur pour simuler un bruit blanc gaussien par la méthode d'inversion, la méthode de Box-Muller et l'approximation normale. On cherche à engendrer un bruit blanc gaussien $(\varepsilon_n)_{n \geq 0}$ où les ε_n suivent la loi normale centrée réduite. Soit $(\hat{\varepsilon}_n)_{n \geq 0}$ le "bruit blanc" obtenu par simulation. Nous nous proposons d'étudier l'influence de la dépendance des variables du générateur sur l'auto-covariance du processus $(\hat{\varepsilon}_n)_{n \geq 0}$.

2. La méthode d'inversion

Rappelons que cette méthode est basée sur le Lemme suivant :

Lemme 2-1 :

Soit X une variable aléatoire de fonction de répartition F et soit U une variable aléatoire de loi uniforme sur l'intervalle (0,1). Alors $F^{-1}(U)$ est de même loi que X. où,

$$F^{-1}(t) = \text{Inf}\{v / F(v) \geq t\}, t \in (0, 1).$$

la suite $(\hat{\varepsilon}_n)_{n \geq 0}$ est définie par :

$$\hat{\varepsilon}_n = \Phi^{-1}(U_n), n \geq 0$$

où Φ^{-1} est la fonction inverse de la fonction de répartition de la loi normale centrée réduite.

Il est clair que $(\hat{\varepsilon}_n)_{n \geq 0}$ est une suite de variables aléatoires gaussiennes centrées réduites et corrélées, cependant le processus $(\hat{\varepsilon}_n)_{n \geq 0}$ n'est pas gaussien :

Proposition 2-2 :

Pour tout $i, j \geq 0, i \neq j$, le couple $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ n'est pas gaussien.

Preuve :

Nous allons raisonner par l'absurde. Considérons,

$$\hat{\varepsilon}_i = \Phi^{-1}(U_i) \text{ et } \hat{\varepsilon}_j = \Phi^{-1}(U_j), q = j - i > 0.$$

Si le couple $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ était gaussien on aurait :

$$E(\hat{\varepsilon}_j | \hat{\varepsilon}_i) = c \hat{\varepsilon}_i \text{ p.s.}$$

où c désigne le coefficient de corrélation de $\hat{\varepsilon}_i$ et $\hat{\varepsilon}_j$. Or pour presque tous les x ,

$$\begin{aligned} E(\hat{\varepsilon}_j | \hat{\varepsilon}_i = x) &= E(\Phi^{-1}(U_j) | \Phi^{-1}(U_i) = x) \\ &= E(\Phi^{-1}(\{a^q U_i\}) | U_i = \Phi(x)) \\ &= \Phi^{-1}(\{a^q \Phi(x)\}) = c x \end{aligned}$$

donc,

$$a^q \Phi(x) - [a^q \Phi(x)] = \Phi(cx).$$

où $[y]$ désigne la partie entière du réel y .

Mais, si $x \in I_k = [\Phi^{-1}(\frac{k}{a^q}), \Phi^{-1}(\frac{k+1}{a^q})]$, on a $[a^q \phi(x)] = k$,
 $0 \leq k \leq a^q - 1$

et

$$\begin{aligned} a^q \Phi(x) - k &= \Phi(cx) \\ a^q \Phi'(x) &= c \Phi'(cx) \\ \frac{a^q e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} &= c \frac{e^{-\frac{(cx)^2}{2}}}{\sqrt{2\pi}} \\ x^2 &= \frac{2(q \text{ Log } a - \text{Log } c)}{1 - c^2} \end{aligned}$$

ce qui est impossible et par conséquent $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ n'est pas gaussien.

c.q.f.d

Remarque : Si $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ était gaussien on aurait : $c = 2 \sin(\frac{\pi}{6a^q - i})$, (voir (10)).

$$\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = \text{Cov}(\Phi^{-1}(U_i), \Phi^{-1}(U_j))$$

Si $j - i = q > 0$ alors :

$$U_j = a^q U_i \pmod{1} = a^q U_i - [a^q U_i]$$

donc ,

$$\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = \int_0^1 \Phi^{-1}(x) \Phi^{-1}(a^q x - [a^q x]) dx$$

Le comportement de cette quantité n'est pas facile à étudier. En effet les approximations usuelles de Φ^{-1} sont peu maniables (voir (1),(6),(9)) et le produit $\Phi^{-1}(x) \Phi^{-1}(a^q x - [a^q x])$ a un mauvais comportement local.

Résultats numériques et simulations :

Les résultats numériques qui suivent sont obtenus en utilisant les approximations de Φ^{-1} données par (11) et (5) et respectivement définies par :

si $0 < u \leq 0.5$:

$$(1) \Phi^{-1}(u) = t - \frac{a_0 + a_1 t}{1 + b_1 t + b_2 t^2} + \varepsilon(u), \quad t = \sqrt{\text{Log} \frac{1}{u^2}},$$

$$|\varepsilon(u)| < 3 \times 10^{-3}.$$

$$(2) \Phi^{-1}(u) = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \varepsilon(u), \quad t = \sqrt{\text{Log} \frac{1}{u^2}},$$

$$|\varepsilon(u)| < 4.5 \times 10^{-4}.$$

$(a_i)_i, (b_i)_i, (c_i)_i$ et $(d_i)_i$ sont des constantes réelles positives.

Les tableaux suivants donnent pour différentes valeurs du paramètre a du générateur les valeurs numériques de la fonction d'auto-covariance théoriques de la suite $(\hat{\varepsilon}_n)_{n \geq 0}$ (qui n'est rien d'autre que la fonction d'auto-corrélation théorique, puisque les $\hat{\varepsilon}_n$ sont de variance un). On remarquera qu'il y a une oscillation de cette fonction autour de 0.

$a = 11$

j	a^j	(1*) $Cov(\hat{\varepsilon}_n, \hat{\varepsilon}_{n+j})$	(2*) $Cov(\hat{\varepsilon}_n, \hat{\varepsilon}_{n+j})$
1	11	0.3463529	0.3552468
2	121	0.8846050	0.9116066
3	1331	-0.2619714	-0.2706538
4	14641	-0.5656095	-0.5841851
5	161051	0.1220383	0.1264923
6	1771561	-2.14109E-03	-2.178001E-03
7	19487171	4.00023E-08	-1.218184E-09

$a = 131$

j	a^j	(1*) $Cov(\hat{\varepsilon}_n, \hat{\varepsilon}_{n+j})$	(2*) $Cov(\hat{\varepsilon}_n, \hat{\varepsilon}_{n+j})$
1	131	0.5702519	-0.5844401
2	17161	-0.7036088	-0.7264198
3	2248091	-8.020136E-07	2.448499E-08
4	294499921	-8.310182E-07	2.53757E-08

(1*) et (2*) sont obtenus respectivement par (1) et (2).

Il est important de remarquer que la suite $(\hat{\varepsilon}_n)_{n \geq 0}$ ne constitue pas un processus gaussien .

Les courbes obtenues (Figures 1, 2 et 3) confirment bien le résultat de la proposition 2-2, elles font apparaître de grandes oscillations de la régression simulée et son écart à la régression d'un couple gaussien.

Régression de $\hat{\varepsilon}_1$ par rapport à $\hat{\varepsilon}_0$ ($a=11$):

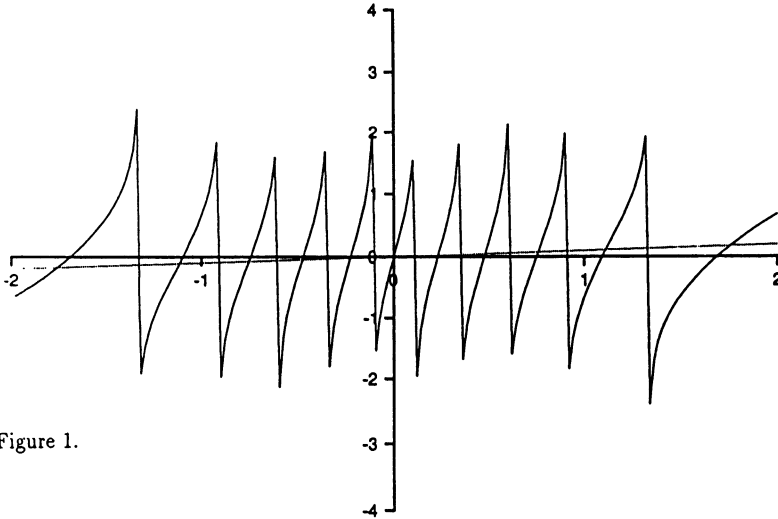


Figure 1.

— $E(\hat{\varepsilon}_1|\hat{\varepsilon}_0)$.
 .. $E(\hat{\varepsilon}_1|\hat{\varepsilon}_0 = x)$ si $(\hat{\varepsilon}_0, \hat{\varepsilon}_1)$ est gaussien (droite en pointillés).

Régression de $\hat{\varepsilon}_1$ par rapport à $\hat{\varepsilon}_0$ ($a=131$):

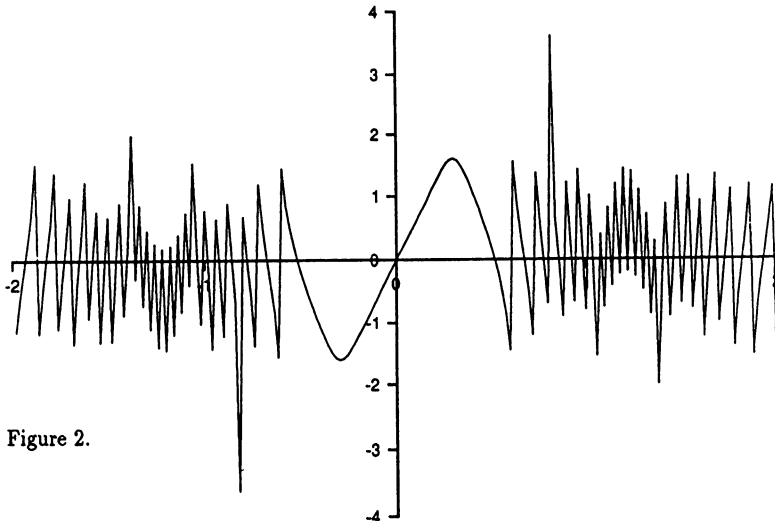


Figure 2.

— $E(\hat{\varepsilon}_1|\hat{\varepsilon}_0)$.
 $E(\hat{\varepsilon}_1|\hat{\varepsilon}_0 = x)$ si $(\hat{\varepsilon}_0, \hat{\varepsilon}_1)$ est gaussien (droite pratiquement confondue avec l'axe des abscisses)

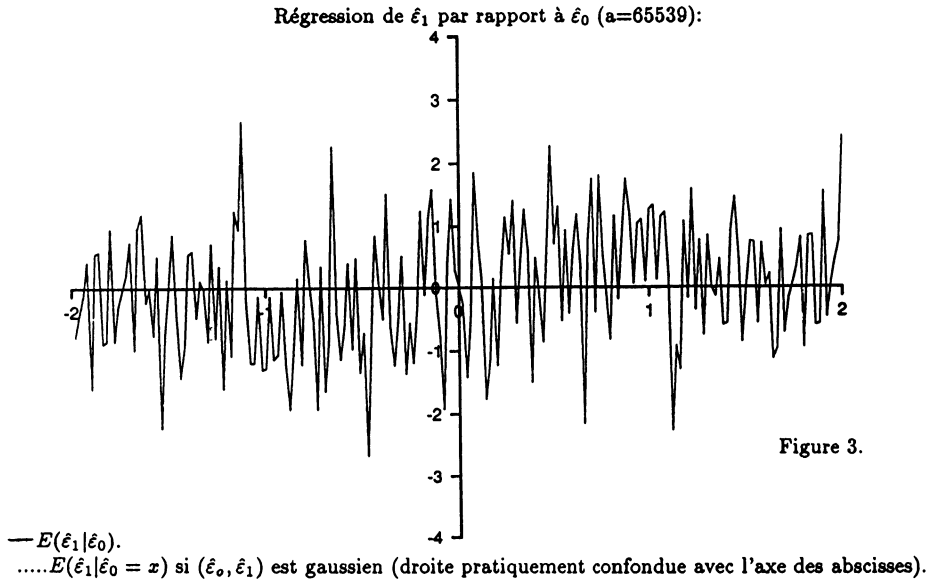


FIGURE 3

3. La transformation de BOX-MULLER

La méthode de *Box-Muller*(1958), permet d'obtenir un couple de variables aléatoires gaussiennes indépendantes en posant :

$$\begin{cases} X = \sqrt{-2 \text{Log } U} \sin 2\pi V \\ Y = \sqrt{-2 \text{Log } U} \cos 2\pi V \end{cases}$$

où U et V sont deux variables aléatoires indépendantes qui suivent la loi uniforme sur $(0,1)$.

Neave (voir (8)), a soulevé le problème de dépendance entre X et Y et a donné une approximation locale de la vraie densité de X . *Golder et Settle* (voir (4)) ont étudié cette méthode pour différents générateurs.

Dans ce qui suit, nous étudierons la fonction d'auto-covariance de la suite $(\hat{\varepsilon}_n)_n$ par cette méthode en utilisant le générateur (T) .

Posons,

$$\begin{cases} \hat{\varepsilon}_{2p} = \sqrt{-2 \text{Log } U_p} \sin 2\pi U_{p+1} \\ \hat{\varepsilon}_{2p+1} = \sqrt{-2 \text{Log } U_p} \cos 2\pi U_{p+1} \end{cases} \quad p \geq 0$$

Il est clair que le processus $(\hat{\varepsilon}_n)_n$ n'est pas gaussien puisque les $\hat{\varepsilon}_{2p}$ et $\hat{\varepsilon}_{2p+1}$ ne le sont pas (U_p et U_{p+1} ne sont pas indépendantes). D'autre part on a (voir 10) :

Proposition 3-1 :

$$1/ \quad E(\hat{\varepsilon}_i) = O\left(\frac{1}{a}\right) = E(\hat{\varepsilon}_j).$$

2/ pour $i \neq j$:

$$\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = O\left(\frac{1}{a^2}\right).$$

3/

$$\text{Var}(\hat{\varepsilon}_i) = 1 + O\left(\frac{1}{a}\right).$$

Résultats numériques : Pour différentes valeurs du paramètre a , chaque case des tableaux suivants donne la valeur théorique de $\text{Cov}(\hat{\varepsilon}_l, \hat{\varepsilon}_{l+n})$ avec $l \in \{i, i+1 \mid i = 2p, p \geq 0\}$ et $n \geq 0$.

	a = 11	a = 11	a = 32	a = 32
	i	i+1	i	i+1
i	0.9867805	3.957891E-02	0.9958125	1.631486E-02
i+1	3.957891E-02	1.011271	1.631486E-02	1.003891
i+2	-9.021959E-04	-5.331706E-03	-1.905297E-04	-6.67363E-04
i+3	6.423485E-02	1.087241E-02	2.247949E-02	2.894646E-03

	a = 131	a = 131	a = 56539	a = 65539
	i	i+1	i	i+1
i	0.9990202	4.847489E-03	0.9999933	2.236567E-05
i+1	4.847489E-03	1.000948	2.236567E-05	0.9999968

Ces résultats confirment bien les résultats théoriques précédents, et compte tenu de la mauvaise approximation gaussienne, il est inutile de donner des graphiques comme ceux des Figures 1, 2, 3.

4. L'approximation normale

Cette méthode est basée sur le théorème de la limite centrale. Sa simplicité vient du fait qu'elle n'utilise pas de fonctions mathématiques présentant des singularités comme dans les cas précédents.

Soit $(U_n)_n$ la suite définie par le générateur (T) et k_0 un entier non nul fixé.

Considérons d'abord les variables :
$$\hat{\varepsilon}'_i = \frac{U_{ik_0} \dots + U_{(i+1)k_0-1} - k_0/2}{\sqrt{k_0/12}}.$$

Nous avons (voir (10)) :

Proposition 4-1 :

$$1/ \text{ Pour } i \neq j : \quad \text{Cov}(\hat{\varepsilon}'_i, \hat{\varepsilon}'_j) = \frac{a^{-(|i-j|+1)k_0+1}}{k_0} \left(\frac{1-a^{k_0}}{1-a} \right)^2.$$

2/

$$\text{Var}(\hat{\varepsilon}'_i) = 1 + \frac{2}{k_0(a-1)} \left(k_0 - 1 - \frac{1}{a-1} (1 - a^{-(k_0-1)}) \right).$$

A partir de ces résultats, nous obtenons le :

Corollaire 4-1 :

$$1/ \lim_{k_0 \rightarrow \infty} \text{Cov}(\hat{\varepsilon}'_i, \hat{\varepsilon}'_j) = \lim_{a \rightarrow \infty} \text{Cov}(\hat{\varepsilon}'_i, \hat{\varepsilon}'_j) \\ = \lim_{|i-j| \rightarrow \infty} \text{Cov}(\hat{\varepsilon}'_i, \hat{\varepsilon}'_j) = 0.$$

$$2/ \lim_{k_0 \rightarrow \infty} \text{Var}(\hat{\varepsilon}'_i) = \frac{a+1}{a-1} \text{ et } \lim_{a \rightarrow \infty} \text{Var}(\hat{\varepsilon}'_i) = 1.$$

Quand a est petit, ces résultats montrent la nécessité d'effectuer une correction pour obtenir des variables de variance asymptotique un. Par conséquent nous posons :

$$\hat{\varepsilon}_n = \sqrt{\frac{a-1}{a+1}} \hat{\varepsilon}'_n, \quad n \geq 0.$$

Pour k_0 assez grand, le théorème central limite multidimensionnel montre alors que le processus $(\hat{\varepsilon}_n)_{n \geq 0}$ est proche d'un bruit blanc gaussien (voir (10)).

Résultats numériques :

Les tableaux suivants présentent les fonctions d'auto-covariance et d'auto-corrélation de la suite $(\hat{\varepsilon}'_n)_{n \geq 0}$; ces résultats montrent bien la supériorité de cette méthode.

$$a = 11, \\ k_0 = 10, \\ \text{Var}(\hat{\varepsilon}'_i) = 1.1780.$$

$$a = 131, \\ k_0 = 10, \\ \text{Var}(\hat{\varepsilon}'_i) = 1.0138.$$

j	$\text{Cov}(\hat{\varepsilon}'_n, \hat{\varepsilon}'_{n+j})$	$\text{Cor}(\hat{\varepsilon}'_n, \hat{\varepsilon}'_{n+j})$
1	0.0110	9.1498E-03
2	4.2409E-13	3.5276E-13
3	1.6350E-23	1.3600E-23
4	6.3039E-34	5.2436E-34
5	2.4304E-44	2.0216E-44

j	$\text{Cov}(\hat{\varepsilon}'_n, \hat{\varepsilon}'_{n+j})$	$\text{Cor}(\hat{\varepsilon}'_n, \hat{\varepsilon}'_{n+j})$
1	7.7515E-04	7.6340E-04
2	5.2080E-25	5.1290E-25
3	3.4991E-46	3.4460E-46

Pour étudier l'auto-régression, nous posons :

$$X = \frac{U_0 + \dots + U_{k_0-1} - k_0/2}{\sqrt{k_0/12}} = \hat{\varepsilon}'_0$$

et

$$Y = \frac{U_{k_0} + \dots + U_{2k_0-1} - k_0/2}{\sqrt{k_0/12}} = \hat{\varepsilon}'_1$$

Alors,

$$E(Y|X = x) = E(\beta(U_0)|\alpha(U_0) = x) , x \in R (*)$$

où,

$$\alpha(t) = \frac{\{t\} + \{at\} + \dots + \{a^{k_0-1}t\} - k_0/2}{\sqrt{k_0/12}}$$

et

$$\beta(t) = \frac{\{a^{k_0}t\} + \{a^{k_0+1}t\} + \dots + \{a^{2k_0-1}t\} - k_0/2}{\sqrt{k_0/12}}$$

$\{t\}$ est la partie fractionnaire du réel t .

Le calcul explicite de (*) étant complexe nous estimons la régression de $\hat{\varepsilon}_{n+1}$ par rapport à $\hat{\varepsilon}_n$ par la méthode du noyau (voir (2), chap.5), l'estimateur est défini par :

$$r_n(x) = \frac{\sum_{i=1}^n \hat{\varepsilon}'_{2i+1} K\left(\frac{x - \hat{\varepsilon}'_{2i}}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - \hat{\varepsilon}'_{2i}}{h_n}\right)}$$

où K est le noyau gaussien :

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} ,$$

$$h_n = \frac{s_n}{n^{1/5}} \text{ et } s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}'_{2i} - \bar{\varepsilon}')^2}$$

Un calcul analogue est effectué pour le processus $(\hat{\varepsilon}_n)_{n \geq 0}$.

L'estimateur r_n est simulé par différents générateurs avec $n = 2000$ et $k_0 = 30$ (voir Figures 4,5 et 6).

Estimateur de la régression simplifié par approximation normale (a =131):

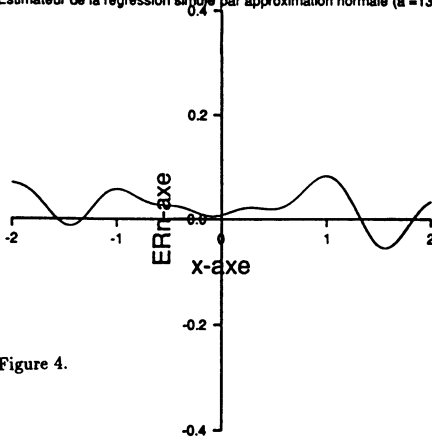


Figure 4.

Estimateur de la régression simplifié par approximation normale (a =65539):

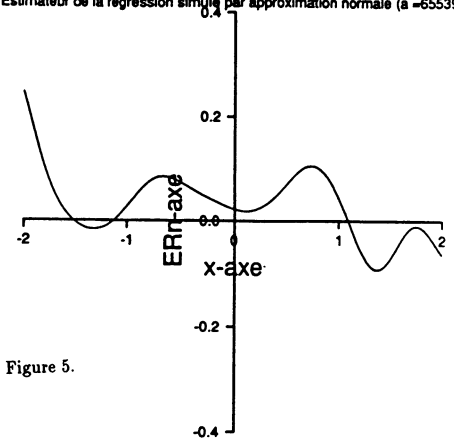


Figure 5.

- Estimateur de $E(\hat{\epsilon}'_1 | \hat{\epsilon}'_0)$
- Estimateur de $E(\hat{\epsilon}_1 | \hat{\epsilon}_0)$ (les deux courbes en continu et en pointillés sont pratiquement confondues).

Estimateur de la régression simplifié par approximation normale (a =11):

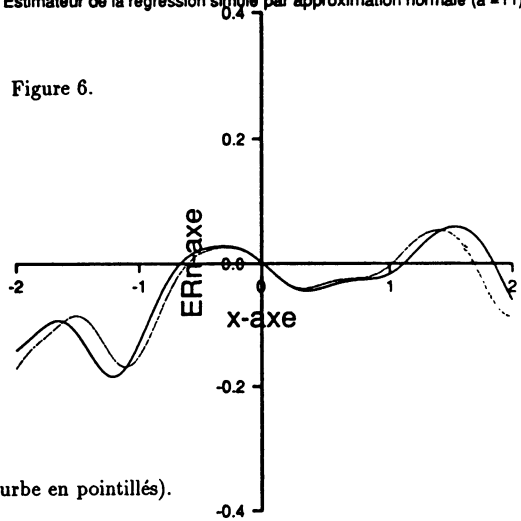


Figure 6.

- Estimateur de $E(\hat{\epsilon}'_1 | \hat{\epsilon}'_0)$.
- Estimateur de $E(\hat{\epsilon}_1 | \hat{\epsilon}_0)$ (courbe en pointillés).

5. Conclusions

1) La méthode d'inversion donne bien une suite de variables aléatoires gaussiennes mais la dépendance des termes du générateur ne permet pas d'engendrer un processus gaussien.

2) La méthode de Box-Muller donne simplement une suite de variables aléatoires corrélées. Elle ne peut engendrer un bruit blanc gaussien que si l'on utilise deux générateurs indépendants de variables aléatoires uniformes.

3) La méthode d'approximation normale est plus performante pour générer un bruit blanc gaussien et plus généralement des variables aléatoires gaussiennes. En particulier la régression semble plus proche de la régression théorique que dans le cas de la méthode d'inversion. Ce résultat est important pour les applications.

Bibliographie

- (1) Bailey, B.J.R. (1981) : *Alternatives to Hastings' approximation to the inverse of the normal cumulative distribution function. Appl. Statist., 30, N° 3, pp. 275-276.*
- (2) Bosq, D. et Lecoutre, J.P (1987) : *théorie de l'estimation fonctionnelle. Ed Economica.*
- (3) Box, J.E.P. and Muller, M.E. (1958) : *A note on the generation of random normal deviates. Ann. Math. Statist., 29, pp. 610-611.*
- (4) Golder, E.R. and Settle, J.G. (1976) : *The Box-Muller method for generating pseudo-random normal deviates. Appl. Statist., 25, N° 1, pp. 12-20.*
- (5) Hastings, C. (1955) : *Approximation for digital computers. Princeton University Press.*
- (6) Hugo, C.H. (1978) : *Approximating the cumulative normal distribution and its inverse. Appl. Statist., 27, N° 1., pp. 76-77.*
- (7) Lehmer, D.H. (1951) : *Mathematical methods in large scale digital computing units. Proceedings of second symposium on large scale digital calculating machinery. Ann. Comput. Lab. Harvard University, V.26, pp. 141-146.*
- (8) Neave, H.R. (1973) : *On using the Box-Muller transformation with multiplicative congruential pseudo-random number generators. Appl. Statist., 22, pp. 92-97.*
- (9) Page, E. (1978) : *Approximation to the cumulative normal function and its inverse for use on a pocket calculator. Appl. Statist., V.26, N° 1, pp. 75-76.*
- (10) Smili, D. (1990) : *Contribution à l'étude probabiliste des générateurs d'échantillons de séries temporelles. Thèse de l'Université Paris 6.*
- (11) Zelen, M. and Severo, N.C. (1966) : *Probability functions. In handbook of mathematical functions (M. Abramowitz and I. A. Stegun, eds). Washington. D.C. : Department of commerce of U.S. Government.*