Fontainebleau / CMM N-653

REMARQUES SUR LA FABRICATION DES ALEAS

G. MATHERON

Mai 1980

REMARQUES SUR LA FABRICATION DES ALEAS

par

G. MATHERON

Table des Matières

REMARQUES SUR LA FABRICATION DES ALEAS

par

G. MATHERON

0 - INTRODUCTION.

Dans ce qui suit, je présente quelques reflexions sur les procédés qui servent à fabriquer ce que l'on appelle d'habitude des "nombres au hasard". Cette terminologie est manifestement impropre, puisque, comme on le sait, la principale caractéristique de ces prétendus "nombres au hasard" est d'être parfaitement déterminés, une fois que l'on connait leur mode de génération. Ils ont seulement l'apparence d'être au hasard, en ce sens qu'ils se révèlent plus ou moins compatibles avec le modèle "variables aléatoires indépendantes et identiquement distribuées" (et, bien sûr, cette compatibilité est jugée selon des critéres bien définis, le plus souvent selon tel ou tel test que fournit la statistique classique). Le coup de dé (en latin : alea) est le prototype et, sans doute, le plus ancien mode de fabrication de ces "nombres au hasard". Ce précédent vénérable me servira de terminologie générale, et dans ce qui suit, j'appellerai Aléa tout procédé de génération de "nombres au hasard".

Les aléas utilisés en pratique sont assez variés, mais ils se ramènent tous, plus ou moins directement, à un procédé unique, toujours le même, qui consiste à prendre la partie décimale d'une fonction donnée. Sous la forme la plus simple, par exemple, on retiendra comme n^{me} nombre x_n les trois ou quatre derniers chiffres significatifs de la valeur de log n fournie par une table de logarithme.

Les procédés les plus usuels utilisent des relations de récurrence du type :

$$x_{n+1} = a\, x_n + b \quad (\text{modulo } m) \tag{0-1}$$

où a, b et m sont des entiers convenablement choisis. Du fait qu'il s'agit de nombres entiers, on vérifie sans peine que x_n est la valeur

modulo m de l'entier y_n fourni par la même relation de récurrence entre entiers naturels, soit $y_{n+1} = a\, y_n + b (y_o \in \mathbb{N})$. Explicitement on trouve :

$$y_n = (y_o + \frac{b}{a-1})\ a^n - \frac{b}{a-1}$$

Ainsi, en désignant par x_o le nombre initial ($0 \leq x_o < m$), la relation (0 - 1) est équivalente à

$$x_n = (x_o + \frac{b}{a-1})\ a^n - \frac{b}{a-1} \qquad \text{(modulo m)}$$

Or ceci se met sous la forme :

$$\frac{x_n}{m} = \text{Déc } f(n)$$

(la notation Déc x désigne la partie décimale du nombre réel x; de même, nous désignerons par Int x, du latin integer, entier, la partie entière du réel x) avec la fonction f définie par :

$$f(n) = \frac{1}{m} [(x_o + \frac{b}{a-1})\ a^n - \frac{b}{a-1}]$$

Il s'agit bien de la partie décimale d'une fonction donnée.

De la même manière, dans le cas d'un coup de dé, le nombre obtenu (1, 2, 3, 4, 5 ou 6) dépend de la partie décimale d'une fonction f, d'ailleurs relativement compliquée, des conditions initiales.

A titre d'illustration, considérons le cas plus simple de la partie de pile ou face. En négligeant la résistance de l'air, en supposant que la pièce tourne autour d'un axe horizontal à raison de ν tours par seconde, et en désignant par a le diamètre de la pièce, la cote au temps t (après l'instant initial $t_o = 0$) du point le plus bas de la pièce est:

$$z(t) = z_o + v\ t + \frac{1}{2}\ g\ t^2 - a\ |\ \sin(2\pi\ \nu\ t + \varphi)\ |$$

La pièce touche donc le sol à l'instant T défini par l'équation $z(T) = 0$. A cet instant T, la pièce a accompli un nombre entier de tours complets

sur elle-même égal à Int(νT), et la valeur (pile ou face) que l'on observera dépend donc uniquement de Déc νT. L'espace à 4 dimensions associé aux conditions initiales z_o, v, ν et φ est ainsi découpé en bandes correspondant alternativement à pile et à face, bandes séparées par les surfaces d'équation :

$$\nu T + \frac{\varphi}{2\pi} = k + \frac{1}{2} \qquad (k \text{ entier})$$

Il s'agit donc bien , ici encore, d'un procédé de décimalisation.

1 - UN MODELE MARKOVIEN.

Dans la relation de récurrence (0-1), prenons b = 0 (ce b ne joue pas de rôle essentiel), et pour nous ramener à des nombres réels compris entre 0 et 1, posons $X_n = x_n/m$. On obtient ainsi la relation de récurrence :

$$X_{n+1} = \text{Déc}(a\, X_n) \qquad (1\text{-}1)$$

Oubliant maintenant l'entier m, nous allons considérer la relation (1-1) pour elle-même : elle définit une application de l'intervalle (0,1) sur lui-même. Si nous considérons X_n comme une VA (sur l'intervalle (0,1) muni de ses boreliens), cette relation (1-1) définit une chaîne de Markov (d'un type un peu particulier, puisque la loi de X_{n+1}, lorsque X_n est donné, est concentrée sur le nombre réel a X_n modulo 1 : mais ce n'est pas une objection).

Cette chaîne de Markov un peu particulière n'en possède pas moins de bonnes propriétés ergodiques. Plus précisément, nous allons établir le théorème suivant :

THEOREME : Soit a un entier > 1, et $x_o \in (0,1)$. Posons $x_n = \text{Déc}(a^n x_o)$ et désignons par

$$\mu_N = \frac{1}{N+1} \sum_{n=0}^{N} \delta_{x_n}$$

la probabilité attribuant le poids 1 / (N+1) à chacun des X_n, n = 0, 1...N. Alors, il existe un ensemble $\mathcal{N}$ négligeable pour la mesure de Lebesgue sur (0,1) tel que la suite μ_N converge

étroitement vers la mesure de Lebesgue sur (0,1) pourvu seulement que $x_0 \notin \mathcal{N}$.

Pour établir ceci, prenons comme espace de probabilité $(\Omega, \mathcal{A}, P)$ l'espace obtenu en élevant à la puissance N l'ensemble $\{0,1,\ldots, a-1\}$ des a premiers entiers munis de la probabilité uniforme (i.e. attribuant le même poids $1/a$ à chacun des a premiers entiers). Ainsi, les éléments $\omega \in \Omega$ sont les suites infinies ω_n, $n=1,2\ldots$, où chaque ω_n prend l'une des valeurs $0,1,\ldots,a-1$. Notons que, pour un n donné, ω_n est, sur cet espace $(\Omega, \mathcal{A}, P)$ une variable aléatoire (avec $P(\omega_n = k) = 1/a$, $k = 0$, $1\ldots a-1$), et surtout que ces VA ω_1, $\omega_2\ldots$ sont mutuellement indépendantes.

Posons alors :

$$X = \sum_{n=1}^{\infty} \frac{\omega_n}{a^n} \tag{1-2}$$

(ce qui revient à interpréter la suite ω_n comme l'écriture en base a d'un nombre réel compris entre 0 et 1). Cette relation (1-2) définit une VA X sur $(\Omega, \mathcal{A}, P)$, dont la loi de probabilité est la loi uniforme sur (0,1). Il est clair que l'on a :

$$\text{Déc}(a\,X) = \sum_{n=1}^{\infty} \frac{\omega_{n+1}}{a^n}$$

Autrement dit, on obtient l'écriture de Déc(a X) en supprimant le premier chiffre de l'écriture de X.

Considérons alors deux entiers p et $N_o \geq 0$, tels que $p < a^{N_o}$, et désignons par $I(p,N_o)$ l'intervalle $[\, p/a^{N_o}, (p+1)/a^{N_o}\, [$ (fermé à gauche, ouvert à droite). Comme on a

$$\text{Déc}\ a^k X = \sum_{n=1}^{\infty} \frac{\omega_{n+k}}{a^n}$$

pour k entier > 0, on voit que $\text{Déc}(a^k X)$ appartient à l'intervalle $I(p,N_o)$ si et seulement si :

$$\sum_{n=1}^{N_o} \omega_{n+k}\, a^{N_o - n} = p$$

Désignons par $A_k = A_k(p,N_o)$ l'évènement correspondant, i.e Déc$_k$ $a^k X \in I(p,N_o)$. Comme les ω_n, $n = 1,2...$ sont mutuellement indépendantes, il en est de même des évènements $A_{k+r\ N_o}$ (k, N_o fixes, $r = 0, 1, 2...$).

Or, la fréquence de visite de $X_n = \text{Déc}(a^n X)$ dans l'intervalle $I(p,N_o)$ est, pour $0 \le n \le N$:

$$\frac{1}{N+1} \sum_{n=0}^{N} 1_{A_n} = \frac{1}{N+1} \sum_{k=0}^{N_o-1} \sum_{r=0}^{r'} 1_{A_{k+r\ N_o}}$$

(r' désigne ici l'entier tel que $k + r' N_o \le N < k + (r'+1) N_o$)

Cette variable apparait donc comme la somme des N_o variables :

$$Y_{k,N} = \frac{1}{N+1} \sum_{r=0}^{r'} 1_{A_{k+r\ N_o}}$$

Or, à k fixé, les évènements $A_{k+r\ N_o}$ ($r = 0, 1,...r'$) sont indépendants, de sorte que la loi forte des grands nombres s'applique, pour $N \to \infty$, à chacune de ces VA $Y_{k,N}$ ($k = 0, 1,...N_o-1$), et on trouve :

$$\lim_{N\to\infty} Y_{k,N} \overset{\text{p.s.}}{=} \frac{1}{N_o a^{N_o}}$$

On en déduit aussitôt que la suite $\frac{1}{N+1} \sum_{n=0}^{N} 1_{A_n}$ converge elle-même presque sûrement vers $1/a^{N_o}$.

Comme ceci a lieu pour tous les intervalles $I(p,N_o)$, on en déduit sans difficulté que (sauf peut être pour x_o appartenant à un ensemble $\mathcal{N}$ négligeable pour la mesure de Lebesgue) la suite des probabilités

$$\mu_N = \frac{1}{N+1} \sum_{n=0}^{N} \delta_{x_n}$$

converge étroitement vers la loi uniforme sur $(0,1)$.

Conséquences : Ce théorème entraine des conséquences suggestives pour notre problème, qui est la fabrication d'aléas.

En effet, la convergence des μ_N vers la loi uniforme sur (0,1) entraine (entre autres) que, toujours pour $x_o \notin \mathcal{N}$, les limites suivantes ont lieu :

$$\frac{1}{N+1} \sum_o^N x_n \to \frac{1}{2} \quad ; \quad \frac{1}{N+1} \sum_o^N x_n^2 \to \frac{1}{3}$$

$$\frac{1}{N+1} \sum_o^N x_n \, x_{n+k} \to C_k = \int_o^1 x \, \text{Déc}(a^k x) \, dx$$

Le calcul explicite de cette dernière intégrale se fait sans difficulté et donne

$$C_k = \frac{1}{4} + \frac{a^{-k}}{12}$$

Ainsi, pour N grand, la covariance centrée (expérimentale) calculée à partir de la suite $x_o, x_1, \ldots, x_N$ différera peu de

$$\sigma_k = \frac{a^{-k}}{12}$$

(sauf, toujours, peut être si $x_o \in \mathcal{N}$). Si l'entier a^k est grand, les nombres successifs $x_o, x_k, \ldots, x_{nk}$ pourront donc être regardés comme une réalisation de VA (à peu près) indépendantes et uniformément distribuées, c'est à dire comme un aléa.

On arrive à des conclusions tout à fait similaires si, au lieu de (1-1), on considère le processus markovien défini par

$$X_{n+1} = \text{Déc}(a \, X_n + C)$$

(a entier > 1, C réel quelconque compris entre 0 et 1). Ici encore (sauf pour x_o appartenant à un ensemble négligeable) la suite

$$\mu_N = \sum_{n=0}^{N} \frac{\delta_{xn}}{N+1}$$

converge étroitement vers la loi uniforme. Pour la covariance centrée empirique :

$$\frac{1}{N+1} \sum_{n=0}^{N} x_n \, x_{n+k} - \frac{1}{4}$$

on obtient (après quelques calculs qui ne méritent pas d'être reproduits) la limite presque sûre :

$$\sigma_k = \frac{1}{a^k}\left(\frac{1}{12} - \frac{\varepsilon_k}{2} + \frac{\varepsilon_k^2}{2}\right)$$

où ε_k est égal à $(a^k - 1)\frac{C}{a-1}$ modulo 1. Il s'agit en somme d'une covariance exponentielle à laquelle se superpose un Zitterbewegung. (Ce dernier disparaissant d'ailleurs lorsque C = 0, c'est à dire dans le cas du processus (1-1).

Le processus inverse : Si l'on prend la loi uniforme comme loi initiale, le processus markovien défini par (1-1) devient un processus stationnaire. Par renversement du sens du temps, il est donc possible d'en déduire un processus inverse, également markovien, soit Y_n : le processus inverse, on le voit facilement, est défini par :

$$Y_{n+1} = \frac{Y_n + \varepsilon_n}{a} \qquad (1\text{-}3)$$

où ε_n est une VA indépendante de Y_n et prenant l'une ou l'autre des valeurs 0, 1,..., a-1 avec la même probabilité 1/a.

Chose curieuse, alors que le processus direct X_n (bien que formellement markovien) était en fait parfaitement déterministe, puisque, lorsque X_n est connu, la relation (1-1) permet de prévoir X_{n+1} avec exactitude, le processus inverse Y_n est, lui, réellement aléatoire (au sens ordinaire, à savoir que l'on ne peut pas prédire Y_{n+1} avec certitude connaissant Y_n).

La figure 1 illustre cette situation dans le cas particulier a = 2. Dans le cas du processus direct, la donnée de X_n détermine univoquement X_{n+1}. Dans le cas du processus inverse, au contraire,

deux valeurs possibles Y'_{n+1} et Y''_{n+1} sont associées à la donnée d'un même Y_n (et ces deux valeurs possibles sont également probables).

Au bout d'un nombre un peu élevé d'itérations, X_{n+k} est toujours strictement déterminé par la donnée de X_n, alors que Y_{n+k} prend l'une ou l'autre de 2^k valeurs possibles avec une égale probabilité : ce qui revient à dire, dès que k est un peu grand, que Y_{n+k} est une VA uniformément distribuée sur (0,1), et cela quelle que soit la valeur initiale Y_n.

Cette dissymétrie très frappante s'atténue un peu si l'on regarde les choses d'un point de vue plus physique. On obtient, en effet, l'écriture du nombre X_{n+k} en base a en supprimant les k premiers chiffres de l'écriture de X_n. Ce procédé est parfaitement déterministe si l'on suppose X_n exactement connu. Mais si (comme ce sera toujours le cas dans un problème physique) on suppose seulement connus les k_o premiers chiffres significatifs de l'écriture de X_n, on aura une information assez précise sur X_{n+k} tant que k restera inférieur à k_o, mais X_{n+k} redeviendra totalement indéterminé dès que k atteindra k_o.

Cas des nombres rationnels.

D'autre part, nous ne devons pas oublier que la convergence étroite de la suite μ_N vers la loi uniforme n'a lieu que presque sûrement, c'est à dire lorsque x_o n'appartient pas à l'ensemble négligeable $\mathcal{N}$. Or, bien que négligeable pour la mesure de Lebesgue, cet ensemble $\mathcal{N}$ n'en contient pas moins tous les rationnels (si $x_o = p_o / q$, p_o et q entiers, il est clair, en effet que x_n sera encore de la forme p_n / q, et par suite la limite de la suite μ_N, si elle existe sera concentrée sur les rationnels de cette forme et ne coïncidera donc pas avec la mesure de Lebesgue. Comme, dans toute application pratique, on ne peut, en réalité, manipuler que des nombres rationnels, on sera en fait toujours dans le cas $x_o \in \mathcal{N}$. On mesure ici la distance qu'il peut y avoir entre un théorème mathématique et la pratique réelle : l'évènement "presque impossible" $x_o \in \mathcal{N}$ est, en réalité, toujours réalisé, l'exception est, en réalité, la règle générale.

Considérons ce qui se passe dans le cas où x_o est de la forme y_o / m (y_o et m entiers, $0 \leq y_o < m$). Quitte à remplacer x_n par

$x'_n = m\, x_n$, nous considérerons plutot le processus à valeurs entières modulo m défini par :

$$X'_{n+1} = a\, X'_n \quad \text{(modulo m)}$$

Nous donnerons plus loin des conditions suffisantes pour que la suite

$$\mu'_N = \frac{1}{N+1} \sum_{n=0}^{N} \delta_{X'_n}$$

converge étroitement vers la loi discrète attribuant le même poids $1/m$ à chacun des entiers $1, 2, \ldots, m-1$. Pour l'instant, nous supposons que cette convergence a lieu effectivement. Le processus markovien inverse Y'_n existe donc et vérifie la relation :

$$Y'_{n+1} = \frac{Y'_n + m\, \varepsilon_n}{a}$$

où ε_n est un entier ($0 \leq \varepsilon_n < a-1$). Or, Y'_n étant donné, l'entier ε_n ne peut prendre n'importe laquelle des valeurs $0, 1, \ldots, a-1$, puisque $Y'_n + \varepsilon_n$ doit être divisible par a. En particulier, si a est un nombre premier $< m$, la donnée de Y'_n détermine ε_n.

Pour éviter une discussion arithmétique inutile, considérons seulement le cas (qui nous retiendra longuement par la suite) où $a = 2$ et où m est impair. Les deux valeurs possibles de ε_n sont 0 et 1. Mais, $Y'_n + m\, \varepsilon_n$ devant être un nombre pair, on aura :

$$\varepsilon_n = 0 \quad \text{si } Y'_n \text{ est pair}$$

$$\varepsilon_n = 1 \quad \text{si } Y'_n \text{ est impair.}$$

Ainsi, (et contrairement à ce qui se passe lorsque $x_0 \notin \mathcal{D}$) dans le cas rationnel le processus inverse est, théoriquement, tout aussi déterministe que le processus direct. Toutefois, alors que dans le cas direct X'_{n+1} dépend, pour l'essentiel, du premier chiffre significatif de l'écriture de X'_n, c'est exactement le contraire dans le cas du processus inverse : Y'_{n+1} est déterminé, pour l'essentiel, par le dernier chiffre significatif de Y'_n. Dans le sens direct,

on a affaire à un déterminisme robuste, peu sensible à de petites erreurs affectant la valeur initiale. Dans le sens inverse, au contraire, le déterminisme (théorique) est particulièrement peu robuste, puisqu'il suffit de modifier le dernier chiffre significatif de Y'_n pour changer l'ordre de grandeur de Y'_{n+1}. Pour un physicien, un déterminisme de ce genre est parfaitement illusoire: si m est de l'ordre de 10^6, et si l'on connaît Y'_n avec une précision de 10^{-5} par exemple (ce qui n'est pas mal) on ne peut rien prédire concernant ε_n; et Y'_{n+1} doit, en fait, être considéré comme aléatoire.

Ainsi, du moins si m est grand, le contraste entre le cas rationnel et le cas général est moindre qu'il ne semblerait. Et, intuitivement, on peut espérer que les propriétés du processus rationnel X'_n ressembleront à celles du processus général X_n.

2 - CHOIX DES ENTIERS a ET m

Dans le cas général du processus défini par (1,1), soit

$$X_k = \text{Déc } a^k X_o$$

il est commode d'introduire les écritures de ces nombres en base a : Si l'écriture de X_o est :

$$X_o = \sum_{n=1}^{\infty} \frac{\varepsilon_n}{a^n} ,$$

celle de X_k sera :

$$X_k = \sum_{n=1}^{\infty} \frac{\varepsilon_{n+k}}{a^n}$$

Elle se déduit de la première en supprimant les k premiers chiffres $\varepsilon_1,\ldots,\varepsilon_k$. Dans le modèle $(\Omega, \mathcal{A}, P)$ du paragraphe précédent, les ε_n constituent une suite de VA indépendantes et uniformément distribuées selon les entiers 0, 1,..., a-1.

Ainsi, la donnée d'<u>une seule VA</u> X_o uniformément distribuée sur le segment (0,1), équivaut à la donnée d'<u>une suite infinie</u> de VA ε_n indépendantes. Cette remarque est importante pour les applications.

D'une part, en effet, il suffit de mettre en mémoire un seul nombre x_o, et non pas séparément chacun des ε_n. D'autre part, X_n et X_{n+k} ne sont qu'asymptotiquement indépendants (leur coefficient de corrélation est $1/a^k$ comme on l'a vu), tandis que ε_n et ε_{n+k} sont (dans ce modèle) rigoureusement indépendants. Dans les procédés usuels, on choisit le nombre a relativement grand, de manière à ce que la corrélation $1/a$ de X_n et X_{n+1} soit à peu près négligeable. Si, au contraire, on choisit comme aléa, non la suite des X_n, mais celle des ε_n, cette précaution devient superflue, et on peut, en particulier, prendre $\boxed{a = 2}$: ce qui nous fera bénéficier de tous les avantages informatiques de l'écriture binaire.

Pour illustrer ces avantages, considérons par exemple la construction d'une FA à une dimension. Le procédé usuel consiste à fabriquer une suite X_n et y effectuer une pondération de la forme :

$$Y_n = \sum_k b_k \, X_{n+k}$$

avec des coefficients b_k convenablement choisis. Il faut ensuite mémoriser chacun des Y_n, en vue des usages ultérieurs.

Si nous utilisons, comme aléa, la suite ε_n, nous prendrons de la même façon :

$$Y_n = \sum_k b_k \, \varepsilon_{n+k}$$

Mais ici ε_{n+k} est simplement le $(n+k)^{\text{ème}}$ chiffre de l'écriture binaire de X_o : Y_n se déduit directement de X_o par un algorithme arithmétique suffisamment simple pour qu'il ne soit pas nécessaire de mémoriser sa valeur numérique : on peut la recalculer très vite chaque fois que le besoin s'en fait sentir.

Ainsi, comme première conséquence de notre décision d'utiliser la suite des ε_n, plutot que celle des X_n, à titre d'aléa, nous pouvons choisir pour a un entier quelconque, en pratique a = 2.

Choix de m :

Puisque, dans la pratique, nous serons toujours dans le cas rationnel, considérons plutot le processus $X'_n = m\, X_n$, défini par $X'_{n+1} = a\, X'_n$ (modulo m), les X'_n étant maintenant des entiers modulo m. Quelle valeur choisir pour l'entier m ?

Tout d'abord, il est clair que cette valeur doit être relativement grande : comme les X'_n ne peuvent prendre que m valeurs au plus, la suite des X'_n (et donc aussi celle des ε_n) est nécessairement périodique, au-delà d'un certain rang, avec une période au plus égale à m (et même à m-1, puisque nous devons éviter la valeur $X'_n = 0$ qui bloquerait définitevement la suite).

De plus, pour avoir un équivalent de la loi limite uniforme du processus général, nous souhaitons plus précisément ceci : quelle que soit la valeur initiale x'_o, la période doit être exactement égale à m-1, et, de n=k à n=k+m-1, la variable x'_n doit prendre une fois et une fois seulement chacune des valeurs 1, 2,...,(m-1).

Autrement dit, la suite a^m modulo m doit être périodique, avec la période m-1. En particulier, on doit donc avoir $a^{m-1} = 1$ modulo m.

Notons tout de suite que ceci n'est possible que si m est un nombre premier. En effet, si tout entier $p \neq 0$ modulo m admet une représentation de la forme a^{n_p} pour un entier n_p $(0 \leq n_n < m-1)$, cet entier p admet également un inverse modulo m, à savoir $p' = a^{m-1-n_p}$, puisque $a^{m-1} = 1$. Ainsi, l'ensemble des entiers non nuls modulo m doit constituer un groupe multiplicatif, et cela a lieu, comme on sait, si et seulement si m est un nombre premier.

Supposons donc m premier, et considérons la suite $X'_{n+1} = a\, X'_n$ modulo m, avec $X'_o = x_o$ donné. Le nombre entier k sera une période si, pour une valeur de n, on trouve

$$a^{n+k}\, x_o = a^n\, x_o + N\, m \quad (N \text{ entier})$$

c'est à dire

$$(a^k - 1)\, a^n\, x_o = N\, m \tag{2-1}$$

Comme m est premier et divise le premier membre de (2-1), il doit diviser soit x_o, soit a^n soit enfin $(a^k - 1)$. Le premier cas est exclus si l'on s'impose la condition $0 < x_o < m$ (le cas $x_o = 0$ modulo m étant manifestement dépourvu d'intérêt, puisqu'alors $X'_n = 0$ pour tout n). D'autre part, si m divise a^n pour un $n > 0$, il divise a lui-même. Mais le cas où a est un multiple de m est évidemment sans intérêt (puisqu'alors, à nouveau, X'_n est identiquement nul). Notons d'ailleurs que la valeur de a^n modulo m ne dépend que de la valeur de a lui-même modulo m (car $(a' + N\ m)^n = a'^n$ modulo m). Nous pouvons donc nous limiter au cas $a < m$. Sous les conditions :

$$1 \leq x_o \leq m-1 \ , \quad 1 < a \leq m-1$$

la relation (2-1) implique donc que m divise $a^k - 1$, soit :

$$a^k = 1 \quad \text{modulo } m \tag{2-2}$$

et la période de la suite sera le plus petit entier k vérifiant cette propriété.

Or, un théorème connu d'arithmétique (qui remonte, sauf erreur, à Fermat) nous indique que, m étant premier et a non multiple de m, a^{m-1} est congru à 1 modulo m.

La démonstration utilise la formule du binome :

$$a^m \equiv ((a-1)+1)^m = (a-1)^m + \sum_{k=1}^{m-1} C_m^k (a-1)^k + 1$$

Comme m est premier et $C_m^k = m(m-1)\ldots(m-k+1) / k !$ entier, k! divise le produit $(m-1)\ldots(m-k+1)$; et C_m^k est un multiple de m (pour $k = 1, 2, \ldots, (m-1)$). Ainsi a^m est congru modulo m à $(a-1)^m + 1$. Pour $a = 2$, 2^m est donc congru à $(2-1)^m + 1 = 2$. Par une récurrence immédiate sur l'entier a, il en résulte que $a^m = a$ modulo m pour tout entier a.

Enfin, m étant premier et a non congru à 0 modulo m, l'entier a admet un inverse a' modulo m : en multipliant par a' la relation $a^m = a$ modulo m, on trouve bien $a^{m-1} = 1$ modulo m.

Toutefois, m - 1, solution de l'équation (2-2), n'est pas forcément le plus petit entier k qui satisfasse cette relation. Notons cependant que la période k est nécessairement un diviseur de m - 1. Car, sur l'ensemble {1, 2,...,m - 1} des entiers modulo m privé de 0, la relation d'équivalence "il existe un entier p tel que $a^p x = x'$ modulo m" définit des classes de même cardinalité k (chacune des classes est constituée des k entiers de la forme $a^p x_o$ modulo m, p = 0, 1,...;k-1).

Les candidats possibles au rôle de période k sont donc les diviseurs de m - 1. Le nombre m - 1 n'est certainement pas premier (sauf si m = 2, cas sans intérêt que nous excluons) puisque m, premier et > 2, est un nombre impair. Posons m-1 = 2q, soit q = (m - 1)/2. Peut-on avoir k = 2, c'est-à-dire $a^2 = 1$ modulo m ? Pour qu'il en soit ainsi, i. e. :

$$a^2 - 1 \equiv (a - 1)(a + 1) = N\, m$$

m doit diviser soit a - 1, soit a + 1. Il en est ainsi si l'on a soit a = 1 modulo m (cas sans intérêt), soit a = m - 1. Pour exclure ces cas sans intérêt, nous prendrons :

$$1 < a < m - 1$$

Dès lors la période sera soit m - 1 = 2q lui-même, soit q, soit enfin un diviseur de q s'il y en a.

Comme $a^{m-1} - 1 = (a^q - 1)(a^q + 1)$ est un multiple de m, m divise l'un ou l'autre des facteurs $a^q - 1$ et $a^q + 1$. Autrement dit, on a soit

$$a^q = 1 \quad \text{modulo m}$$

soit

$$a^q = -1 \quad \text{modulo m}.$$

Dans le premier cas, la période sera au plus égale à q, donc certainement inférieure à m - 1. Dans le second cas, elle sera soit 2q = m - 1, soit un diviseur de q : si q = (m - 1)/2 est lui-même un nombre premier, cette dernière éventualité est exclue, et la condition $a^q = -1$ modulo m est nécessaire et suffisante pour avoir la période k = m - 1.

En résumé, nous trouvons donc qu'un entier m répondra à nos exigences (engendrer une suite périodique de période exactement égale à m - 1) pourvu qu'il vérifie les 3 conditions suivantes :

~ m est premier (et impair)

~ $q = \frac{m-1}{2}$ est également premier

~ a^q est congru à -1 modulo m

La recherche des nombres m vérifiant ces trois conditions n'est pas particulièrement difficile.

Si nous écartons les petites valeurs (m = 3, 5, 7, 11) sans intérêt, nous notons d'abord que m et q ne peuvent être multiples ni de 2, ni de 3, ni de 5. Les valeurs possibles pour q, modulo 30, sont à priori :

q : 1 , 7 , 11 , 13 , 17 , 19 , 23 , 29

Les valeurs correspondantes de m = 2q + 1, toujours modulo 30, sont respectivement :

m : 3 , 15 , 23 , 27 , 5 , 9 , 17 , 29

Seules, la 3ème, la 7ème et la 8ème sont admissibles.

Au total, les valeurs à examiner seront donc :

pour q : 11 + 30k ; 23 + 30k ; 29 + 30k

pour m : 23 + 60k ; 17 + 30(2k + 1) ; 29 + 30(2k + 1)

En fonction de k, elles forment trois suites que l'on peut appeler suite en 3, suite en 7 et suite en 9. Voici quelques valeurs de m vérifiant les 2 premières conditions (m et $\frac{m-1}{2}$ premiers):

.../

Suite en 3 : m = 23 + 60k		
k	m	$2^{\frac{m-1}{2}}$ modulo m
0	23	+1
1	83	-1
4	263	+1
6	383	+1
8	503	+1
9	563	-1
14	863	+1
16	983	+1
21	1 283	-1
25	1 523	-1
30	1 823	+1
34	2 063	+1
1 022	61 343	+1
1 028	61 703	+1
5 000	300 023	+1
5 026	301 583	+1
33 338	2 000 303	+1 ?

Suite en 7 : m = 17 + 30(2k + 1)		
k	m	$2^{\frac{m-1}{2}}$ modulo m
0	47	+1
1	107	-1
2	167	+1
3	227	-1
5	347	-1
7	467	-1
9	587	-1
14	887	+1
19	1 187	-1
21	1 307	-1
22	1 367	+1
24	1 487	+1
31	1 907	-1
33	2 027	-1
1 010	60 647	+1
1 025	61 547	-1
5 010	300 647	+1
5 024	301 487	+1
33 339	2 000 387	-1 ?

Suite en 9 : m = 29 + 30(2k + 1)		
k	m	$2^{\frac{m-1}{2}}$ modulo m
0	59	-1
2	179	-1
5	359	+1
7	479	+1
11	719	+1
13	839	+1
16	1 019	-1
21	1 319	+1
23	1 439	+1
26	1 619	-1
33	2 039	+1
1 008	60 539	-1
1 012	60 779	-1
5 021	301 319	+1
5 032	301 979	-1
33 352	2 001 179	-1 ?

L'inventaire est exhaustif jusqu'à 2 000. Les valeurs suivantes (les plus intéressantes en vue des applications) résultent de sondages effectués par valeurs croissantes de k à partir de k = 1000 (m de l'ordre de 60 000), k = 5 000 (m de l'ordre de 300 000) et k = 33 333 (m de l'ordre de 2 millions).

Dans la colonne de droite figure la valeur de $2^{\frac{m-1}{2}}$ modulo m : comme on l'a vu, cette valeur ne peut être que +1 ou -1, et les cas qui nous intéressent sont uniquement ceux où cette valeur est -1. Les calculs ont été effectués avec une simple calculatrice de poche (à l'aide d'un algorithme rectifiant les erreurs d'arrondi : cette rectification n'était plus possibles pour les valeurs de m de l'ordre de 2 millions, d'où le point d'interrogation).

On note que les valeurs -1 sont associées aux valeurs impaires de k, dans les cas des suites en 3 et en 7, et, au contraire, aux valeurs paires dans le cas de la suite en 9 : il s'agit probablement d'un théorème général (mais je n'ai pas cherché à le démontrer). C'est sur la foi de ce théorème non démontré que j'ai indiqué, avec un point d'interrogation, les valeurs de $2^{\frac{m-1}{2}}$ modulo m pour m de l'ordre de 2 millions (toutes les autres ont été effectivement vérifiées).

Pour les applications numériques ultérieures, j'ai retenu le plus souvent la valeur

$$\boxed{m = 301\ 979}$$

3 - CONTROLES

Ayant choisi a = 2 et (par exemple) m = 301 979, nous obtenons (à partir de $x_o = 1$ ou de toute autre valeur) une suite

$$x'_n = 2^n x_o \quad \text{(modulo m)}$$

de période m - 1 prenant une fois et une fois seulement chacune des valeurs 1, 2,...,m-1 lorsque n varie d'un n_o quelconque à $n_o + m - 2$.

Il sera souvent commode de remplacer x'_n par

$$x_n = \frac{x'_n}{m}$$

qui est un nombre compris entre 0 et 1. Notre objectif est de contrôler les propriétés de la suite ε_n définie par

$$\varepsilon_n = \text{Int } 2\, x_n = \begin{cases} 0 & \text{si } x'_n < m/2 \\ 1 & \text{si } x'_n > m/2 \end{cases}$$

a - Distribution de x'_n et de ε_n

Lorsque n varie de 1 à m-1, x'_n prend exactement une fois chacune des valeurs comprises entre 1 et m-1. La loi limite des x'_n est donc la distribution uniforme sur ces m-1 entiers.

En posant, comme plus haut, $q = \frac{m-1}{2}$, on a toujours

$$x'_{n+q} = m - x'_n$$

puisque 2^q est congru à -1 : la période est 2q, mais la longueur utile de la suite est seulement $q = \frac{m-1}{2}$, puisque x'_{n+q} est congru à $-x'_n$.

En ce qui concerne les ε_n, on trouve pour la même raison :

$$\varepsilon_{n+q} = 1 - \varepsilon_n \qquad (3\text{-}1)$$

Cette relation nous garantit que, sur 2q valeurs successives de ε_n, il y aura toujours exactement q valeurs 0 et q valeurs 1. Mais elle nous indique aussi que le coefficient de corrélation ρ_q de ε_n et ε_{n+q} sera toujours égal à -1 :

$$\rho_q = -1 \qquad (3\text{-}2)$$

b - Loi des longueurs des séries de succès.

Si les ε_n étaient indépendants, la probabilité d'avoir

$$\varepsilon_n = \varepsilon_{n+1} = \varepsilon_{n+p-1} = 0$$

serait exactement

$$Q(p) = 2^{-p}$$

(d'après (3-1), on peut indifféremment raisonner sur les séries de 0 consécutifs ou sur les séries de 1).

Or, si x'_n est égal à un entier k, l'évènement ci-dessus (p 0 consécutifs) est réalisé si et seulement si :

$$2^p \, k < m$$

Lorsque k varie de 1 à m-1, le nombre de réalisations de cet évènement est exactement $\text{Int}(m\, 2^{-p})$. par suite, nous aurons, asymptotiquement la probabilité

$$Q(p) = \frac{1}{m-1} \text{Int}\left(\frac{m}{2^p}\right)$$

(au lieu de 2^{-p}) pour l'occurence de p 0 consécutifs. Comme m est grand, cette valeur ne diffère pas substantiellement de la valeur théorique 2^{-p} : en ce qui concerne la loi des séries de succès (ou d'échecs) consécutifs, l'aléa ε_n est bien conforme à ce que l'on est en droit d'exiger.

c - <u>La covariance des ε_n</u>

Désignons par C_p la covariance de ε_n et ε_{n+p}, lorsque n varie de 1 à m - 1. Elle est égale au nombre des réalisations de l'évènement $\varepsilon_n = \varepsilon_{n+p} = 0$ divisé par m - 1. L'idéal serait d'avoir

$$C_o = \frac{1}{2} \; ; \quad C_p = \frac{1}{4} \qquad (p = 1, 2, \ldots)$$

Nous savons déjà que la première relation $C_o = 1/2$ est effectivement réalisée. Par contre, on ne pourra pas avoir $C_p = 1/4$ pour pour tout $p \geq 1$. En effet, on a déjà vu que l'on a nécessairement :

$$C_{\frac{m-1}{2}} = 0 \; ; \quad C_{m-1} = \frac{1}{2}$$

ce qui correspond à des coefficients de corrélation égaux à -1 et à

+ 1 respectivement. Toutefois, $\frac{m-1}{2}$ étant assez grand (de l'ordre de 150 000 dans notre exemple), ces anomalies ne seront pas genantes en pratique, si l'on n'utilise que quelques milliers ou dizaines de milliers de valeurs consécutives.

D'une manière générale, pour p entier ≥ 1, on a $x'_{n+p} = 2^p x'_n$ modulo m. Désignons par $R_p = R$ la valeur de 2^p modulo m, soit

$$2^p = R + Nm \qquad (0 \leq R < m)$$

(m étant impair, R ne prend que les valeurs 1, 2,...,m-1). Lorsque n varie de 1 à m - 1, x'_n prend une fois chacune des valeurs k = 1, 2, ..., m-1. L'évènement qui nous intéresse ($\varepsilon_n = \varepsilon_{n+p} = 0$) est réalisé par les valeurs de k qui vérifient les 2 conditions suivantes :

$$k < \frac{m}{2} \; ; \quad s\,m \leq k\,R < s\,m + \frac{m}{2} \qquad (s \text{ entier})$$

Les valeurs possibles pour s sont tous les entiers compris entre 0 (inclus) et R/2 (exclus), soit :

$$s = 0,\ 1,\ 2,\ldots\ \mathrm{Int}\left(\frac{R-1}{2}\right)$$

et la condition imposée à k s'écrit :

$$s\,\frac{m}{R} \leq k < (s + \frac{1}{2})\,\frac{m}{R}$$

Par suite, le nombre des occurences de l'évènement qui nous intéresse est égal à

$$\sum_{s=0}^{\mathrm{Int}\left(\frac{R-1}{2}\right)} \left[\mathrm{Int}(s + \frac{1}{2})\,\frac{m}{R} - \mathrm{Int}\; s\,\frac{m}{R}\right]$$

D'où la valeur cherchée de la <u>covariance</u> C_p :

$$(3\text{-}3) \qquad \sum_{s=0}^{\mathrm{Int}\left(\frac{R-1}{2}\right)} \left(\mathrm{Int}\,(s + \frac{1}{2})\,\frac{m}{R} - \mathrm{Int}\,\frac{s\,m}{R}\right)$$

$$(R = 2^p \text{ modulo } m)$$

Avec m = 301 979, les 11 premières valeurs sont les suivantes :

$$C_1 = 0,2499983$$
$$C_2 = 0,2500017$$
$$C_3 = 0,2499983$$
$$C_4 = 0,2499950$$
$$C_5 = 0,2500050$$
$$C_6 = 0,2500016$$
$$C_7 = 0,2499917$$
$$C_8 = 0,2500016$$
$$C_9 = 0,2499851$$
$$C_{10} = 0,2500083$$
$$C_{11} = 0,2500050$$

On reste remarquablement près de la valeur théorique, qui est 0,25. Mais pour des valeurs de p plus élevées, des divergences plus importantes finissent par se manifester.

La plus sérieuse survient pour R = 3. La valeur correspondante est p = 51 790 :

$$2^p = 3 \text{ (modulo m)} \quad \text{pour } p = 51\ 790$$

Pour cette valeur, on trouve :

$$C_p = 0,33333$$

ce qui correspond à un coefficient de corrélation ($\rho = \frac{1}{3}$) nullement négligeable.

De même,

pour $2^p = 5$ (modulo m) ; $C_p = 0,299997$
" $2^p = 7$ " , $C_p = 0,28571$
" $2^p = 9$ " , $C_p = 0,27778$
" $2^p = 11$ " , $C_p = 0,27272$
" $2^p = 13$ " , $C_p = 0,26923$

D'une manière générale, il semble bien que les fortes valeurs de C_p soient associées aux très faibles valeurs de $R = 2^p$ modulo m.

Il est vraisemblable que le maximum de C_p correspond à R = 3. Comme cette valeur est associée à une valeur déjà très élevée de p (p = 51 790), elle n'introduira pas, en pratique, de perturbation perceptible.

Comme 2ème exemple, voici les valeurs de C_p lorsque m = 59. Le tableau ci-dessous doit être prolongé, pour p > 29, selon la relation $C_{p+\frac{m-1}{2}} = \frac{1}{2} - C_p$:

C_1 = 0,24138	C_{16} = 0,25862
C_2 = 0,25862	C_{17} = 0,24138
C_3 = 0,24138	C_{18} = 0,29310
C_4 = 0,22414	C_{19} = 0,24138
C_5 = 0,24138	C_{20} = 0,25862
C_6 = 0,29310	C_{21} = 0,17241
C_7 = 0,24138	C_{22} = 0,25862
C_8 = 0,32759	C_{23} = 0,20690
C_9 = 0,24138	C_{24} = 0,25862
C_{10} = 0,25862	C_{25} = 0,27586
C_{11} = 0,20690	C_{26} = 0,25862
C_{12} = 0,25862	C_{27} = 0,24138
C_{13} = 0,24138	C_{28} = 0,25862
C_{14} = 0,25862	C_{29} = 0
C_{15} = 0,24138	

Bien que la suite soit très courte (m - 1 = 58), les valeurs de la covariance sont déjà proches de la valeur souhaitée 0,25. Il est notable que la plus forte valeur observée (C_8 = 0,3276) soit du même ordre de grandeur que dans le cas de la suite précédente, beaucoup plus longue (m = 301 979, C_p = 0,33333 pour p = 51 790).

Ces résultats sont, en somme, très encourageants. On objectera toutefois que les covariances calculées ci-dessus ne sont effectivement réalisées que si la suite x'_n utilisée comporte exactement m - 1

termes (soit à peu près 300 000 dans notre exemple).

En pratique, on utilisera le plus souvent de quelques centaines à quelques dizaines de milliers de termes seulement, et les covariances expérimentales pourront fluctuer notablement par rapport à la valeur théorique fournie par la formule (3-3).

d - Fluctuations de la covariance des ε_n.

Le tableau suivant montre l'ordre de grandeur de ces fluctuations dans le cas d'une suite de N = 100 valeurs consécutives, pour différentes valeurs initiales x_o.

p \ x_o	141592	71828	134438
0	.49	0.46	0.61
1	.22	.20	.38
2	.24	.16	.37
3	.25	.21	.35
4	.26	.18	.37
5	.21	.19	.38
6	.22	.22	.43
7	.24	.23	.38
8	.23	.19	.41
9	.26	.23	.31
10	.21	.20	.36
11	.22	.19	.35
12	.21	.26	.35
13	.26	.24	.37
14	.21	.16	.38
15	.22	.23	.32
16	.21	.26	.35
17	.20	.25	.35
18	.18	.22	.32
19	.24	.22	.32
$\overline{C}_p$	.22.58	.21.26	0.3632
C_o^2	.2401	.2116	0.3721

On note d'importantes différences d'une série à l'autre. Mais (et c'est là sans doute le plus important), pour chaque série, les valeurs de la covariance C_p ($p > 1$) tendent à se regrouper autour de C_o^2.

e - Construction de fonctions aléatoires.

Mais les tests les plus intéressants consistent sans doute à construire des réalisations de fonctions aléatoires à partir des ε_n, et à comparer les variogrammes expérimentaux obtenus à leur contrepartie théorique.

Sur la figure 2 figurent les variogrammes obtenus à partir de la fonction

$$Y_n = \sum_{k=0}^{15} \varepsilon_{n+k}$$

n variant de 1 à 1 600. Les valeurs expérimentales de $\gamma(h)$ pour h variant de 1 en 1 (de 1 à 9) sont prises sur N = 1 600 couples de points. Pour h variant de 2 en 2 (de 2 à 18), on n'utilise qu'un couple sur 2, soit N = 800, et seulement un couple sur 4 (N = 400) pour h variant de 4 en 4 (de 4 à 36).

Le schéma triangulaire est assez bien reproduit. On notera toutefois des fluctuations relativement importantes autour du palier. La variance expérimentale (3,66) est d'ailleurs légèrement différente de la valeur théorique ($\sigma^2 = 4$). Mais l'ampleur de ces fluctuations ne semble pas dépasser ce qu'autorise la statistique classique. De fait, la portée étant égale à 16, avec 1 600 données nous disposons en réalité, grosso modo, de 100 échantillons indépendants. D'où, sur la valeur expérimentale de la variance un écart type relatif de l'ordre de $\sqrt{\frac{2}{100}} = 0,14$: l'écart observé (3,66 au lieu de 4 ne dépasse guère 2 écarts type).

On notera aussi que le fait de retenir un couple sur deux ou sur quatre seulement n'affecte pas beaucoup la valeur expérimentale du variogramme, comme on le voit en comparant les valeurs trouvées pour h = 4 ou h = 8 avec N = 1 600, 800 ou 400 couples seulement.

N = 1 600	N = 800	N = 400
$\gamma 1 = 0,2631$	$\gamma 2 = 0,5225$	$\gamma 4 = 0,961$
$\gamma 2 = 0,5225$	$\gamma 4 = 0,9631$	$\gamma 8 = 1,716$
$\gamma 3 = 0,7540$	$\gamma 6 = 1,3597$	$\gamma 12 = 2,511$
$\gamma 4 = 0,9497$	$\gamma 8 = 1,7444$	$\gamma 16 = 3,312$
$\gamma 5 = 1,1528$	$\gamma 10 = 2,1594$	$\gamma 20 = 3,380$
$\gamma 6 = 1,3403$	$\gamma 12 = 2,5294$	$\gamma 24 = 3,416$
$\gamma 7 = 1,5303$	$\gamma 14 = 2,8769$	$\gamma 28 = 3,535$
$\gamma 8 = 1,7266$	$\gamma 16 = 3,2981$	$\gamma 32 = 3,587$
$\gamma 9 = 1,9372$	$\gamma 18 = 3,4706$	$\gamma 36 = 3,752$

La figure 3 montre des résultats analogues obtenus dans le cas du schéma sphérique. La fonction utilisée est ici

$$Z_n = \sum_{k=0}^{15} \left(\frac{1}{2} + k\right) \varepsilon_{n+k}$$

D'une manière générale, la covariance centrée (théorique) σ_p de la fonction

$$Z_n = \sum_{k=0}^{N-1} (\alpha + \beta k) \varepsilon_{n+k}$$

est donnée par :

$$\frac{\sigma_p}{\sigma_\varepsilon^2} = (N - p)\left(\alpha^2 - \alpha\beta + \frac{\beta^2}{6}\right) + N^2\left(1 - \frac{p}{N}\right)\left(\alpha\beta - \frac{\beta^2}{2}\right)$$

$$+ \beta^2 \frac{N^3}{3}\left(1 - \frac{3}{2}\frac{p}{N} + \frac{1}{2}\frac{p^3}{N^3}\right)$$

(pour $p \leq N$, et $\sigma_p = 0$ pour $p > N$).

On a ici $\sigma_\varepsilon^2 = \frac{1}{4}$, $N = 16$, $\alpha = \frac{1}{2}$ et $\beta = 1$. La covariance (théorique) correspondante est donc (pour $p \leq 16$)

$$\sigma_p = \frac{1024}{3}\left(1 - \frac{3}{2}\frac{p}{16} + \frac{1}{2}\frac{p^3}{16^3}\right) - \frac{1}{3}\left(1 - \frac{p}{16}\right)$$

Elle n'est pas tout à fait identique à la covariance sphérique, mais n'en diffère pas de manière numériquement significative.

La technique utilisée pour construire Z_n repose sur des relations de récurrence :

~ Les X_n sont calculés de proche en proche par la formule $X_{n+1} = 2\,X_n$ modulo m.

~ Les $Y_n := \sum_{k=0}^{15} \varepsilon_{n+k}$ sont calculés par récurrence en prenant

$$Y_{n+1} = Y_n + \varepsilon_{n+16} - \varepsilon_n$$

(Y_o fait l'objet d'un calcul explicite préalable).

~ pour Z_n, enfin, on a la relation

$$Z_{n+1} - Z_n = 16\,\varepsilon_{n+16} - \frac{Y_n + Y_{n+1}}{2}$$

ou, sous forme moins symétrique mais plus pratique :

$$Z_{n+1} = Z_n + 15{,}5\,\varepsilon_{n+16} - Y_n + \frac{\varepsilon_n}{2}$$

Comme on n'a besoin de connaître Z_n qu'à une constante près, il n'est même pas nécessaire de calculer explicitement Z_o.

Numériquement, à partir d'une suite de 1 600 valeurs de Z_n, le variogramme, calculé de 4 en 4 en prenant un couple sur quatre seulement (c'est-à-dire 400 couples) a fourni les valeurs suivantes (voir aussi Figure 3) :

	Expérimental	(Théorique)
$\gamma 4$ =	123,4	(125,2)
$\gamma 8$ =	229,2	(234,4)
$\gamma 12$ =	303,6	(311,7)
$\gamma 16$ =	340,2	(341)
$\gamma 20$ =	345,4	"
$\gamma 24$ =	352,3	"
$\gamma 28$ =	349,3	"

Ces résultats paraissant exceptionnellement bons, j'ai modifié la valeur initiale (x_o = 216 916) et recommencé avec x_o = 141 592, en utilisant cette fois N × 20 valeurs et en prenant seulement un couple sur 20, de manière à avoir pour chaque point du variogramme N données à peu près indépendantes. Avec N = 100 et N = 500, les résultats ont été les suivants : (voir Figure 4)

	N = 500	N = 100	Théorique
γ1	31.52	32.26	31.94
γ2	65.53	70.06	63.62
γ3	89.25	76.12	94.81
γ4	123.36	116.12	125.25
γ5	160.16	135.8	154.69
γ6	182.14	150.2	182.87
γ7	213.04	165.8	209.56
γ8	252.51	206.7	234.50
γ9	272.97	228.1	257.44
γ10	308.08	270.9	278.12
γ11	328.48	294.7	296.31
γ12	341.36	308.1	311.75
γ13	330.24	272.6	324.19
γ14	331.33	245.2	333.37
γ15	340.62	250.0	339.06
γ16	340.86	251.6	341
γ17	344.10	266.7	"
γ18	343.45	274.4	"
γ19	338.08	250.2	"
γ20	341.45	247.8	"

Dans le cas N = 100, les valeurs expérimentales sont trop basses à partir du point 13. Mais ce défaut est compensé dans le cas N = 500, et cela bien que les 500 données incorporent les 100 précédentes.

Le dernier essai (Figure 6) concerne <u>le variogramme linéaire</u> : pour chacun des 30 points du variogramme expérimental, le nombre des données est de N = 700. Comme on a pris un couple sur 30, ces 700 données sont, pour chaque couple, des données indépendantes, d'où

l'ajustement presque parfait sur la droite $\gamma(h) = h/8$.

Variogramme exponentiel.

Pour construire la FA $Y_n = \sum_{k=0}^{\infty} \rho^k \varepsilon_{n-k}$, j'utilise la relation de récurrence $Y_{n+1} = \rho Y_n + \varepsilon_n$ (partant de $Y_o = 0$, le régime stationnaire est atteint en moins de 100 itérations pour $\rho \leq 0.9$). Avec $\rho = 0,8$, $N = 17 \times 500$, et en retenant un point sur 17, de manière à avoir 500 couples pratiquement indépendants pour chaque point du variogramme, j'ai obtenu les résultats suivants : (voir aussi Figure 5).

Variogramme exponentiel (ρ=0,8)		
Valeur	Expérimental	Théorique
m	2.535	2.5
σ^2	0.6887	0.6944
$\gamma 1$	0.1409	0.1389
$\gamma 2$	0.2592	0.2500
$\gamma 3$	0.3412	0.3389
$\gamma 4$	0.4199	0.4100
$\gamma 5$	0.4773	0.4689
$\gamma 6$	0.5286	0.5124
$\gamma 7$	0.5807	0.5488
$\gamma 8$	0.5856	0.5779
$\gamma 9$	0.6372	0.6012
$\gamma 10$	0.6139	0.6199
$\gamma 11$	0.6178	0.6348
$\gamma 12$	0.6639	0.6467
$\gamma 13$	0.6599	0.6563
$\gamma 14$	0.6697	0.6639
$\gamma 15$	0.6570	0.6700
$\gamma 16$	0.6629	0.6749
$\gamma 17$	0.6746	0.6788

4 - CONCLUSIONS

Au total, la suite ε_n, définie par :

$$\varepsilon_n = \text{Int}(2x_n / m) \qquad x_n = 2^n \text{ modulo } m = 301\ 979$$

semble bien pouvoir jouer le rôle d'aléa : ses propriétés sont au moins aussi bonnes que celles des aléas fabriqués par des procédés plus usuels, et elle se recommande par une plus grande simplicité. En particulier, lorsqu'il s'agit de combiner entre elles plusieurs fonctions aléatoires du type

$$Y_n = \sum_k b_k \varepsilon_{n+k} \qquad (4\text{-}1)$$

(par exemple, dans la méthode des bandes tournantes), il peut y avoir intérêt à utiliser le caractère franchement déterministe de la relation (4-1) : plutot que de fabriquer au préalable et mettre en mémoire des réalisations de chacune de ces FA, il peut y avoir intérêt à recalculer chaque $Y_n = f(n)$ chaque fois que l'on en a besoin.

Il existe d'ailleurs d'autres possibilités sur lesquelles, pour terminer, je me contenterai de donner quelques indications. Alors que les ε_n peuvent être considérées comme des VA à peu près indépendantes, les $x_n = 2^n$ modulo m constituant plutot une réalisation d'une FA stationnaire dont le corrélogramme est $\rho_p = 2^{-p}$. En particulier, les x_n ne peuvent pas être considérées comme des VA indépendantes. Mais les x_n présentent sur les ε_n (qui ne prennent que les 2 valeurs 0 et 1) l'avantage d'être uniformément distribuées sur les $m - 1 = 301978$ premiers entiers. Si, au lieu de x_n, on considère la variable $y_n = x_n / m$, on obtient une excellente approximation de la distribution uniforme sur le segment (0,1) (à partir de laquelle, par anamorphose, on peut obtenir n'importe quelle distribution continue). Peut-on combiner les avantages des ε_n (indépendance) et des x_n ou y_n (loi uniforme) ? Celà est certainement possible. De fait, considérons la suite :

$$x_n = a^n \text{ modulo } m$$

où l'entier a n'est plus nécessairement égal à 2. Elle conduit à une

distribution uniforme sur les m - 1 premiers entiers, et à un corrélogramme $\rho_p = a^{-p}$: si a est grand, on aura à la fois l'indépendance (approximative), et la distribution uniforme sur (0,1) de la suite $y_n = x_n / m$.

Il faut toutefois s'assurer que la suite x_n prend effectivement une fois et une seule chacune des valeurs 1, 2,...,m - 1 lorsque n varie de 0 à m - 2. Autrement, comme on l'a vu au paragraphe 2, m étant choisi dans l'une ou l'autre des suites en 3, en 7 ou en 9, il faut vérifier que l'on a :

$$a^{\frac{m-1}{2}} = -1 \text{ (modulo m)}$$

En essayant différents entiers a, on met en évidence des propriétés arithmétiques assez curieuses, qui pourraient sans doute faire l'objet de démonstrations rogiureuses (ici, je me suis contenté de vérifications numériques pour les valeurs de m trouvées au paragraphe 2).

Par exemple, pour a = 3, on trouve $3^{\frac{m-1}{2}} = +1$ modulo m, pour <u>toutes</u> les valeurs de m des suites en 3, en 7 et en 9.

Pour a = 5, chose plus curieuse, on trouve $5^{\frac{m-1}{2}} = +1$ pour toutes les valeurs de la suite en 9, et, au contraire, $5^{\frac{m-1}{2}} = -1$ pour toutes les valeurs des suites en 3 et en 7.

Pour les nombres premiers ultérieurs (7, 11, 13, etc...) on n'observe rien de semblable , les valeurs +1 et -1 alternant de manière imprévisible dans chacune des 3 suites.

Du fait que $5^{\frac{m-1}{2}} \equiv 1$ pour toutes les valeurs de m de la suite en 9, on aura également $10^{\frac{m-1}{2}} \equiv -1$ si, de plus m vérifie $2^{\frac{m-1}{2}} \equiv -1$ modulo m (par exemple, pour m = 301 979). Pour a = 100, 10^4, 10^6, etc..., on trouvera toujours $a^{\frac{m-1}{2}} \equiv +1$, mais, ce qui est plus intéressant; on aura également :

$$a^{\frac{m-1}{2}} \equiv -1 \quad \text{pour a} = 1\,000,\ 10^5,\ 10^7 \text{ etc...}$$

Or 1 000 et, plus encore, 100 000 est un nombre déjà élevé. La suite

$$y_n = x_n / m \quad ; \quad x_n = (100\ 000)^n \text{ modulo } m$$

$$(m = 301\ 979)$$

semble donc vérifier toutes les conditions voulues (indépendance et distribution uniforme).

Je n'ai fait que des tests assez sommaires, qui devront sans doute être repris, mais les premiers résultats sont encourageants : voici les valeurs obtenues pour le variogramme de la fonction

$$Z_n = \sum_{k=0}^{n} y_k$$

(chacun des 30 points est calculé sur 200 couples indépendants) : (voir aussi Figure 7)

$\gamma 1$	$\gamma 2$	$\gamma 3$	$\gamma 4$	$\gamma 5$	$\gamma 6$	$\gamma 7$	$\gamma 8$	$\gamma 9$	$\gamma 10$
.0418	.0938	.1297	.1616	.2041	.2419	.2802	.3493	.3968	.4156
$\gamma 11$	$\gamma 12$	$\gamma 13$	$\gamma 14$	$\gamma 15$	$\gamma 16$	$\gamma 17$	$\gamma 18$	$\gamma 19$	$\gamma 20$
.4513	.5372	.5936	.6417	.7317	.7893	.8089	.8923	.8960	.9052
$\gamma 21$	$\gamma 22$	$\gamma 23$	$\gamma 24$	$\gamma 25$	$\gamma 26$	$\gamma 27$	$\gamma 28$	$\gamma 29$	$\gamma 30$
.9342	.9972	1.0904	1.1516	1.1809	1.204	1.2198	1.2284	1.3328	1.4112

Concernant les y_n eux-mêmes, un test portant sur 1 000 valeurs a donné la distribution suivante :

classe	0.0	0.1	0.2	0.3	0.4	0.5
fréquence	99	83	116	98	106	
classe	0.5	0.6	0.7	0.8	0.9	1.0
fréquence	99	112	81	92	112	

avec une moyenne $m = 0.5365$ et une variance $\sigma^2 = 0.0899$ (valeurs théoriques : $m = 0.5$ et $\sigma^2 = 0.0833$).

Sur cette même suite de 1 000 points, les 10 premières valeurs du variogramme sont les suivantes :

0.0857 , 0.0860 , 0.0859 , 0.0879 , 0.0823

0.0853 , 0.0841 , 0.0837 , 0.0889 , 0.0860

Tout ceci est bien conforme à ce que l'on attendait.

ANNEXE - LA DECIMALISATION

Dans tout ce qui précède, on a quelque peu perdu de vue le point de vue initial, selon lequel tout procédé de fabrication d'aléa repose sur un algorithme du type x_n = Déc f(n) - ou plutot, on s'est limité au cas où la fonction f était de la forme $f(n) = a^n / m$. Il s'agit bien, d'ailleurs, du cas le plus intéressant pour les applications. Pourtant, dans cette annexe, je voudrais donner quelques indications plus générales.

Dans ce qui suit, la fonction Déc x est considérée comme une fonction périodique, de période égale à 1 (donc Déc (-x) = 1 - Déc x). Sous forme de développement en série de Fourier, on a ainsi :

$$\text{(A-1)} \qquad \text{Déc } x = \frac{1}{2} - \sum_{p=1}^{\infty} \frac{\sin 2\pi p x}{\pi p}$$

Comme toute fonction périodique, Déc x peut être regardée comme une fonction aléatoire (formellement, il suffit de remplacer Déc x par Déc (x + ω) où ω est une VA uniformément distribuée sur (0,1)). A ce titre, elle admet une covariance centrée g(h), dont le développement est donné par :

$$\text{(A-2)} \qquad g(h) = \sum_{p=1}^{\infty} \frac{\cos 2\pi p h}{2\pi^2 p^2}$$

(en particulier, $g(o) = \sigma^2 = 1/12$). Cette covariance périodique g(h) coïncide sur le segment (0,1) avec le polynome du second degré :

$$\text{(A-3)} \qquad g(h) = \frac{1}{12} - \frac{1}{2} h(1-h) \qquad (0 \leq h \leq 1)$$

Par elle-même, cette covariance est trop particulière (et trop périodique) pour présenter un intérêt du point de vue des applications. Mais on peut, par quelques transformations simples, en déduire d'autres covariances plus intéressantes. Par exemple, en appliquant la méthode des <u>bandes tournantes</u> à la (pseudo) fonction aléatoire Déc x, nous o obtiendrons, dans R^2 ou dans R^3, des réalisations de FA admettant une covariance du type :

$$\text{(A-4)} \qquad \sigma(h) = \int g(t\,h)\, f(t)\, dt$$

(où f est la densité, concentrée sur (0,1) du cosinus d'une direction aléatoire, soit $f(t) = 1$ ou $(2/\pi)(1 - t^2)^{-1/2}$ selon que l'on est dans $\mathbb{R}^3$ ou $\mathbb{R}^2$).

Plus généralement, si les t_i sont des VA indépendantes admettant la même loi f, la FA

$$Y(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{Déc}(t_i\, x)$$

admettra à peu près la covariance (1-4) (en pratique, plutot que de tirer au sort des t_i indépendantes, on prendra plutot les n demi quantiles définis par $1 - F(t_i) = (i - 1/2)/n$, i=1, 2...n).

Une autre manière de procéder est la suivante : soit $\varphi(x)$ une fonction quelconque définie (par exemple) sur l'intervalle (0,1), et prolongée sur la droite réelle par périodisation ($\varphi(x+n) = \varphi(x)$ pour n entier positif ou non).

Choisissons deux nombres A et L (relativement grands) et posons :

(A-5) $$Y(x) = \text{Déc}\ (A\ \varphi(\frac{x}{L}))$$

Nous obtenons une fonction périodique , de période L. Supposons, pour abréger la discussion, que la fonction φ admette une dérivée φ' et que la courbe $y = \varphi(x)$ puisse être assimilée à sa tangente au point x sur un petit intervalle entourant ce point. On aura

$$A\ \varphi\ (\frac{x + \delta\, x}{L}) = A\ \varphi\ (\frac{x}{L}) + 1$$

pour $\delta\, x = L\ /\ (A\ \varphi')$. Si $\delta\, x$ est assez petit (i.e. A assez grand) pour que la courbe soit assimilable à sa tangente sur un intervalle de cette longueur, Y(x) aura "localement" la covariance $g(h\ /\ \delta\, x)$ et, globalement, une covariance centrée qui différera peu de

$$\sigma(h) = \frac{1}{L} \int_0^L g\ (\frac{A\ \varphi'(x)}{L}\ h)\ dx$$

Ceci est bien de la forme (A - 4) si l'on prend comme loi f celle de la variable aléatoire $t = \frac{A}{L}\ \varphi'(x)$, x étant uniformément distribuée sur (0,L).

Exemple 1 : Si $\varphi(x) = \cos 2\pi x$, à un facteur d'échelle près, la loi f est $(2 / \pi)\ (1 - t^2)^{-1/2}$ sur (0,1) : autrement dit, la covariance correspondante est celle que l'on obtient en appliquant la méthode des bandes tournantes (dans R^2) à la fonction Déc x.

Exemple 2 : de même, si la fonction à décimaliser est la fonction périodique égale à h(1 - h) sur (0,1), on obtiendra par décimalisation la même covariance qu'en appliquant à Déc x la méthode des bandes tournantes dans R^3 (la loi f est ici, en effet, la loi uniforme).

Expression explicite du variogramme obtenu par décimalisation.

A la covariance g(h) est associé le variogramme (de période 1) défini par

$$\gamma(h) = \frac{1}{2} h(1 - h) \qquad (0 \leq h \leq 1) \tag{A-6}$$

Si t est une VA (que nous pouvons toujours supposer positive puisque la fonction $\gamma(h)$ est paire) admettant la densité f(t), proposons-nous d'évaluer le variogramme $\tilde{\gamma}$, obtenu par décimalisation, dont l'expression exacte est :

$$\tilde{\gamma}(h) = \int_0^\infty \gamma(t\,h)\, f(t)\, dt$$

Compte tenu de (A-6), nous trouvons explicitement :

$$\tilde{\gamma}(h) = \frac{1}{2} \sum_{N=0}^{\infty} \int_{\frac{N}{h}}^{\frac{N+1}{h}} (h\,x - N)(1 - h\,x + N)\, f(x)\, dx \tag{A-7}$$

Introduisons les notations suivantes :

$$\begin{cases} 1 - F(x) = \displaystyle\int_x^\infty f(t)\, dt \\ V(x) = \displaystyle\int_x^\infty (1 - F(t)\, dt = \int_x^\infty (t - x)\, f(t)\, dt \end{cases}$$

(avec d'ailleurs $V(o) = m = \displaystyle\int_0^\infty t\, f(t)\, dtn$ et de même :

$$\int_0^\infty V(x)\,dx = \frac{1}{2}\,m_2 = \int_0^\infty \frac{t^2}{2} f(t)\,dt \quad).$$

On voit alors sans peine, d'après (A-7), que le variogramme $\tilde{\gamma}(h)$ peut encore s'écrire :

$$\tilde{\gamma}(h) = \frac{1}{2}\,(m\,h - h^2\,m_2) + h \sum_1^\infty V\left(\frac{N}{h}\right)$$

ce que nous mettrons plutot sous la forme :

$$\tilde{\gamma}(h) = h\left[\frac{1}{2}V(o) + \sum_1^\infty V\left(\frac{N}{h}\right)\right] - h^2 \int_0^\infty V(x)\,dx \qquad \text{(A-8)}$$

Cette expression (A-8) ressemble étrangement à celle d'une variance d'estimation transitive. De fait, si C(h) est un covariogramme transitif (dans l'espace à une seule dimension), la variance d'estimation associée à la maille régulière a est $\sigma^2(a) = \delta_a(C)$, où l'opérateur δ_a est défini par :

$$\delta_a\,C = a\,C(o) + 2\sum_{N=1}^{\infty} C(Na) - \int_{-\infty}^{+\infty} C(h)\,dh$$

Autrement dit, en appliquant cet opérateur à la fonction V avec a = 1/h, on trouve, d'après (A-8) :

$$\tilde{\gamma}(h) = \frac{h^2}{2}\,\delta\,\frac{1}{h}\,V \qquad \text{(A-9)}$$

Nous pouvons ainsi transposer les résultats connus concernant la valeur approchée d'une variance d'estimation $\sigma^2(a)$ pour les petites mailles a pour obtenir le comportement de $\tilde{\gamma}(h)$ aux grandes valeurs de h.

Ce comportement résultera de la somme de deux termes : un terme régulier, lié au comportement de V(t) au voisinage de t = 0, et un Zitterbewegung, lié à la manière dont V(t) tend vers 0 lorsque t tend vers l'infini.

Au voisinage de t = 0, on a

$$V(t) = \frac{1}{2} m_2 - \int_0^t (1 - F(x))\, dx \simeq \frac{1}{2} m_2 - t$$

Le terme principal est -t : pour a petit, $\delta_a V \simeq \frac{a^2}{6}$, et donc, pour h grand :

$$\tilde{\gamma}(h) \simeq \frac{1}{12}$$

De fait, $\frac{1}{12}$ est la variance de la VA uniformément distribuée sur (0,1), et c'est bien cette loi que l'on observe, pour A assez grand, en décimalisant A $\varphi(\frac{x}{L})$.

Pour obtenir le terme suivant de la partie régulière, supposons que la densité f(t) soit équivalente à $B\, t^{\lambda}$ au voisinage de t = 0. Alors F(t) est équivalent à $B\, t^{1+\lambda} / (1 + \lambda)$ et :

$$V(t) \simeq \frac{1}{2} m_2 - t + \frac{B}{(1 + \lambda)(2 + \lambda)} t^{2+\lambda}$$

D'où :

$$\delta_a V = - \frac{a^2}{6} + \frac{B\, T_{\lambda+2}}{(1 + \lambda)(2 + \lambda)} a^{3+\lambda}$$

(avec $T_\lambda = - \sin \frac{\lambda \pi}{2} \frac{\Gamma(1 + \lambda)}{2^{\lambda-1} \pi^{1+\lambda}} \sum_1^\infty \frac{1}{p^{1+\lambda}}$)

On en déduit :

$$\tilde{\gamma}(h) \simeq \frac{1}{12} + \frac{B\, T_{\lambda+2}}{2(1 + \lambda)(2 + \lambda)} \frac{1}{h^{1+\lambda}} \qquad \text{(A-10)}$$

Par exemple, pour $\lambda = 0$, c'est-à-dire dans le cas où la densité prend en t = 0 une valeur f(o) = B finie non nulle, le raccordement de $\tilde{\gamma}(h)$ à son asymptote est en 1/h, c'est-à-dire très lent. D'une manière générale, ce raccordement sera d'autant plus rapide que λ sera plus grand, donc que f(t) tendra plus rapidement vers 0 lorsque $t \to 0$. Si, au contraire, f(o) est infini, $\tilde{\gamma}(h)$ tendra très

lentement vers sa valeur limite, ou même n'admettra plus de limite du tout. (Nous verrons ci-dessous que la portée intégrale associée au variogramme $\tilde{\gamma}$ est proportionnelle à f(o) : nulle pour $\lambda > 0$, infinie pour $\lambda < 0$, elle ne prend une valeur finie que dans le cas ou f(o) est finie et non nulle).

A la partie régulière (A-10) de $\tilde{\gamma}(h)$, pour h grand, se rajoute un Zitterbewegung : celui-ci est absent si f(t) a un raccordement bien régulier avec l'axe des t. C'est le cas par exemple, d'une loi exponentielle. Pour

$$f(t) = b\, e^{-bt} \quad (t \geq 0)$$

on trouve, en effet, la formule (exacte)

$$\tilde{\gamma}(h) = \frac{h}{2b} - \frac{h^2}{b^2} + \frac{h}{b}\,\frac{e^{-\frac{b}{h}}}{1 - e^{-\frac{b}{h}}}$$

Par contre, si la loi t est concentrée sur un intervalle fini, par exemple (0,1), le terme fluctuant sera d'autant plus important que le raccordement de f(t) avec l'axe des t sera plus brutal en t = 1. Tout ceci ayant été étudié dans le cas des variances d'estimation transitive, je me contente de donner deux exemples :

Exemple 1 : Sur la Figure 8 on peut voir le variogramme de la "FA" obtenue en décimalisant A cos b x. La loi f a une densité en $1/\sqrt{1 - t^2}$, donc très irrégulière en t = 1, et le Zitterbewegung est particulièrement intense et prolongé. Comme ici f(o) est finie et non nulle, il y a, théoriquement, une portée finie, mais $\tilde{\gamma}(h)$ se raccorde très lentement à son asymptote (en 1/h). Les valeurs expérimentales (non reproduites sur la figure) s'ajustent remarquablement à la courbe théorique, ce qui n'a évidemment ici rien de bien surprenant.

Exemple 2 : Ici, la loi f est uniforme sur (0,1), ce qui correspond, à un facteur d'échelle près, à la décimalisation d'une parabole périodisée. A partir de la formule générale (A-8), on trouve, après quelques calculs élémentaires, l'expression suivante pour $\tilde{\gamma}$:

$$\tilde{\gamma}(h) = \frac{1}{12}\,\frac{N}{N+\varepsilon} + \frac{1}{N+\varepsilon}\left(\frac{\varepsilon^2}{4} - \frac{\varepsilon^3}{6}\right)$$

(N = Int h , ε = Déc h)

On note que l'on a exactement $\tilde{\gamma}(h) = \frac{1}{12}$ pour les valeurs entières et demi-entières de h (ε = 0 ou 1/2). La figure 9 montre l'allure générale de ce variogramme : le Zitterbewegung est encore sensible, mais s'amortit nettement plus vite que dans l'exemple précédent.

Exemple 3 : Ici, $f(t) = 5(1-t)^4$ $(0 \le t \le 1)$, de manière à avoir un très bon raccordement en t = 1. De fait, il n'y a pas de Zitterbewegung apparent, comme on peut le voir sur la figure 10.

La densité spectrale

Compte tenu du développement (A-2) de g(h), la covariance $\sigma(h) = \frac{1}{12} - \tilde{\gamma}(h)$ de la formule (A-4) se met sous la forme :

$$\sigma(h) = \sum_{p=1}^{\infty} \frac{\Phi(2\pi p h)}{2\pi^2 p^2}$$

où Φ est la transformée de Fourier de la loi f symétrisée (i.e. la partie réelle de la fonction caractéristique de la loi f). On en déduit que la transformée de σ(h), c'est-à-dire la densité χ(u) de la mesure spectrale associée à cette covariance, est :

$$\chi(u) = \frac{1}{4\pi^2} \sum_{p=1}^{\infty} \frac{f\left(\frac{|u|}{p}\right)}{p^3}$$

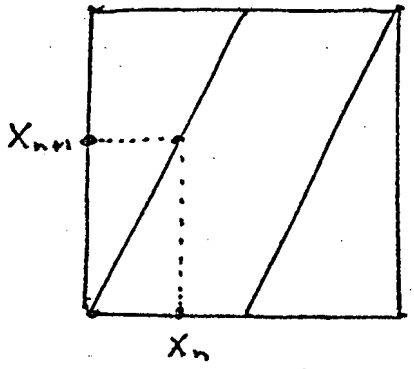
La valeur de la portée intégrale en résulte (plutot que par l'intégrale $\int_{-\infty}^{+\infty} \sigma(h)\, d\,h$, qui peut ne pas exister si σ(h) n'est pas absolument convergente, comme dans le cas σ(h) = cos h, il convient de définir cette portée comme la limite pour L infinie de

de définir cette portée comme la limite pour L infinie de

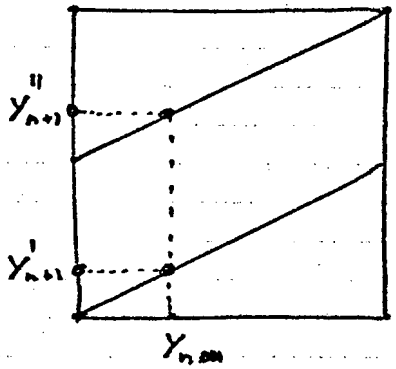
$$\frac{2}{L}\int_{o}^{L} (L - h)\ \sigma(h)\ d\ h\ ).$$

Cette limite existe et vaut $\chi(o)$ dès que la mesure spectrale admet une densité $\chi(u)$ continue au moins sur un voisinage de l'origine. Dans le cas qui nous occupe, on trouve donc :

$$\text{Portée} = \frac{f(o)}{4\pi^2} \quad \sum_{1}^{\infty} \frac{1}{p^3}$$

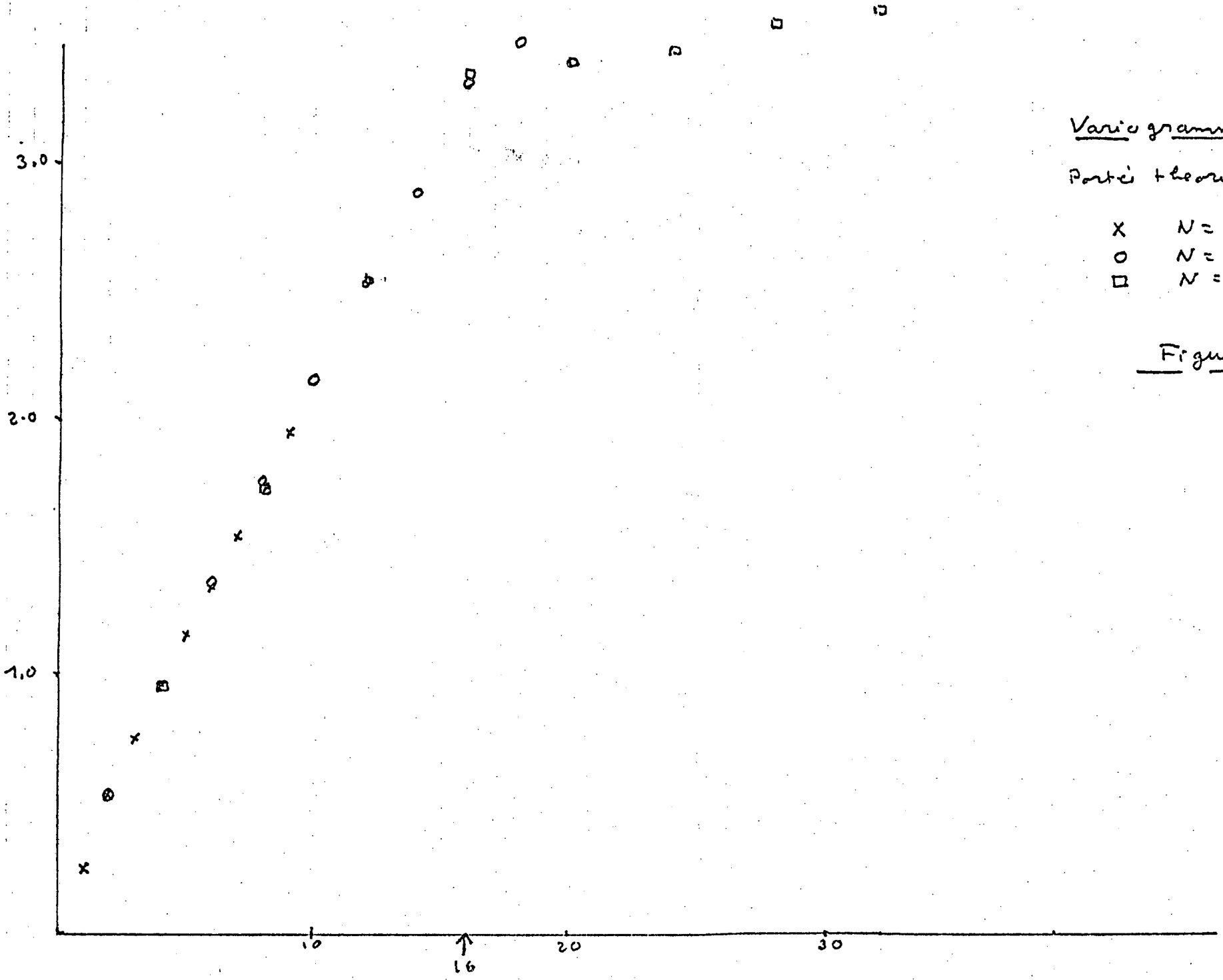


Processus direct



Processus inverse

Figure 1

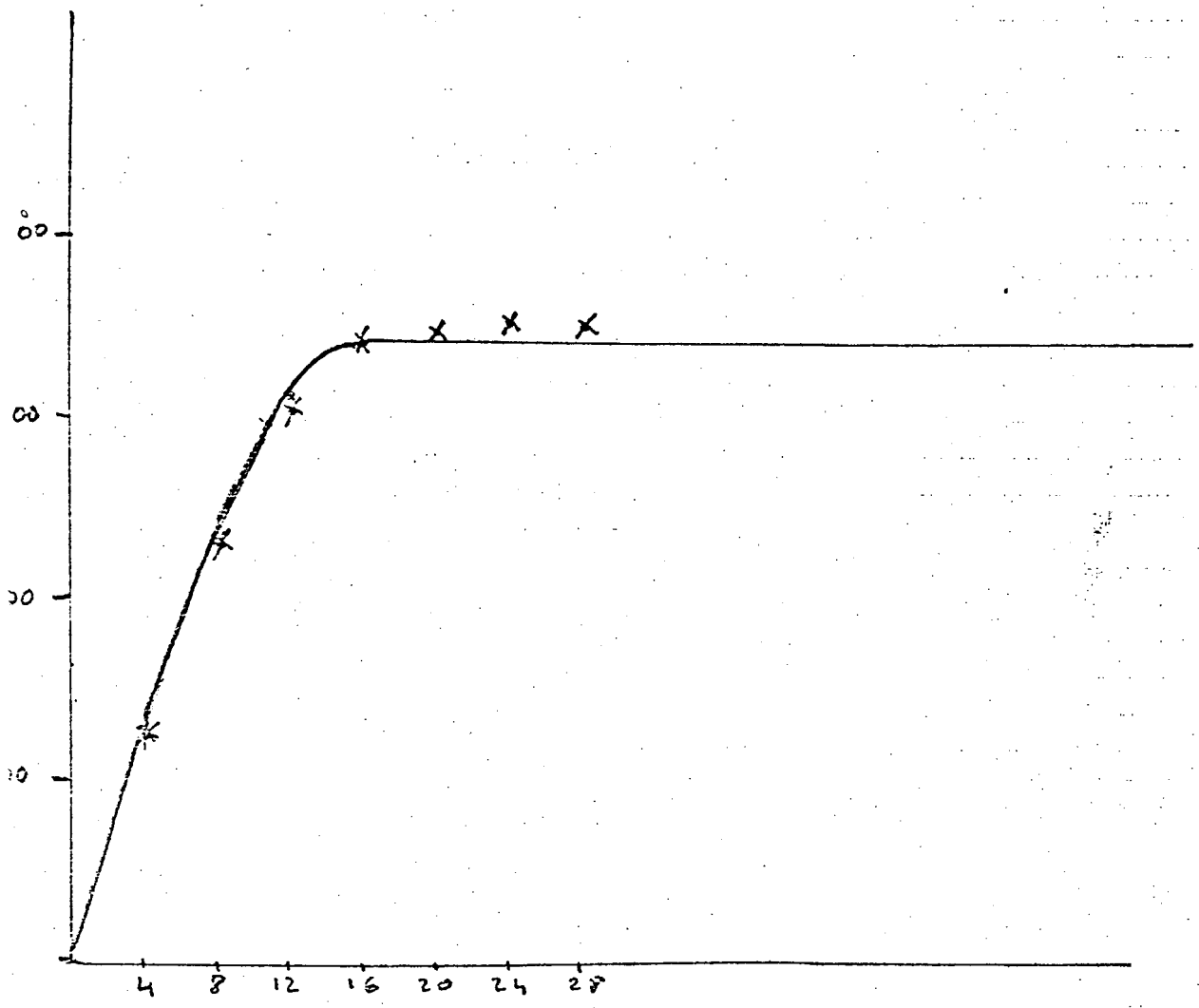


Variogramme triangulaire

Portée théorique : 16

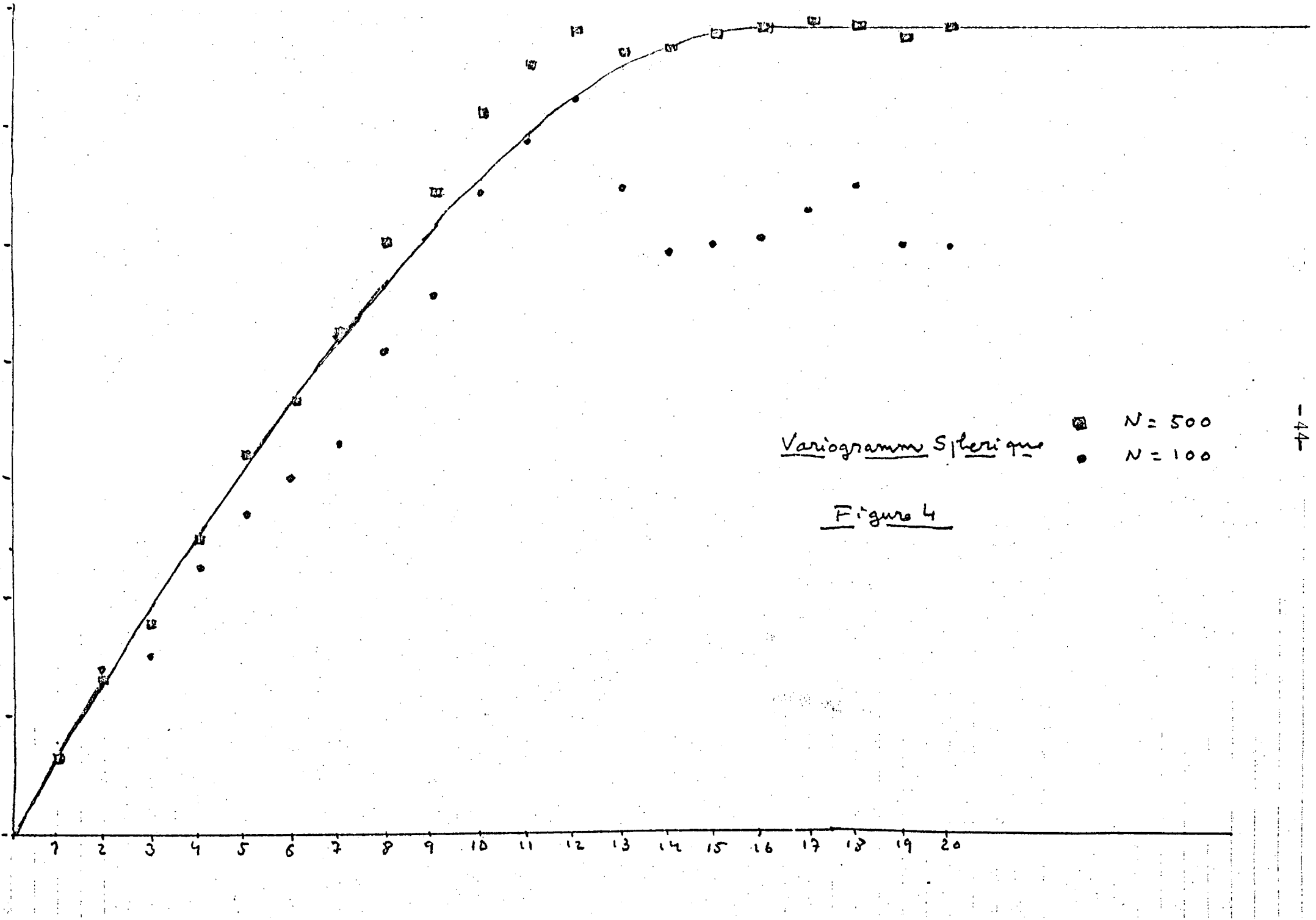
X N = 1600
O N = 800
□ N = 400

Figure 2



Variogramme sphérique

Figure 3



Variogramme Sphérique

Figure 4

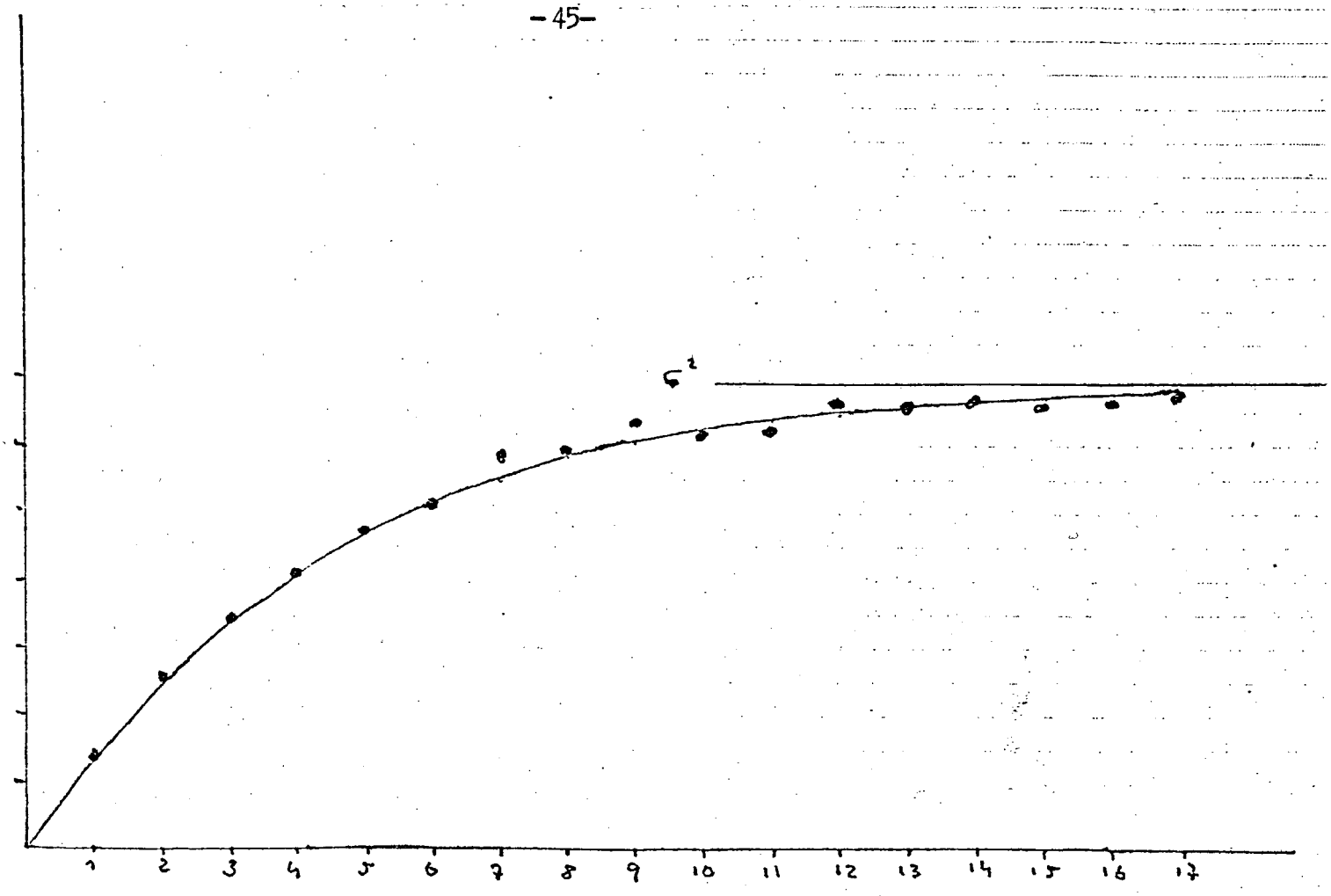


Figure 5 Variogramme exponentiel

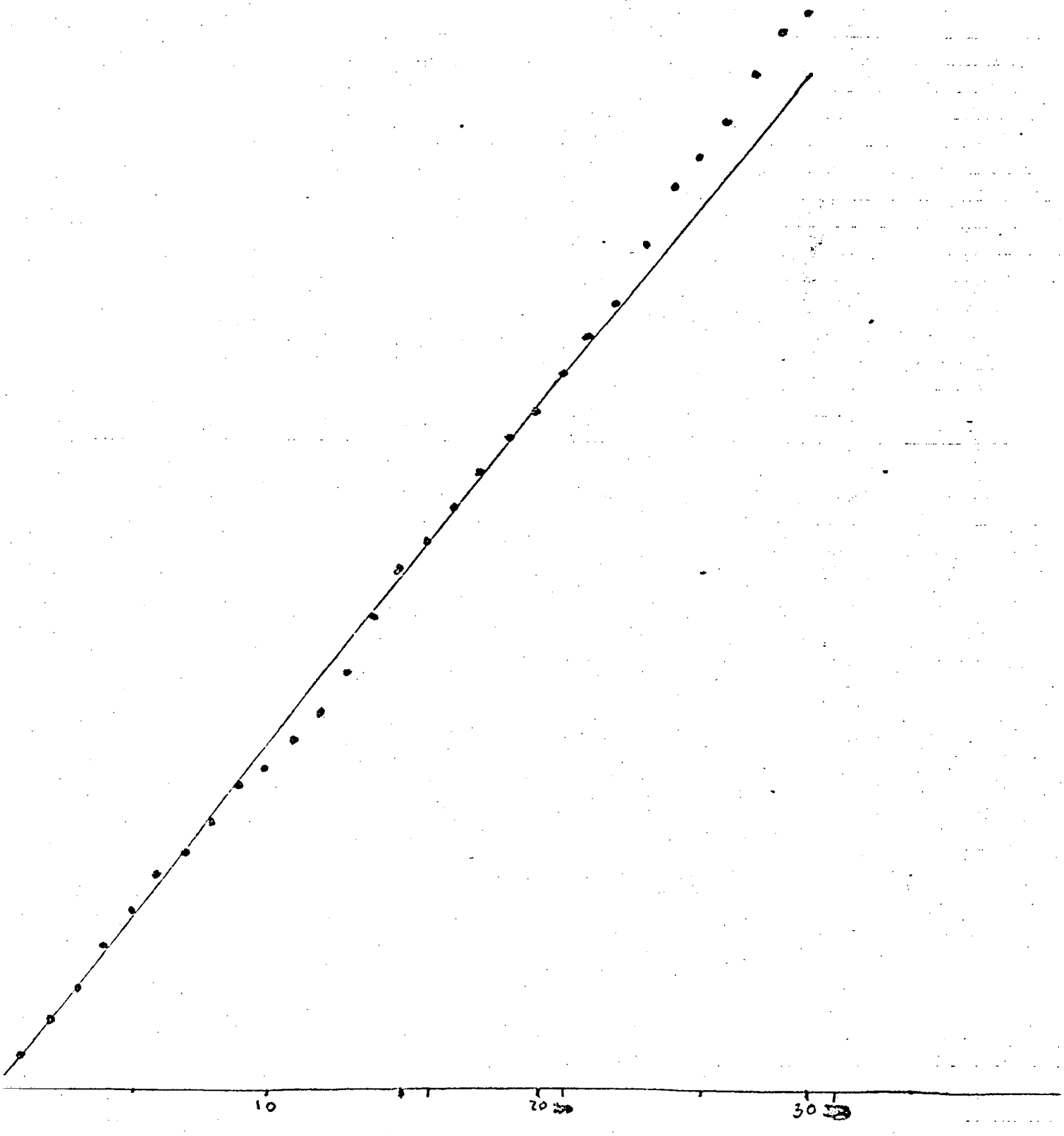


Figure 6 Variogramme linéaire (N=700)

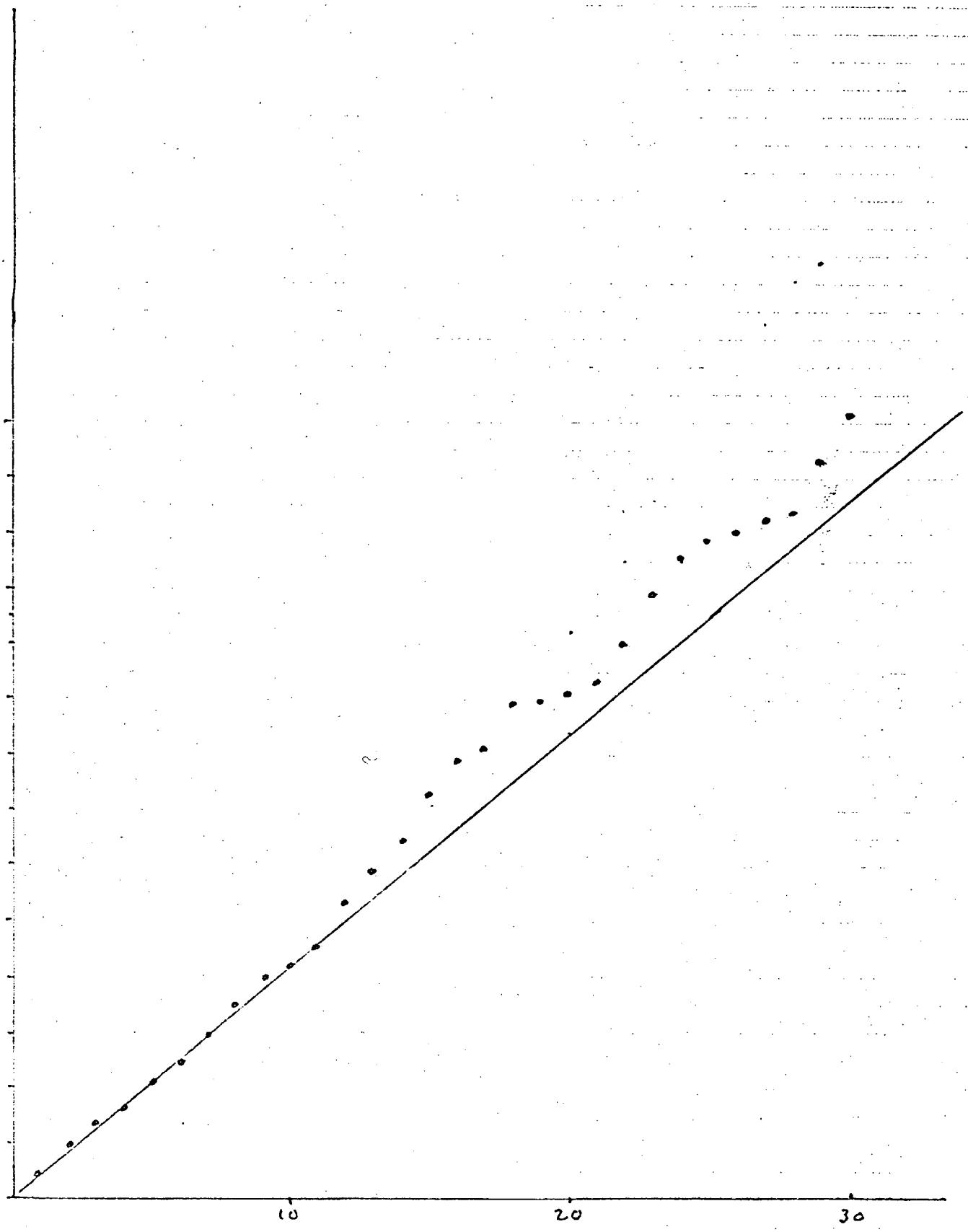


Figure 7 Variogramme linéaire, N = 200

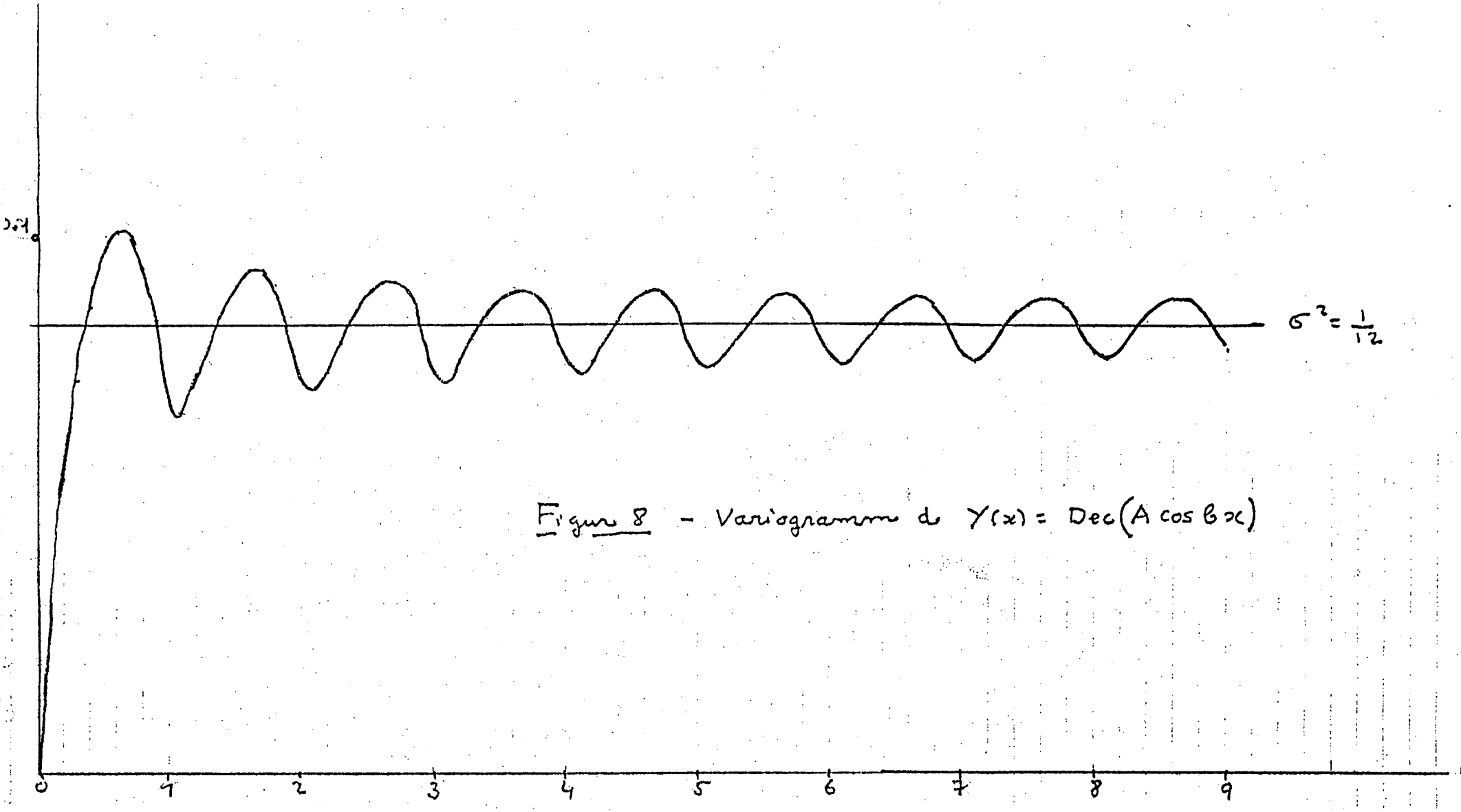


Figure 8 - Variogramme de $Y(x) = Dec(A \cos Bx)$

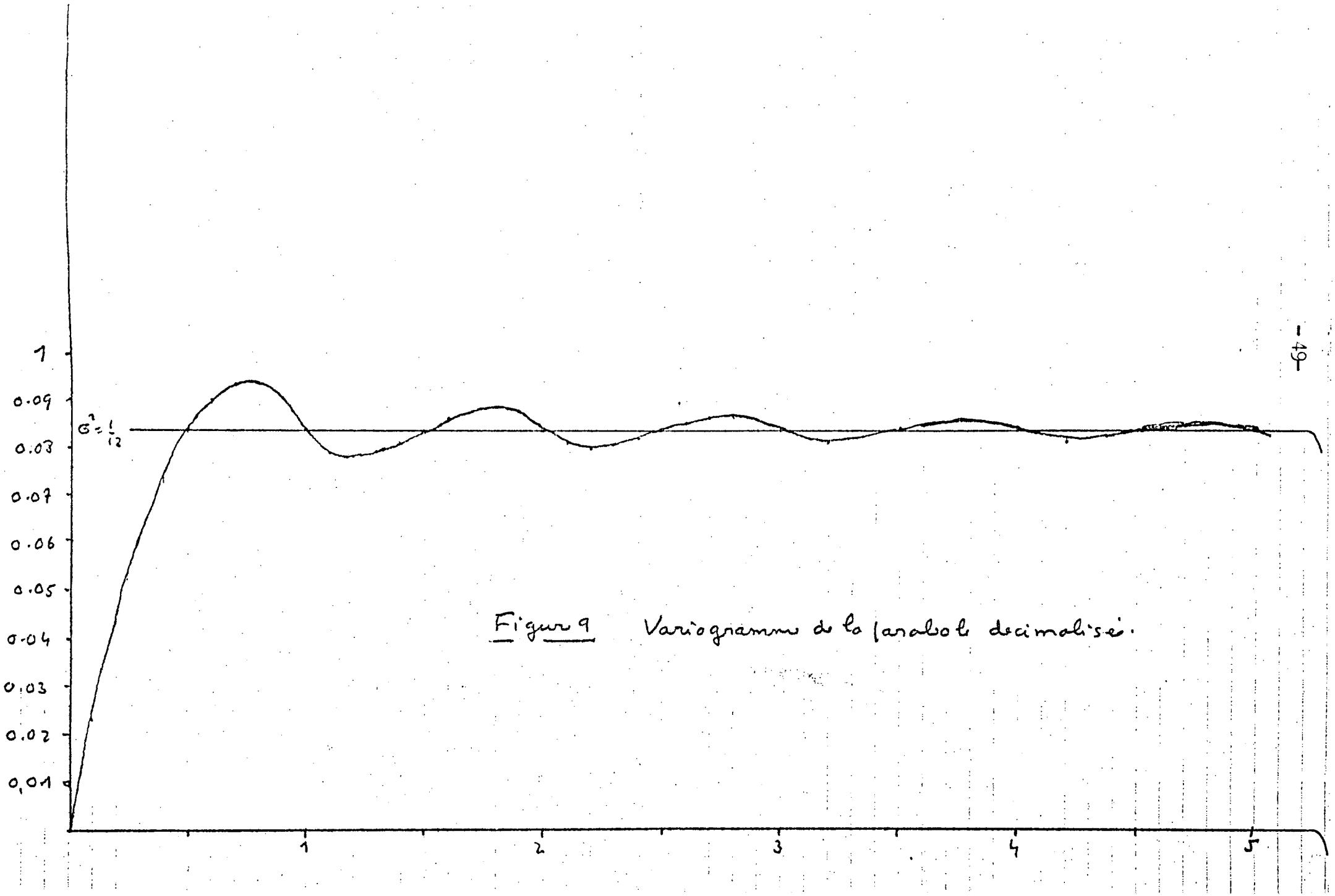


Figure 9 Variogramme de la parabole décimalisée.

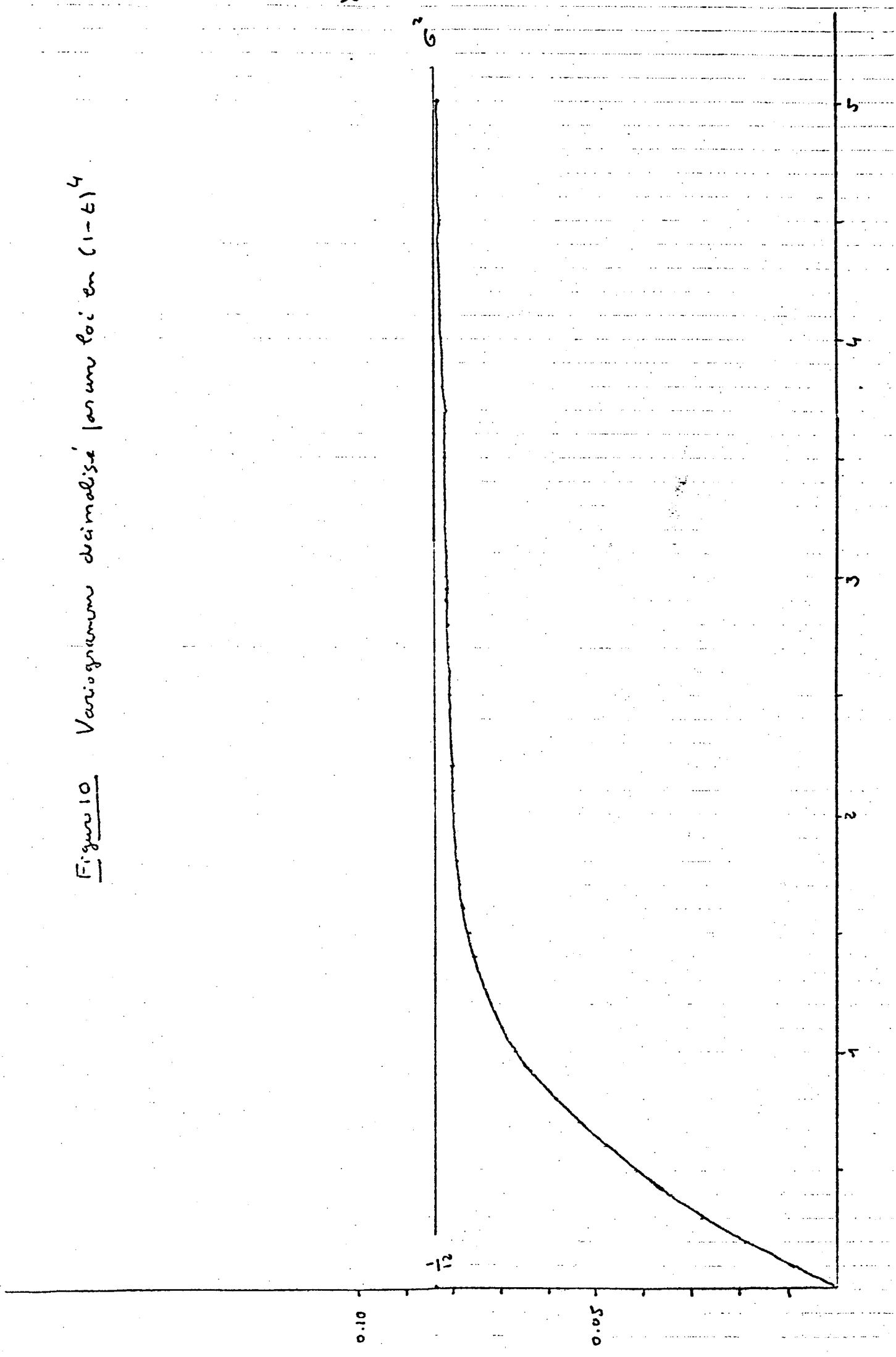


Figure 10 Variogramme dénormalisé pour $\nu = 4$

COMPLEMENTS

C - COMPLEMENTS SUR LES FONCTIONS DE COVARIANCE

J'indique ici quelques formules exactes ou approchées relatives aux fonctions de covariance des processus :

$$X_n = 2^n \text{ modulo } m \quad ; \quad \varepsilon_n = \text{Int } \left(\frac{2X_n}{m}\right)$$

La formule théorique de la covariance du processus ε_n a été établie dans le paragraphe 2. La tabulation numérique (dans le cas m = 301 979) met en évidence une relation très simple, qui reste valable avec une excellente approximation jusqu'à des distances k de l'ordre de quelques milliers (voir tableau 1 ci-après) : le coefficient de corrélation ρ_k de ε_{n+k} est, à très peu de choses près :

(C-1)
$$\rho_k \simeq \begin{cases} 0 & \text{si } R_k \text{ est pair} \\ 1/R_k & \text{si } R_k \text{ est impair} \end{cases}$$
$$(R_k = 2^k \text{ modulo } m)$$

Cette relation , toutefois, ne peut pas subsister pour des valeurs de R de l'ordre de m ou de m/2, ne serait-ce qu'à cause des relations :

$$\begin{cases} \rho_{m-1-k} = \rho_k \\ \rho_{\frac{m-1}{2} - k} = \rho_{\frac{m-1}{2} + k} = -\rho_k \end{cases}$$

La première de ces deux relations signifie que ρ_k reste invariant si l'on remplace k par -k modulo m-1 (et résulte de $2^{m-1} = 1$ modulo m). Or, remplacer k par -k modulo (m-1) revient à remplacer R_k par son inverse R' modulo m. Autrement dit : si deux entiers k et k' vérifient $R_k R_{k'} = 1$ modulo m, alors $\rho_k = \rho_{k'}$:

$$R_k R_{k'} = 1 \text{ modulo } m \Rightarrow \rho_k = \rho_{k'}$$

La seconde relation découle de $2^{\frac{m-1}{2}} = -1$ modulo m. Posons, pour abréger :

$$q = \frac{m-1}{2}$$

de sorte que $X_{n+q} = m - X_n$, $\varepsilon_{n+q} = 1 - \varepsilon_n$, et, de la même manière :

$$R_{k+q} = m - R_k$$

Ainsi, changer R_k en $m - R_k$ conduit à remplacer ρ_k par $-\rho_k$:

$$R_k + R_{k'} = m \Rightarrow \rho_k + \rho_{k'} = 0$$

En particulier, pour $k = 0$, on a $\rho_o = 1$, $R_o = 1$, et donc, avec $k' = q$, on trouve

$$\rho_q = -1 \quad , \quad R_q = m - 1$$

relation évidemment incompatible avec $\rho_q = 1/R_q$ (pour R_k voisin de m, la règle sera donc

$$\rho_k = \begin{cases} 0 & \text{si } R_k \text{ impair} \\ -\dfrac{1}{m - R_k} & \text{si } R_k \text{ pair} \end{cases}).$$

Explication de la règle C-1.

On peut, sinon démontrer, du moins comprendre l'origine de cette règle extraordinairement simple, à l'aide du raisonnement suivant :

Soit a un entier modulo m, et k l'entier modulo m-1 tel que :

$$a = 2^k \text{ modulo } m$$

A côté du processus X_n, nous pouvons considérer le processus :

$$Y_n = a^n \text{ modulo } m$$

de sorte que l'on a $Y_n = X_{n\,k}$. Dans l'écriture en base a du nombre Y_n/m

$$\frac{Y_n}{m} = \sum_{p=1}^{\infty} \frac{\eta_{n+k}}{a^p}$$

TABLEAU 1

Corrélogramme de ε_n

R_k	ρ_k	$\rho_k \times R_k$
1	1	1
2	-0.000007	
3	.333329	0.99999
4	.000007	
5	.199995	0.99997
6	-0.000007	
7	.142851	0.99996
8	-0.000007	
9	.111114	1.00003
10	- .000007	
11	.090901	0.99991
12	.000033	
13	.076926	1.00004
14	- .000007	
15	.066660	0.99991
16	- .000020	
17	.058845	1.00037
18	- .000007	
19	.052620	0.99977
20	.000060	
21	.047613	0.99987
22	.000020	
23	.043440	0.99913
24	- .000007	
25	.040010	1.00024
26	.000007	
27	.037042	1.00015
28	.000086	
29	.034486	1.00009
30	- .000007	
31	.032254	0.99926
32	.000020	
54	- .000033	
55	.018154	0.99845
63	.015875	1.00015
64	.000007	
127	.000786	0.99841
128	- .000033	
255	- .003927	1.00150
256	.000007	
511	.001954	0.99838
512	- .000060	
1023	.001027	1.05018
1024	.000033	

les η_n successifs constituant un aléa : on peut les assimiler à des VA indépendantes uniformément distribuées sur les entiers 0,1, 2,...,a-1.

Considérons maintenant la quantité

$$\varepsilon_{n\,k} = \text{Int}\left(\frac{2\,X_{n\,k}}{m}\right) = \text{Int}\left(2\,\frac{Y_n}{m}\right)$$

en notant que $2\,Y_n/m$ se met sous la forme :

$$\text{(C-2)} \qquad \frac{2\,Y_n}{m} = \frac{2\,\eta_1}{a} + \frac{2}{a}\,\frac{Y_{n+1}}{m}$$

Deux cas sont à distinguer selon que a est pair ou impair.

Si <u>a est pair</u>, les valeurs possibles de $2\,\eta_1/a$ sont

$$0\ ,\ \frac{2}{a}\ ,\ \frac{4}{a}\ ,\ldots,\ 1-\frac{2}{a}\ ,\ 1\ ,\ 1+\frac{2}{a},\ldots$$

Comme $\frac{2}{a}\,\frac{Y_{n+1}}{m}$ est $< \frac{2}{a}$, on voit que, dans ce cas, la partie entière de $\frac{2}{a}\,\frac{Y_n}{m}$ ne dépend que du premier chiffre η_1 et non des suivants :

$$\varepsilon_{n\,k} = \text{Int}\,\frac{2\,\eta_1}{m}$$

Par suite, pour <u>a pair</u>, $\varepsilon_{n\,k}$ est indépendant de Y_{n+1}, et donc aussi indépendant de $\varepsilon_{(n+1)\,k} = \text{Int}\,(2\,Y_{n+1}\,/\,m)$. Cela explique la règle empirique $\rho_k = 0$ si $R_k = a$ est pair.

Si <u>a est impair</u>, la situation est
possibles de $2\,\eta_1\,/\,a$ sont, en effet, cette fois :

$$0\ ,\ \frac{2}{a}\ ,\ \frac{4}{a}\ ,\ldots,\ 1-\frac{3}{a}$$

Cette fois, dans le cas médian ($\eta_1 = \frac{a-1}{2}$, i.e. $\frac{2\,\eta_1}{a} = 1 - \frac{1}{a}$), la valeur additionnelle $2\,Y_{n+1}\,/\,a_n$, comprise entre 0 et

2/a est susceptible de modifier la partie entière de $2\,Y_n/m$. En supposant les η_n indépendants, le décompte des configurations possibles donne alors

$$E\,(\varepsilon_{n\,k}\,\varepsilon_{(n+1)\,k}) = \frac{a-1}{4\,a} + \frac{1}{a}\left(\frac{a-1}{2\,a} + \frac{1}{2\,a}\right)$$

$$= \frac{a+1}{4\,a}$$

Cette covariance non centrée correspond à un coefficient de corrélation $\rho = 1/a$: Mais $a = R_k$ et $\rho = \rho_k$. D'où la règle empirique C-1.

Covariance du processus X_n.

Passons maintenant à la covariance du proessus X_n, ou plutot de :

$$X'_n = \frac{X_n}{m} = \text{Déc}\left(\frac{2^n}{m}\right)$$

Lorsque n varie de 0 à m - 2, X_n prend exactement une fois chacune des valeurs $p = 1,2,\ldots, m-1$. On en déduit sans peine la moyenne et la variance de X'_n :

$$E\,X'_n = \frac{1}{2} \quad ; \quad \sigma^2 = \frac{1}{12} - \frac{1}{6\,m}$$

Désignons par

$$C_k = E(X'_n\,X'_{n+k})$$

la covariance (non centrée) et par R_k l'entier 2^k modulo m, de sorte que :

$$C_k = \frac{1}{m^2(m-1)} \sum_{n=0}^{m-2} (2^n)_m\,(R_k\,2^n)_m$$

(la notation $(a)_m$ désigne l'entier a modulo m. Notons que si

$(2^n)_m = p$, on a aussi $(R_k\ 2^n)_m = (p\ R)_m$. Comme 2^n prend exactement une fois chacune des valeurs $p = 1, 2, \ldots, m-1$, on a donc également

$$(C\text{-}2) \qquad C_k = \frac{1}{m^2(m-1)} \sum_{p=1}^{m-1} p\ (p\ R_k)_m$$

En introduisant la variable $N = \text{Int}\ (\frac{p\ R_k}{m})$, qui varie de $N = 0$ à $N = \text{Int}\ \frac{(m-1)\ R}{m} = R - 1$, la formule (C-2) peut encore s'écrire :

$$C_k = \frac{1}{m^2\ (m-1)} \left[\sum_{N=0}^{R-2} \sum_{p=1\ +\ \text{Int}\ \frac{N\,m}{R}}^{\text{Int}\ \frac{(N+1)m}{R}} p(p\ R - N\ m) + \sum_{1\ +\ \text{Int}\ \frac{(R-1)m}{R}}^{R-1} p(p\ R - (R-1)\ m) \right]$$

Dans la sommation de l'expression $p\ (p\ R - N\ m) = p^2 R - N\ m\ p$, le terme $p^2\ R$, qui ne dépend pas de N, est sommé en p de $p = 1$ à $R - 1$, ce qui donne :

$$\frac{R}{m^2(m-1)} \sum_{p=1}^{R-1} p^2 = R\ (\frac{1}{3} - \frac{1}{6\ m})$$

Posant $K_n = \text{Int}\ (\frac{n\ m}{R})$, la sommation du terme suivant, qui est $-\ N\ m\ p$, va donner :

$$\frac{-\ 1}{2\ m^2(m-1)} \left[\sum_{n=0}^{R-2} n\ m\ (K_{n+1} + K_{n+1}^2 - K_n - K_n^2) + \right.$$

$$\left. + (R-1)\ m^2\ (m-1) - m\ (R-1)\ K_{R-1}\ (1 + K_{R-1}) \right]$$

$$= -\ \frac{1}{2\ m(m-1)} \left[m(m-1)(R-1) - \sum_{n=1}^{R-1} K_n\ (1+K_n) \right]$$

et au total, on obtient la formule intéressante :

TABLEAU 2

Corrélogramme de X_n ($\rho_k \simeq 1/R_k$)

k	R_k	ρ_k	$\rho_k \times R_k$
0.	1	1	1
1	2	0.499995	0.99999
51 790	3	.333329	0.99997
2	4	.249996	0.99999
123 942	5	.200000	1.00000
51 791	6	.166669	1.00007
250 332	7	.142863	1.00004
3	8	.124989	0.99991
103 580	9	.111096	0.99987
123 943	10	.100015	1.00015
178 246	11	.090905	0.99996
51 792	12	.083355	1.00025
67 105	13	.076902	0.99972
250 333	14	.071456	1.00039
175 732	15	.066698	1.00046
4	16	.062485	0.99978
244 876	17	.058833	1.00016
103 581	18	.055543	0.99978
181 100	19	.052625	0.99988
123 944	20	.050047	1.00094
144	21	.047669	1.00106
178 247	22	.045458	1.00007
24 163	23	.043440	0.99913
51 793	24	.041667	1.00002
247 884	25	.039984	0.99960
67 106	26	.038446	0.99959
155 370	27	.037019	0.99951
250 334	28	.035788	1.00206
293 535	29	.034435	0.99861
175 733	30	.033413	1.00309
75 135	31	.032227	0.99905
5	32	.031252	1.00007
155 371	54	.018475	0.99763
210	55	.018149	0.99821
51 934	63	.015877	0.99896
6	64	.015616	0.99943
92 893	127	.007870	0.99952
7	128	.007778	0.99562
118 630	255	.003886	0.99093
8	256	.003879	0.99308
266 583	511	.003009	1.08644
9	512	.001910	0.97779
3 193	1 023	.001013	1.03671
10	1 024	.000985	1.00822

(C-3)
$$C_k = R\left(\frac{1}{3} - \frac{1}{6m}\right) - \frac{R-1}{2} + \frac{1}{2m(m-1)} \sum_{n=1}^{R-1} K_n(1+K_n)$$

$$\left(K_n = \text{Int}\ \frac{n\,m}{R};\ R = R_k = 2^k \text{ modulo } m\right)$$

En première approximation, et si R n'est pas trop grand vis à vis de m, on peut prendre $K_n \simeq \frac{n\,m}{R}$, et il reste

$$C_k \simeq -\frac{R}{6} - \frac{R}{6m} + \frac{1}{2} + \frac{R-1}{4(m-1)} + \frac{m}{m-1}\ \frac{(R-1)(2R-1)}{12\,R}$$

Comme m = 301 979 est assez grand, on peut négliger les termes en 1/m, d'où :

$$C_k \simeq \frac{1}{4} + \frac{1}{12\,R}$$

D'autre part, la moyenne étant 1/2, la covariance centrée est $C_4 - 1/4 \simeq 1/(12\,R)$. Comme enfin la variance est $\sigma^2 = 1/12 - 1/(6m)$, il vient à la même approximation

(C-4)
$$\rho_k \simeq \frac{1}{R_k} \qquad (R_k = 2^k \text{ modulo } m)$$

Explication de la règle C-4.

Cette formule donne une valeur remarquablement approchée, au moins pour R_k petit (au plus égal à quelques milliers) comme on peut le voir sur la table 2 ci-jointe.

Ce résultat se comprend assez bien : comme m est grand, le processus X'_n (arithmétique), doit être relativement proche du processus général $X_n = 2^n x_o$ ($x_o \notin \mathcal{N}$), dont la covariance est $\tilde{\rho}_k = 2^{-k}$. On s'attend donc à avoir $\rho_k \simeq 2^{-k}$ pour k petit. De fait, $R_k = 2^k$ pour $k \leq 18$. Pour $k > 18$, la valeur 2^{-k} du processus continu est remplacé par la valeur - très différente - $1/R_k$.

De fait, prenons un entier a différent de 2 et considérons le processus $Y_n(a) = a^n$ modulo m : pour la même raison que ci-dessus, on s'attend, pour k petit, à ce que le coefficient de corrélation $\rho_k(a)$ de ce processus $Y_n(a)$ soit $\simeq a^{-k}$.

Or, l'entier a admet la représentation

$$a = 2^{n_a} \quad \text{modulo m}$$

pour un entier n_a convenable, et il en résulte : $Y_n(a) = X_{nn_a}$ et

$$\rho_k(a) = \rho_{kn_a} \simeq \frac{1}{a^k}$$

En particulier, pour k = 1, $\rho_{n_a} \simeq 1/a$. Mais justement, a est 2^{n_a} modulo m, c'est-à-dire R_{n_a}. Il faut donc bien que l'on ait $\rho_n = 1/R_n$, au moins pour n petit.

Il est tout-à-fait instructif de constater sur la valeur numérique que ρ_k est bien à peu près égal à $1/R_k$: les mêmes valeurs numériques, ordonnées en k, auraient donné l'impression de fluctuations imprévisibles. Ordonnées en R_k, au contraire, elles mettent en évidence une relation fonctionnelle très simple.

Limitations de la règle C-4.

Cette relation remarquable $\rho_k = 1/R_k$ ne peut pas, cependant, rester valable pour toutes les valeurs de k. De fait, comme dans le cas du processus ε_n, on a les relations :

$$\begin{cases} \rho_{m-1-k} = \rho_k \\ \rho_{q-k} = \rho_{q+k} = -\rho_k \end{cases} \qquad \left(q = \frac{m-1}{2}\right)$$

d'où résulte ici encore :

$$R_k\, R_{k'} = 1 \text{ modulo m} \Rightarrow \rho_k = \rho_{k'}$$

$$R_k + R_{k'} = m \Rightarrow \rho_k + \rho_{k'} = 0$$

TABLEAU 3

Correspondance $R_k = 2^k$ modulo m → k

R_k	k	R_k	k
3	51 790	149	90 817
5	123 942	151	232 270
7	250 332	157	50 204
11	178 246	163	26 830
13	67 105	167	211 706
17	244 876	173	287 495
19	181 100	179	288 466
23	24 163	181	135 325
29	293 535	191	81 169
31	75 135	193	276 819
37	173 415	197	161 852
41	70 919	199	260 428
43	299 778	211	301 004
47	253 155	223	41 083
53	177 694	227	275 523
59	141 061	229	196 129
61	216 791	233	122 071
67	10 039	239	266 355
71	32 393	241	267 938
73	16 251	251	103 330
79	206 304	257	253 908
83	165 541	263	131 924
89	290 726	269	252 549
97	117 206	271	263 879
101	265 013	277	19 726
103	212 650	281	108 089
107	290 433	283	261 096
109	206 006	293	220 098
113	239 057		
127	92 893		
131	246 478		
137	53 465		
139	298 243		

m = 59

k	R_k	ρ_k de ε_n	ρ_k de X_n
0	1	1	1
1	2	-0.0345	0.4737
50	3	.3103	.3103
2	4	.0345	.2377
6	5	.1724	.2014
51	6	-.0345	.1833
18	7	.1724	.1143
3	8	-.0345	.0671
42	9	.0345	.0345
7	10	-.0345	.1833
25	11	.1034	.0163
52	12	.1724	.2014
45	13	-.0345	-.0345
19	14	-.0345	-.0200
56	15	.0345	.2377
4	16	-.1034	-.0163
40	17	.1724	.1143
43	18	-.0345	-.0054
38	19	.0345	-.1107
8	20	.3103	.3103
10	21	.0345	.0200
26	22	.0345	-.0671
15	23	-.0345	-.0054
53	24	-.0345	.0526
12	25	.0345	-.0381
46	26	.0345	-.0381
34	27	.0345	-.0526
20	28	.0345	-.1107
28	29	.0345	-.4737
57	30	-.0345	+.4737

En particulier, $\rho_q = -1$ avec $R_q = m - 1$. Si R_k est voisin de m, la règle C-4 doit donc être remplacée par :

$$\rho_k = -\frac{1}{m - R_k}$$

Lorsque R_k est voisin de $q = \frac{m-1}{2}$, ces règles ne peuvent plus s'appliquer. De fait, $R_k = q$, i.e. $2 R_k = 2q = m - 1$ correspond à $2^{k+1} = -1$ modulo m, c'est-à-dire à $k = q-1$. Or $\rho_{q-1} = -\rho_1$ est voisin de 1/2 :

$$R_{q-1} = q \quad , \quad \rho_{q-1} \simeq -\frac{1}{2}$$

valeur incompatible avec les règles $1/R_k$ et $-1/(m-R_k)$.

A titre d'exemple, le tableau ci-après (que l'on doit compléter selon la règle $\rho_{k'} = -\rho_k$ pour $R_{k'} = m - R_k$; R_k, variant de 30 à 58) donne les valeurs numériques de ρ_k pour les deux processus ε_n et X_n dans le cas m = 59 : on voit que les règles C-1 et C-4 s'appliquent assez bien pour $R_k \leq 7$. J'ignore si on peut en conclure, plus généralement, qu'elles restent valables jusque vers m/10 (soit k = 30 000 dans le cas m = 301 979). (Voir Tableau des valeurs numériques ci-contre).

Généralisation de la règle C-4.

Entre les processus ε_n et X_n, il existe une relation simple que l'on peut écrire

$$\varepsilon_n = 2\frac{X_n}{m} - \frac{X_{n+1}}{m}$$

La covariance $\tilde{\sigma}_k$ de ε_n et ε_{n+k} se déduit donc de la covariance σ_k de X_n et X_{n+k} selon la relation

$$\tilde{\sigma}_k = \frac{1}{m^2}(5\,\sigma_k - 2\,\sigma_{k-1} - 2\,\sigma_{k+1})$$

ou encore, $\tilde{\rho}_k$ et ρ_k désignant les corrélogrammes des processus ε_n et X_n :

(C-5) $$\tilde{\rho}_k = \frac{1}{3}\,(5\,\rho_k - 2\,\rho_{k-1} - 2\,\rho_{k+1})$$

Supposons $R_k = 2^k$ modulo m assez petit devant m pour que les règles C-1 et C-4 s'appliquent à ρ_k et aussi à ρ_{k+1} (puisque $R_{k+1} = 2\,R_k$ est lui aussi petit devant m).

Alors, si R_k est pair, on a $\tilde{\rho}_k = 0$, d'après (C-1), et $\rho_k = 1/R_k$, $\rho_{k+1} = 1/2\,R_k$, d'après (C-4) : la relation (C-5) ci-dessus se réduit à $\rho_{k-1} = 2/R_k = 2\,\rho_k$. Celà était prévisible puisque, R_k étant pair, on a $R_{k-1} = R_k/2$, et que ce nombre est petit devant m.

Mais si R_k est impair, on obtient un résultat moins trivial. Cette fois, en effet :

$$\tilde{\rho}_k = \rho_k = \frac{1}{R_k} \quad ; \quad \rho_{k+1} = \frac{1}{2\,R_k}$$

et (C-5) donne

$$\rho_{k-1} = \frac{1}{2\,R_k} = \frac{1}{2}\,\rho_k = \rho_{k+1}$$

La valeur correspondante de R_{k-1} est $(R_k + m)/2$ puisque R_k est impair. D'où la règle :

(C-6) Pour R_k voisin de $\frac{m}{2}$, $\rho_k = \dfrac{1}{4(R_k - \frac{m}{2})}$

Plus généralement, on peut conjecturer la règle suivante :

(C-7) S'il existe deux entiers b et β <u>petits devant m</u> et <u>premiers</u> entre eux tels que :

$$b\,R_k = \beta \quad \text{modulo } m$$

alors

$$\rho_k = \frac{1}{b\,\beta} \qquad (b > 0,\ \beta \text{ de signe quelconque})$$

Pour β = 1, cette règle est certainement vraie : en effet, b et R_k sont inverses modulo m. Donc, si $b = 2^{k'}$ modulo $m = R_{k'}$, on a

$\rho_k = \rho_{k'}$. Mais $\rho_{k'} = 1/R_{k'} = 1/b$, puisque b est petit. Donc $\rho_k = 1/b$.

Si b = 2 et β impair, $b\ R_k = \beta$ entraine $R_k = (m+\beta)/2$ et donc $\rho_k = 1/2\beta$ d'après la règle (C-6).

Dans le cas général, la règle (C-7) doit pouvoir s'établir en raisonnant sur les processus $Y_n(b) = (b)^n$ modulo m et $Y_n(\beta) = \beta^n$ modulo m.

Selon cette règle, les seules valeurs de k pour lesquelles ρ_k prendrait des valeurs significativement différentes de 0 seraient celles pour lesquelles on peut écrire :

$$R_k = \frac{a\ m + \beta}{b}$$

avec des entiers a, b, β tels que :

β et b petits devant m ($\beta > 0$ ou non , $b > 0$)

β et b premiers

et on aurait alors $\rho_k = 1/b\beta$.

Justification de la règle (C-7)

Supposons qu'il existe deux entiers b et β petits et premiers entre eux tels que

(i) $$b\ R_k = \beta \quad \text{modulo m}$$

Alors, en négligeant des termes du premier ordre en 1/m, on a :

(ii) $$\boxed{\rho_k \simeq \frac{1}{b\ \beta}}$$

Pour le voir, cherchons à évaluer la covariance (non centrée)

$$C_k = \frac{1}{m^2(m-1)} \sum_{p=1}^{m-1} p(R\ p)_m$$

Notons d'abord ceci : si b' est l'inverse de b modulo m, on a $R = b'\beta$. Alors :

$$C_k = \frac{1}{m^2(m-1)} \sum_{p=1}^{m-1} p(\beta\ b'p)_m$$

Lorsque p varie de 1 à m-1, $(b'p)_m$ prend une fois chacune des valeurs 1,2,...,m-1. Donc, en posant $p' = (b'p)_m$, soit $p = (p'b)_m$, on trouve

(iii) $$C_k = \frac{1}{m^2(m-1)} \sum_{p=1}^{m-1} (bp)_m\ (\beta p)_m$$

Autrement dit

$$C_k = < Z_b, Z_\beta >$$

avec les VA Z_b et Z_β définies par

$$Z_b = \frac{(bp)_m}{m} \quad ; \quad Z_\beta = \frac{(\beta p)_m}{m}$$

(p uniforme sur 1,2,...,m-1). Comme b et β sont petits vis à vis de m, on peut remplacer ces VA par

$$Z'_b = \text{Déc}\ (bx) \quad ; \quad Z'_\beta = \text{Déc}\ (\beta x)$$

(x uniforme sur l'intervalle (0,1)). La covariance C_k définie en (iii) est alors approchée (avec une erreur de l'ordre de 1/m) par:

(iv) $$C'_k = \int_0^1 \text{Déc}\ (bx)\ \text{Déc}\ (\beta x)\ dx$$

Pour calculer l'expression (iv), partons du développement

$$\text{Déc}\ x = \frac{1}{2} - \frac{1}{2i\pi} \sum_{k \neq 0} \frac{e^{-2i\pi kx}}{k}$$

D'où

$$Z'_b(x) = \frac{1}{2} - \frac{1}{2i\pi} \sum_{k \neq 0} \frac{e^{-2i\pi kbx}}{k}$$

$$Z'_\beta(x) = \frac{1}{2} - \frac{1}{2i\pi} \sum_{k' \neq 0} \frac{e^{-2i\pi k'\beta x}}{k'}$$

La covariance $C'_k = \langle Z'_b Z'_\beta \rangle$ est donc :

$$C'_k = \frac{1}{4} + \frac{1}{4\pi^2} \sum \frac{1}{k\,k'}$$

expression où la sommation porte sur les couples (k,k') tels que $k\,b = k'\,\beta$. Comme b et β sont premiers entre eux, ces couples sont de la forme $k = \beta p$, $k' = bp$. Par suite :

$$C'_k = \frac{1}{4} + \frac{1}{b\,\beta} \sum_{p=1}^{\infty} \frac{1}{4\pi^2 p^2} = \frac{1}{4} + \frac{1}{12\,b\beta}$$

Au premier ordre en 1/m, on a $C_k = C'_k$, $E(X_n) = 1/2$ et Var X_n = 1/12 (en réalité : $\frac{m-2}{12\,m}$).

Donc, à la même approximation, on a bien

$$\rho_k \simeq \frac{1}{b\,\beta}$$

et la règle (C-7) se trouve justifiée.

D - LE PROCESSUS $X_{n+1} = (X_n)^2$ modulo m.

Au lieu de la relation de récurrence $X_{n+1} = 2\,X_n$ (m), nous allons maintenant considérer le processus obtenu en prenant les carrés modulo m :

(D-1) $$X_{n+1} = (X_n)^2 \text{ modulo } m$$

Le processus continu correspondant, avec des variables Y comprises entre 0 et 1, serait :

$$Y_{n+1} = \text{Déc } (m\ Y_n^2)$$

Ses propriétés ne sont pas très bonnes : pourvu que la valeur initiale x_o n'appartienne pas à un certain ensemble négligeable pour la mesure de Lebesgue sur (0,1), Y_n converge vers 0 lorsque n tend vers l'infini.

Mais dans le cas arithmétique, qui correspond à la relation (D-1), on a toujours $x_o \in \mathcal{K}$. Si le nombre m possède les bonnes propriétés habituelles, la suite X_n définie en (D-1) est cyclique et et doit pouvoir servir d'aléa.

Dans ce qui suit, je suppose :

(D-2) m premier, $q = \frac{m-1}{2}$ premier, $2^q = -1$ modulo m

Dans le cas m = 301 979 qui nous intéresse plus particulièrement, on a :

$q = \frac{m-1}{2} = 150\ 989$; $\frac{q-1}{2} = 75\ 944$ n'est pas premier, mais :

$$\frac{q-1}{4} = 37\ 747 \text{ est premier}$$

De plus, à la relation $2^q = -1$ modulo $\underline{m}$ s'ajoute la relation

$$2^{\frac{q-1}{2}} = -1 \text{ modulo } \underline{q} \quad \text{(et non modulo m)}$$

qui nous servira dans un instant.

Avec un nombre m vérifiant les conditions (D-2), tout entier k (non nul) modulo m admet une représentation unique

$$k = 2^p \text{ modulo m}$$

où p est un entier <u>modulo m-1</u> = 2q. Comme m-1 est pair, la parité est une propriété qui passe au quotient modulo m-1. On voit ainsi que k est un carré parfait modulo m (i.e. $k = x^2$ modulo m pour un entier x) si et seulement si p est un entier <u>pair</u>. Si p = 2p' est pair, k admet les deux racines

$$x_1 = 2^{p'} \qquad x_2 = 2^{p'+q} \qquad (q = \frac{m-1}{2})$$

($x_2 = -x_1$ modulo m). De ces deux racines, une et une seule est un carré parfait. Ainsi; pourvu seulement que la valeur initiale x_o soit déjà elle-même un carré modulo m, on peut prolonger le processus vers la gauche, en prenant pour X_{-n-1} celle des deux racines de X_{-n} qui est un carré parfait : il suit de là que la suite des X_n est <u>cyclique</u>.

La question qui se pose est celle de la période de cette suite : est-elle égale à la valeur maximale $\frac{m-3}{2} = q - 1$, autrement dit le processus X_n, n = 0,1,...,q-1 prend-il une fois et une fois seulement chacune des valeurs 2^{2p} (p = 1,2,...,q-2) ?

Ou encore, en remarquant que $4 = 2^2$ est un carré, on peut se demander si la suite modulo m

$X_n = 2^{(2^n)}$, n = 1,2,...,q-2 prend une fois et une fois seulement chacune des valeurs 2^{2p}, p = 1,2,...,q-2, autrement dit :

la suite 2^n modulo m-1 (n = 1,2,...) est-elle cyclique avec la période q - 1 ?

Posons donc

$$Z_n = 2^n \text{ modulo } m-1 \ , \quad n = 1,2...$$

On aura $Z_{n+k} = Z_n$ si n et k vérifient la condition

$$2^n(2^k - 1) = N(m - 1) = 2\ N\ q$$

Comme q est premier et différent de 2, il ne divise pas 2^n et doit donc diviser $2^k - 1$. D'autre part, 2^n est pair (puisque nous avons exclu la valeur n = 0), de sorte que la condition ci-dessus équivaut à

(D-3) $$2^k - 1 = 0 \quad \text{modulo } q$$

Ainsi, la suite Z_n est cyclique et sa période est le plus petit entier k vérifiant (D-3). Les candidats possibles sont q-1 et ses diviseurs. On a bien $2^{q-1} = 1$ modulo q. Avec m = 301 979, les facteurs de q-1 sont 2, 4 (qui ne vérifient pas D-3) et $\frac{q-1}{2}$, $\frac{q-1}{4}$: comme $2^{\frac{q-1}{2}} = -1$ modulo q, ces deux derniers facteurs ne vérifient pas non plus D-4, de sorte que la période k est bien égale à q - 1. Ainsi :

n variant de n = 0 à n = q -2, la suite 2^n modulo q prend une fois et une seule chacune des valeurs 1,2,...,(q-1), et donc la suite

$$Z_n = 2(2^n \bmod q) = 2^{n+1} \bmod 2q \quad (n = 0,1,\ldots,q-2)$$

prend une fois et une seule chacune des valeurs paires (0 exclu), 2, 4,...,2(q-1) = m-3 modulo m-1. En particulier, à une translation près, la suite obtenue par exponentiation, soit $2^{Z_n} \bmod m$ s'identifie à la suite X_n, n = 1,2,... définie en D-2 (à une translation près), qui est donc bien cyclique et admet la période maximale q-1.

Cette suite peut-elle servir d'aléa ? De prime abord, vu la manière dont elle est fabriquée, il serait surprenant que sa covariance prenne des valeurs significatives. D'autre part, les X_n prenant exactement une fois chacune des valeurs correspondant à un carré parfait, la loi de distribution n'est pas nécessairement uniforme sur ces q-1 entiers, mais ne doit pas différer beaucoup de la loi uniforme. Je n'ai fait que les tests sommaires suivants (sur la suite $X'_n = X_n/m$):

Pour N = 400 valeurs successives :

moyenne 0.4783 , variance σ^2 = 0.07623

(pour une loi uniforme : σ^2 = 1/12 = 0.08333)

et les 30 premiers points du variogramme :

0.08111	0.08072	0.08445
.08048	.08145	.07954
.07285	.07605	.07973
.08928	.08287	.07591
.07957	.07530	.08725
.08247	.08412	.08346
.07938	.08110	.08023
.08263	.08615	.07741
.08432	.07695	.07530
.08399	.07906	.07906

Pour N = 1 000 valeurs successives (à partir d'une autre valeur initiale, soit x_o = 1 024 au lieu de 232 348)

moyenne = 0.48944 variance σ^2 = 0.08118

Histogramme

classe	fréquence	classe	fréquence
0 à 0.1	.106	0.5 à 0.6	.102
0.1 à 0.2	.098	0.6 à 0.7	.118
0.2 à 0.3	.106	0.7 à 0.8	.087
0.3 à 0.4	.097	0.8 à 0.9	.096
0.4 à 0.5	.102	0.9 à 1	.088

Ces résultats semblent encourageants.