LES METHODES DE MONTE-CARLO

1. - INTRODUCTION

On peut définir les méthodes de Monte-Carlo comme l'utilisation d'expériences sur des nombres aléatoires pour la résolution d'un problème.

Un premier type de problèmes pour lequel cette méthode est applicable est évidemment celui des problèmes probabilistes : on en fait une simulation directe si le problème est directement posé en termes de nombres aléatoires, ou une simulation indirecte comme pour les problèmes du deuxième type.

Ce deuxième type est celui des problèmes déterministes, qui n'ont donc à priori aucun rapport avec les phénomènes aléatoires. L'utilisation des techniques probabilistes pour leur résolution se fait alors de la façon suivante : l'étude théorique du problème posé conduit à une formulation dont on constate qu'elle décrit également un processus aléatoire apparemment sans relation ; la simulation directe du problème probabiliste correspondant, permet alors de résoudre le problème posé.

Il va de soi que les résultats obtenus par ces méthodes présentent une incertitude due à leur origine même ; mais il y a lieu à ce sujet de remarquer qu'on peut, par les techniques du calcul des probabilités, déterminer en fonction du nombre et de la nature des expériences faites, un intervalle de confiance pour la solution ; sans tomber dans un optimisme inconsidéré, on doit aussi signaler que l'utilisation des calculateurs électroniques permet de réaliser dans des temps raisonnables un grand nombre d'expériences, et d'atteindre ainsi une précision acceptable ; enfin pour bien des problèmes, les méthodes classiques de l'analyse numérique ne fournissent elles aussi que des résultats approchés et il y a lieu alors de comparer l'efficacité de ces méthodes et celle des méthodes de Monte-Carlo.

2. - FONDEMENTS DES METHODES DE MONTE-CARLO

Les méthodes de Monte-Carlo ont leur racine profonde, du point de vue probabilité dans la loi forte des grands nombres.

Rappelons que la loi des grands nombres exprime que : X_1 , X_2 , .. X_n désignant n variables aléatoires indépendantes (deux à deux) obéissant toutes à la même loi de probabilité de moyenne m et d'écart type σ , la variable aléatoire $Y_n = \frac{X_1 + X_2 + .. + X_n}{n}$, moyenne arithmétique de ces variables, tend en probabilité quand $n \to \infty$, vers la moyenne stochastique m, et que la différence a pour écart type $\frac{\sigma}{\sqrt{n}}$. Cela entraîne l'existence d'une sous-suite extraite de la suite Y_n qui converge presque sûrement vers m ; mais cette propriété ne serait pas suffisante pour justifier l'utilisation pratique des méthodes de Monte-Carlo. C'est la loi forte des grands nombres qui fournit cette justification.

Elle affirme que la convergence de la suite Y_n est presque sûre, c'est-à-dire qu'il y a une probabilité égale à 1 pour que $Y_n = \frac{X_1 + X_2 + .. + X_n}{n}$ converge au sens de l'analyse vers m.

L'inégalité de Bienaymé-Tchebichef permet une évaluation statistique de l'erreur commise lorsque l'on a effectué n expériences, mais la tendance vers la loi normale indiquée par le théorème central limite de Liapounoff permet d'atteindre un résultat plus précis dans ce domaine. Rappelons ce théorème : la loi de probabilité de la variable $Y_n = \frac{X_1 + X_2 + .. + X_n}{n}$ tend vers la loi normale de moyenne m et d'écart type $\frac{\sigma}{\sqrt{n}}$ (convergence au sens de Bernoulli).

Signalons aussi dès maintenant qu'une approche non probabiliste de ces méthodes a été faite (Bass et Guilloud, 1958), fondée sur la notion de suite uniformément dense. Nous y reviendrons dans la suite (§ 4.6).

3. - LES NOMBRES ALEATOIRES

3.1. - Généralités.

Le problème de l'aiguille de Buffon (1777) constitue

un exemple historique d'utilisation d'une méthode de Monte-Carlo (utilisation d'expériences probabilistes pour le calcul de la valeur de π), mais ce n'est pas ce genre de procédés qui vaut aux méthodes de Monte-Carlo leur succès actuel ; ainsi que nous l'avons dit au § 1 , les calculateurs électroniques permettent de réaliser dans des temps raisonnables le très grand nombre d'expériences nécessaires pour obtenir des résultats significatifs, mais ces expériences ne peuvent porter que sur des nombres.

On doit substituer à une variable aléatoire un ensemble de valeurs réelles ayant les propriétés statistiques de la variable aléatoire. Ces valeurs sont appelées nombres aléatoires parce qu'elles auraient pu être produites par le hasard. à l'aide d'un processus aléatoire adéquat.

Il est donc nécessaire de disposer de "sources" de nombres aléatoires satisfaisant à différentes fonctions de distribution. En fait, comme nous le verrons, ce sont les nombres dont la distribution est la distribution rectangulaire réduite -nombres tirés "au hasard" sur l'intervalle (0,1)- qui jouent un rôle fondamental ; il est en effet possible d'obtenir à partir de ceux-là des suites de nombres aléatoires obéissant à toute autre distribution.

Nous distinguerons :

Les nombres aléatoires, obtenus par des processus physiques qui sont aléatoires au sens strict.

Les nombres pseudo-aléatoires, calculés sur le calculateur utilisé suivant une règle bien spécifiée qui soit telle qu'aucun test statistique raisonnable ne puisse détecter d'écart significatif avec l'aléatoire. L'avantage de ces nombres est la reproductibilité de leur suite dans le cas où l'on veut effectuer des calculs de contrôle ; cette reproductibilité même montre bien qu'il ne s'agit pas de nobmres aléatoires au sens strict !

Enfin les nombres quasi-aléatoires, pour lesquels on sait qu'ils ne satisfont pas certains tests statistiques et possèdent uniquement les propriétés statistiques particulières qui intéressent la résolution du problème.

3.2. - Obtention de nombres aléatoires obéissant à la distribution rectangulaire réduite.

Nous dirons peu de chose de la fabrication des nombres

aléatoires au sens strict qui est d'usage peu courant dans la pratique.

Le nombre à engendrer est représenté dans l'ordinateur par n positions binaires dont les valeurs 0 ou 1 doivent être aléatoires avec probabilité égale à 1/2. (En fait on n'obtient ainsi, puisque n est fini, qu'une approximation de la distribution rectangulaire réduite par une distribution en escalier, à $1/2^n$ près, mais il ne s'agit là que des approximations usuelles en calcul numérique).

Deux types principaux de dispositifs physiques sont utilisés pour fixer les valeurs des positions binaires de façon aléatoire :

l'émission de particules radioactives,

le bruit de fond dans un circuit électronique.

(on trouvera des détails sur ce sujet dans le livre de Shreider, 1966). Les ordinateurs spécialisés dans l'utilisation des méthodes de Monte-Carlo peuvent comporter un de ces dispositifs, mais leur entretien est délicat et l'on doit en permanence s'assurer du caractère réellement aléatoire du phénomène utilisé. Cela explique leur emploi très limité.

Les tables de la Rand Corporation (1955) ont été obtenues par des procédés de ce type.

3.3. - Nombres pseudo-aléatoires pour la loi rectangulaire réduite.

On construit habituellement une suite de nombres pseudo-aléatoire par un processus de récurrence de la forme

$$x_n = f(x_{n-1}, x_{n-2}, \dots x_{n-k})$$

tel que la suite obtenue ait les propriétés statistiques voulues.

Une des premières méthodes pour engendrer une suite de nombres pseudo-aléatoires uniformément répartis sur (0,1) a été la méthode des carrés médians de Metropolis et Von Neumann (cf. Forsythe, 1951), qui s'est révelée peu satisfaisante.

Nous allons examiner une méthode de congruence couramment utilisée pour sa simplicité d'emploi sur les calculateurs électroniques, dont l'idée semble due à Lehmer (1951).

La relation de récurrence est la congruence,

(1) $x_i \equiv a x_{i-1}$ (modulo m), congruence multiplicative,

généralisée par Greenberger (1961) sous la forme

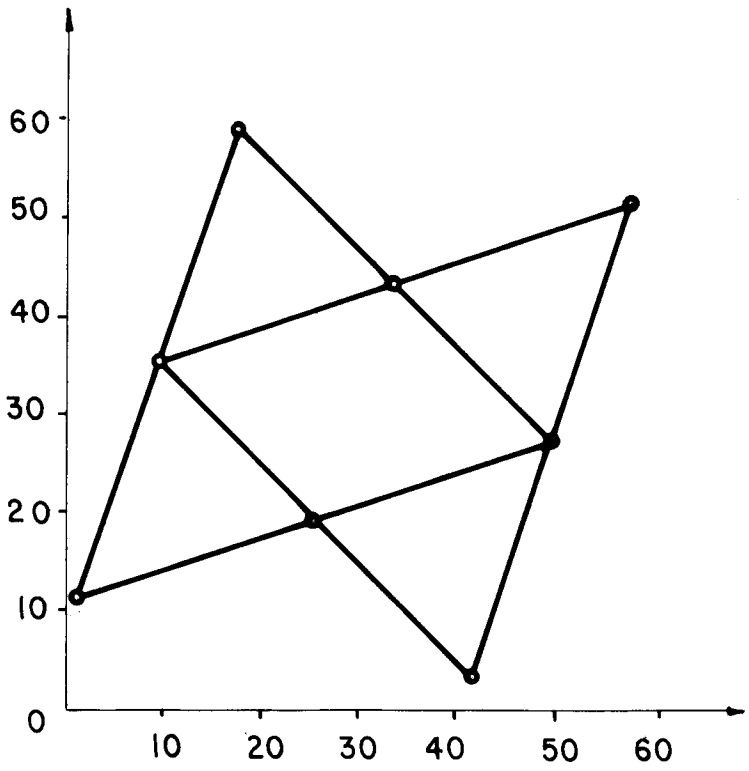

figure 1

(2) $x_i \equiv ax_{i-1} + c$ (modulo m).

Les nombres x_i/m sont utilisés comme suite pseudo-aléatoire uniformément répartie sur (0,1).

La commodité de la méthode tient au fait qu'en prenant pour m l'entier correspondant à la capacité de la machine - 2^p pour un ordinateur binaire à p digits - l'opération de congruence s'effectue automatiquement.

La récurrence (1) a une période maximale égale à m/4 (on obtient deux sur quatre des nombres impairs compris entre 0 et m -1) si x_0 est impair et a de la forme $a = 8k \pm 3$. La condition de période maximale assure une distribution uniforme des m/4 nombres entre 0 et (m - 1), mais ne renseigne en rien sur la corrélation entre les termes de la suite : les termes successifs de la suite sont évidemment corrélés puisqu'il existe entre eux une relation fonctionnelle, la relation de récurrence de définition, mais ce qu'il importe de savoir c'est dans quelles utilisations cette corrélation peut se révéler néfaste ; ce point est capital : par exemple pour le calcul d'intégrales afin que de bons résultats soient obtenus sans utiliser tous les termes de la suite, ou lorsque les nombres sont extraits de la suite par groupes pour former les coordonnées d'un vecteur pseudo-aléatoire. Marsaglia (1968) a récemment donné un exemple des fortes corrélations pouvant exister dans ce type de suite :

pour a = 11, m = 64, x_1 = 49, les termes de la suite (1) sont :
49, 27, 41, 3, 33, 43, 25, 19, 17, 59, 9, 35, 1, 11, 57, 51,
qui , si on les groupe par deux pour former les coordonnées de huit points pseudo-aléatoires dans le carré 64 x 64, donnent la disposition indiquée à la figure 1 !

Le calcul direct du coefficient de corrélation entre termes successifs pour cette suite nous a donné : $\rho = \frac{33}{337} \# 0,1$.

Nous donnons à titre d'exemple de l'utilisation courante de ces méthodes, le programme RANDU de la bibliothèque I.B.M. pour le système 360. Ce programme correspond à la méthode de congruence multiplicative avec :
$m = 2^{31}$ (impliqué par la capacité de la machine)
$a = 65539 = 8192 \times 8 + 3$.

Greenberger (1961) a donné un encadrement de la corrélation existant entre deux nombres consécutifs obtenus par la formule (2) ; le résultat nous paraît peu utilisable. D'autre

```
C
C     ..................................................................
C
C        SUBROUTINE RANDU
C
C        PURPOSE
C           COMPUTES UNIFORMLY DISTRIBUTED RANDOM REAL NUMBERS BETWEEN
C           0 AND 1.0 AND RANDOM INTEGERS BETWEEN ZERO AND
C           2**31. EACH ENTRY USES AS INPUT AN INTEGER RANDOM NUMBER
C           AND PRODUCES A NEW INTEGER AND REAL RANDOM NUMBER.
C
C        USAGE
C           CALL RANDU(IX,IY,YFL)
C
C        DESCRIPTION OF PARAMETERS
C           IX - FOR THE FIRST ENTRY THIS MUST CONTAIN ANY ODD INTEGER
C                NUMBER WITH NINE OR LESS DIGITS. AFTER THE FIRST ENTRY,
C                IX SHOULD BE THE PREVIOUS VALUE OF IY COMPUTED BY THIS
C                SUBROUTINE.
C           IY - A RESULTANT INTEGER RANDOM NUMBER REQUIRED FOR THE NEXT
C                ENTRY TO THIS SUBROUTINE. THE RANGE OF THIS NUMBER IS
C                BETWEEN ZERO AND 2**31
C           YFL- THE RESULTANT UNIFORMLY DISTRIBUTED, FLOATING POINT,
C                RANDOM NUMBER IN THE RANGE 0 TO 1.0
C
C        REMARKS
C           THIS SUBROUTINE IS SPECIFIC TO SYSTEM/360
C           THIS SUBROUTINE WILL PRODUCE 2**29 TERMS
C           BEFORE REPEATING
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           POWER RESIDUE METHOD DISCUSSED IN IBM MANUAL C20-8011,
C           RANDOM NUMBER GENERATION AND TESTING
C
C     ..................................................................
C
      SUBROUTINE RANDU(IX,IY,YFL)
      IY=IX*65539
      IF(IY)5,6,6
    5 IY=IY+2147483647+1
    6 YFL=IY
      YFL=YFL*.4656613E-9
      RETURN
      END
```

part, pour juger de la qualité des suites de nombres pseudo-aléatoires on effectue des tests statistiques comme par exemple celui donné par Hamnersley et Handscomb (1967) : test du χ^2 sur l'uniformité de la distribution des 10 000 premiers termes d'une suite de 2^{27} nombres pseudo-aléatoires obtenue par la récurrence (1).

3.4. - Nombres aléatoires ou pseudo-aléatoires pour une loi de distribution quelconque.

Soit $f(x)$ une fonction réelle non négative de la variable réelle x, possédant une intégrale de Stieltjes égale à 1. Soit η une variable aléatoire uniformément répartie sur (0,1). Il est aisé de montrer que la variable aléatoire ξ définie par la relation

$$(3) \qquad \eta = \int_{-\infty}^{\xi} f(x)\, dx ,$$

admet pour densité de probabilité la fonction $f(x)$.

Ce théorème fournit une méthode de construction de nombres aléatoires (ou pseudo-aléatoires) S_i correspondant à une loi de distribution donnée $f(x)$ à partir de nombres aléatoires (ou pseudo-aléatoires) uniformément répartis sur (0,1) R_i :

S_i est défini par l'inversion de la relation

$$R_i = \int_{-\infty}^{S_i} f(x)\, dx.$$

On trouve dans [1] au chapitre VII, dans le cas où il s'agit de nombres aléatoires, l'étude de certaines difficultés techniques liées au fait que les nombres R_i ... ne peuvent avoir qu'une distribution approximativement uniforme, étant représentés par un nombre fini de positions binaires.

On y donne également une autre méthode pour obtenir la variable aléatoire ξ à partir de la variable aléatoire η.

Il est à signaler que dans le cas où l'on cherche à obtenir une variable normale aucune des deux méthodes n'est applicable en pratique ; on utilise le théorème central limite (~~central limit theorem~~) rappelé au § 2, en considérant une somme d'une dizaine de variables aléatoires uniformément distribuées.

3.5. - Simulation d'événements aléatoires indépendants

Il est possible à partir d'une variable aléatoire uniformément distribuée sur (0,1) de simuler un événement A de probabilité p. Il suffit en effet de définir l'événement A comme l'événement :

$$\{\xi \leqslant p\} \; ;$$

on aura alors

$$P(A) = p$$

$$P(\bar{A}) = 1 - p \qquad \bar{A} = \text{contraire de } A.$$

De même on simulera un groupe de n événements incompatibles et complémentaires $(\sum_i p_i = 1)$.

Pour la simulation d'une succession d'événements et en particulier d'une chaîne de Markov on procède de la façon suivante :

Soit une chaîne de Markov définie par la matrice des probabilités de passage

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ & & & \\ p_{k1} & p_{k2} & & p_{kk} \end{pmatrix} \; ;$$

On part d'un état initial 0 avec les probabilités de passage initiales p_{01}, p_{02}, ... p_{0k}. Simulant les k événements correspondants, incompatibles et complémentaires, comme indiqué plus haut, on détermine par le tirage d'un nombre aléatoire (ou pseudo-aléatoire) uniformément répartis sur (0,1), l'état i dans lequel se trouve le système. Le processus recommence alors à partir de l'état i avec les probabilités de passage p_{i1}, p_{i2}, .. p_{ik}, etc.

Là encore le fait que les nombres aléatoires ne comportent qu'un nombre fini de positions binaires et ne soient donc uniformément répartis que de façon approchée peut conduire à effectuer certaines corrections sur les méthodes indiquées ci-dessus, exposées dans [1] chapitre VII.

3.6. - Nombres quasi-aléatoires

Nous réservons de parler de ces nombres quand leur intérêt apparaîtra (§ 4.6).

4. - METHODES DE MONTE-CARLO POUR LE CALCUL DES INTEGRALES DEFINIES.

Une des applications les plus importantes des méthodes de Monte-Carlo est le calcul des intégrales définies simples ou multiples. Les techniques sont les mêmes quelle que soit la multiplicité de l'intégrale et nous les exposerons sur l'exemple de l'intégrale simple.

4.1. - Calcul de l'intégrale $J = \int_0^1 f(x)\,dx$

4.1.1. - Première méthode : calcul de la valeur moyenne

ξ désignant la variable aléatoire uniformément répartie sur (0,1), et ξ_1 , ξ_2 , ξ_N étant N valeurs aléatoires de cette variable, la loi des grands nombres indique que si $N \to \infty$, $\dfrac{f(\xi_1) + f(\xi_2) + .. + f(\xi_N)}{N}$ tend vers l'espérance mathématique de $f(x)$, $E\,f(x) = \int_0^1 f(x)\,dx = J$.

On en déduit une méthode de calcul de l'intégrale J : pour une suite de n valeurs de la variable aléatoire ξ , on accumule $f(\xi_1) + f(\xi_2) \ldots + f(\xi_i)$ et on calcule à la fin

$$(4) \qquad \theta_1 = \frac{\sum_1^N f(\xi_i)}{N} \text{ , estimation de } J.$$

4.1.2. - Deuxième méthode : calcul de l'aire

Supposons, pour simplifier l'exposé, que la fonction $f(x)$ à intégrer soit comprise entre 0 et 1.

J n'est autre que l'aire comprise entre la courbe représentative de $f(x)$, l'axe des x et les verticales $x = 0$, $x = 1$; ou encore le rapport de cette aire à l'aire du carré (1 x 1). Cette constation conduit à une autre méthode d'évaluation de l'intégrale J.

Considérons (ξ, η) un point aléatoire uniformément réparti dans le carré (1×1), et soient (ξ_i, η_i) une suite de N valeurs aléatoires de ce point : sur ces N points, N' sont situés au-dessous de la courbe $y = f(x)$. Il est évident que si $N \to \infty$, N'/N tend vers le rapport des aires indiqué plus haut c'est-à-dire vers J.

En pratique il suffit de tester, pour chaque point (ξ_i, η_i), la condition $\eta_i \leqslant f(\xi_i)$ et d'augmenter N' d'une unité si la réponse est affirmative. On calcule à la fin

(5) $$\theta_2 = \frac{N'}{N}, \text{ estimation de } J.$$

Cette méthode est aussi appelée méthode du tirage noir ou blan.

Les méthodes d'estimation de l'erreur (cf. §4.2), permettent de montrer que l'estimation (4) est de meilleure qualité que l'estimation (5). Pourtant celle-ci est parfois plus avantageuse, en particulier lorsque la fonction f est définie par une relation implicite, pour les calculs d'aires intérieures etc...

4.1.3. - Remarques

Nous pensons que ces deux exemples de méthodes de calcul expliquent l'importance que nous avons attachée au § 3.3 à la corrélation interne dans les suites de nombres pseudo-aléatoires.

On peut également remarquer d'ores et déjà certaines caractéristiques générales des méthodes de Monte-Carlo:

elles nécessitent d'effectuer un grand nombre de calculs simples identiques ;

elles utilisent peu de mémoires dans le calculateur pour le stockage des résultats intermédiaires.

4.2. - Estimation statistique de l'erreur

La tendance vers la loi normale indiquée par le théorème central limite permet une évaluation statistique de l'erreur plus commode qu'au moyen de l'inégalité de Bienaymé.

Désignons par $E\theta$ l'espérance mathématique de la variable approximativement normale θ et par $\mu_2\theta$ sa variance, et soit x_p défini par

$$\frac{2}{\sqrt{2\pi}} \int_{-\infty}^{x_p} e^{-t^2/2}\, dt - 1 = p,$$

la probabilité de l'inégalité

(6) $|\theta - E\theta| \leqslant x_p \sqrt{\mu_2 \theta}$ est approximativement égale à p.

On prend en pratique un intervalle de confiance pour p = 95 %.

La variable aléatoire θ est la moyenne de n variables ζ_i dont l'espérance mathématique est égale à J : On a

$$E\zeta = E\theta = J \quad \text{et}$$

(7) $$\mu_2 \theta = \frac{\mu_2 \zeta}{n}$$

d'où

(8) probabilité p pour que $|\theta - J| \leqslant x_p \frac{\sqrt{\mu_2 \zeta}}{\sqrt{n}}$

La variance $\mu_2 \zeta$ peut être évaluée au fur et à mesure du calcul au moyen des estimateurs habituelles en statistique (cf. [3]).

Ainsi d'après la formule (8), la précision est en $1/\sqrt{n}$. Quant à l'efficacité d'une méthode qui est le facteur déterminant pour le choix, elle se mesure par l'inverse du temps de calcul nécessaire pour atteindre une précision (statistique) donnée ; temps qui est proportionnel à N nombre d'expériences nécessaires et au temps de calcul nécessité par chaque expérience (fonction de la complexité de celle-ci).

4.3. - Techniques de réduction de la variance.

L'intervenction du facteur $\sqrt{\mu_2 \zeta}$ au numérateur dans la formule (8) montre qu'il y a intérêt à réduire la variance $\mu_2 \zeta$ pour augmenter la précision atteinte avec un même nombre d'expériences. Diverses techniques sont utilisées pour cela qu'il nous paraît intéressant d'exposer rapidement afin d'en dégager les idées importantes.

4.3.1. - Utilisation des variables de contrôle ("extraction of the regular part")

L'idée qui est à la base de ce procédé, et qui est une règle générale concernant les méthodes de Monte-Carlo,

est qu'il y a toujours intérêt lorsque cela est possible à effectuer certaines parties des intégrations par des quadratures de façon à n'avoir sur celles-ci que les erreurs d'arrondi et à ne faire porter les intégrations par les techniques de Monte-Carlo que sur des quantités plus faibles relativement à la valeur de l'intégrale cherchée.

Ici pour le calcul de l'intégrale J on fait la décomposition

$$J = \int_0^1 g(x)\,dx + \int_0^1 \big(f(x) - g(x)\big)\,dx \tag{9}$$

où $g(x)$ est une fonction assez simple pour pouvoir être intégrée formellement et qui suit cependant d'assez près les variations de $f(x)$ de façon que la deuxième intégrale, à calculer par la méthode de Monte-Carlo, soit petite par rapport à J. $g(x)$ est appelée la variable de contrôle de $f(x)$.

Définissant la variable aléatoire ζ par :

$$\zeta = f(\xi) - g(\xi) + I,$$

$$\theta_3 = \frac{1}{N} \sum_{i=1}^{N} \big(f(\xi_i) - g(\xi_i)\big) + I$$

est l'estimation de J.

Cette méthode correspond à une diminution de la variance de ζ, et de θ à N fixé, par rapport à celle du § 4.1.1.

Dans la pratique on prend souvent pour $g(x)$ les premiers termes du développement en série de $f(x)$.

4.3.2. - Echantillonnage suivant l'importance

Ce procédé correspond à l'idée intuitive qu'il y a intérêt dans l'évaluation de l'intégrale à utiliser plus de points dans les zones qui ont le plus d'importance.

On emploie une variable non uniformément répartie.

Soit ξ une variable aléatoire dont la densité de probabilité sur $(0,1)$ est $p(x)$ $\left(\int_0^1 p(x)\,dx = 1\right)$.

L'intégrale J peut être transformée en

$$J = \int_0^1 \frac{f(x)}{p(x)}\, p(x)\,dx \tag{11}$$

et apparaît comme l'espérance mathématique de la fonction ζ de ξ :

(12) $$\zeta = \frac{f(\xi)}{p(\xi)}$$

(13) $$E\zeta = J$$

ainsi

(14) $$\theta_4 = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\xi_i)}{p(\xi_i)}$$ est une estimation de J.

On montre par le calcul des variations et compte tenu des exigences pratiques, qu'il y a intérêt pour minimiser $\mu_2 \zeta$ à prendre une variable aléatoire ξ dont la densité de probabilité soit à peu près proportionnelle à la valeur absolue de la fonction à intégrer :

$$\frac{p(x)}{|f(x)|} \simeq \text{cte}$$

ce qui correspond bien à l'idée intuitive de départ.

4.3.3. - Echantillonnage stratifié

L'idée de base est la même qu'au § précédent mais l'application consiste à tronçonner l'intervalle d'intégration en intervalle plus petits, $l_1 + l_2 + .. + l_n = 1$, sur chacun desquels on effectue l'intégration séparément comme au § 4.1.1. Pour un nombre total N de valeurs de la variable aléatoire utilisées, on cherche à déterminer les nombres N_k à utiliser pour chaque intégration ($\sum_1^n N_k = N$) pour minimiser la variance.

Les raisonnements et les conclusions sont du même type qu'au § 4.3.2. Dans la pratique on prend N_k proportionnel à l_k ce qui conduit à une meilleure variance que par l'application de la méthode du § 4.1.1 globale.

4.3.4. - Variables antithétiques

Signalons encore l'utilisation de variables antithétiques dont l'importance tend à s'affirmer (voir à ce sujet : Hammersley et Morton (1956), Hammersley et Mauldon (1956), Handscomb (1958)). Un exemple particulier en est la symétrisation de l'intégrande :

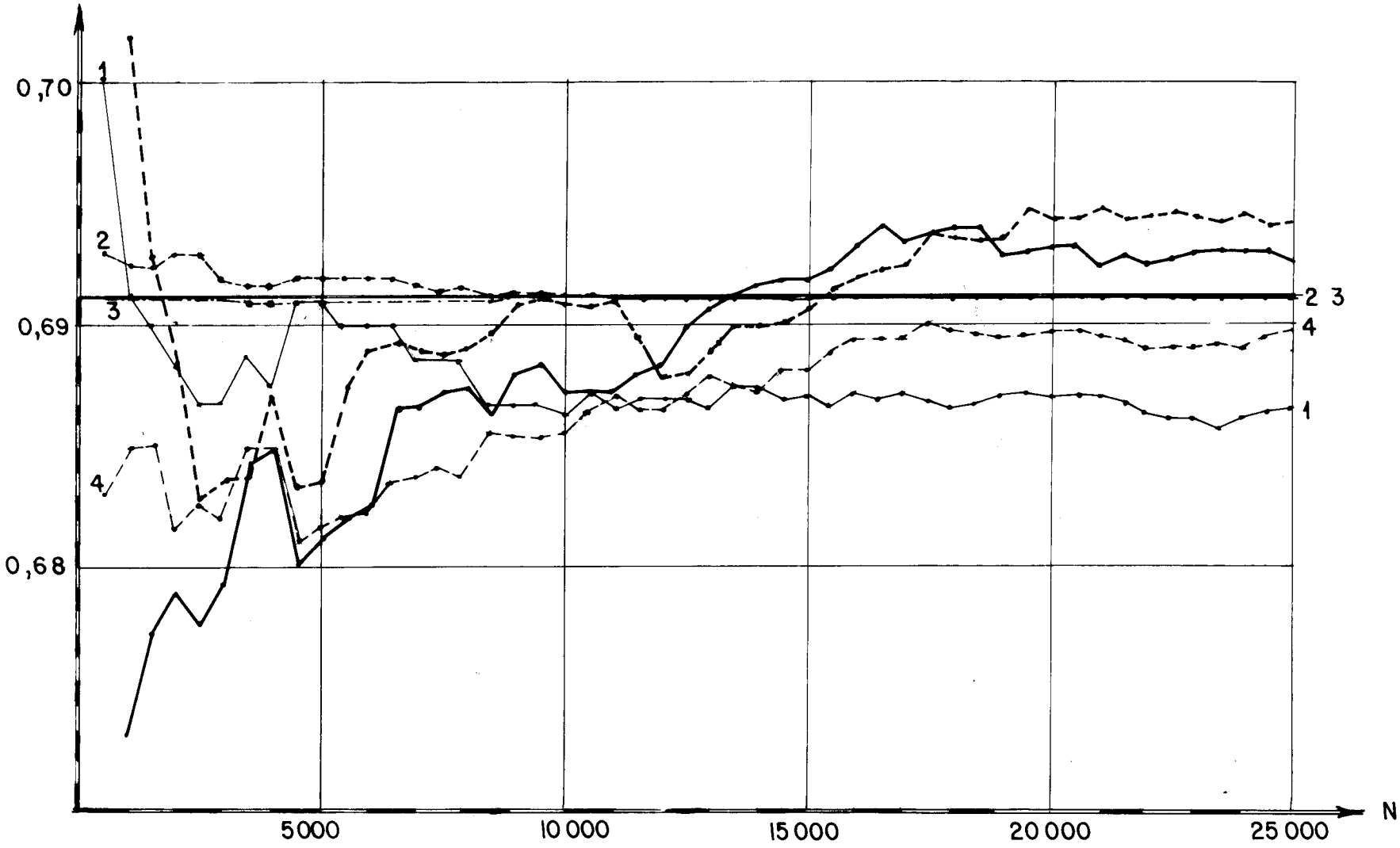

figure 2

ξ étant une variable aléatoire uniformément distribuée sur (0,1), on considère

(15) $\zeta = \frac{1}{2}\left(f(\xi) + f(1 - \xi)\right)$ dont l'espérance mathématique est $E\zeta = J$.

(16) $\theta_6 = \frac{1}{2N} \sum_1^N \left(f(\xi_i) + f(1 - \xi_i)\right)$ est une estimation de l'intégrale J.

4.3.5. - Un exemple de calcul.

Nous donnons sur la figure 2 les résultats d'un calcul que nous avons effectué.

L'intégrale à calculer est $J = \int_0^1 \sqrt{1 - x^2} \cos x \, dx$, dont la valeur exacte est connue.

Nous avons utilisé la méthode du § 4.1.1 deux fois avec les nombres pseudo-aléatoires fournis par le programme RANDU déjà présenté en prenant deux valeurs différentes pour x_0 dans la formule (1) ($x_0 = 87$ et $x_0 = 1357973$) : les courbes 1 et 4 correspondent à ces calculs.

La méthode de la variable de contrôle avec les mêmes nombres pseudo-aléatoires ($x_0 = 1357973$) a donné les résultats de la courbe 2 : la variable de contrôle est $(1 - x^2)$, premiers termes du développement en série de la fonction à intégrer.

En appliquant simultanément la méthode de la variable de contrôle et l'échantillonnage suivant l'importance on obtient les résultats de la courbe 3, ($x_0 = 1357973$) : variable de contrôle $(1 - x^2)$; densité de probabilité de la variable aléatoire $= 5x^4$, proportionnelle au troisième terme du développement de la fonction à intégrer.

Les temps de calcul ont été

pour	1	38 s
	4	39 s
	2	63 s
	3	74 s.

On remarque l'efficacité des méthodes de réduction de la variance : à temps de calcul égal ($N \# 12000$) 2 et 3 fournissent des évaluations bien meilleures de J que 1 et 4 pour $N = 25\,000$. On voit qu'en prenant 25000 termes de la suite de 2^{29} termes fournie par RANDU on obtient par la méthode directe la valeur de l'intégrale à 1 % près environ.

4.4. - Calcul des intégrales multiples

Il est clair que les méthodes présentées aux § 4.1.1 et 4.1.2, sont applicables aussi bien pour le calcul des intégrales multiples que pour celui des intégrales simples :

Soit à calculer l'intégrale

$$J = \int_0^1 \int_0^1 .. \int_0^1 f(x_1 , x_2 , .. x_n) \, dx_1 \, dx_2 .. dx_n .$$

Soit ξ la variable aléatoire uniformément répartie sur (0,1), et ξ_1, ξ_2 , ξ_i ... une suite de valeurs aléatoires ou pseudo-aléatoires, de cette variable.

Pour la méthode du § 4.1.1, on prélève dans cette suite les valeurs par groupes de n , pour être les coordonnées d'un point aléatoire du n-cube, (1 x 1 x .. x 1).

Pour la méthode du § 4.1.2, (tirage blanc ou noir), on prélève les valeurs par groupes de n +1 : n coordonnées d'un point aléatoire du n-cube, et la (n + 1) ème valeur pour être comparée à la valeur en ce point de la fonction à intégrer.

En ce qui concerne la réduction de la variance, les techniques indiquées dans les paragraphes précédents sont générales. On ajoute à cela un principe analogue à ce qui a été dit au § 4.3.1. : il y a toujours intérêt à effectuer analytiquement les intégrations partielles pour lesquelles cela est possible.

D'autre part dans le cas où les intégrales présentent des singularités, la méthode d'échantillonnage suivant l'importance doit être appliquée et on prend alors une variable aléatoire dont la densité de probabilité présente le même type de singularités, que la fonction à intégrer.

4.5. - Utilisation des méthodes d'intégration de Monte-Carlo.

Le caractère fondamental des méthodes de Monte-Carlo est le caractère probabiliste de la convergence exprimé par la loi forte des grands nombres : on peut trouver N tel que si n , nombre d'expériences, $\geqslant$ N, la probabilité pour que le résultat trouvé diffère de la solution exacte de moins de ε donné, soit supérieure ou égale à p donné < 1 (95 % par exemple).

Cette différence avec les méthodes classiques ne doit pourtant pas, du point de vue pratique, les faire écarter sans une étude approfondie.

Nous avons vu que l'ordre de convergence de la méthode de Monte-Carlo est en $1/\sqrt{N}$, ce qui est assez faible. Mais il est remarquable que ce résultat est valable quelle que soit la multiplicité de l'intégrale, ce dont on peut tirer les conclusions suivantes :

lorsque l'on fait croître la multiplicité de l'intégrale à effectuer, le nombre d'opérations nécessité par la méthode de Monte-Carlo pour une précision donnée, augmente seulement en fonction de la complexité de l'intégrande et du nombre de variables aléatoires uniformément distribuées à utiliser, c'est-à-dire, en gros, proportionnellement à la multiplicité n de l'intégrale ;

par contre avec les méthodes classiques de l'analyse numérique le nombre d'opérations augmente en K^n.

Cela explique l'intérêt des méthodes de Monte-Carlo pour le calcul des intégrales multiples.

On trouve dans (1) , chapitre II, les renseignements suivants : pour une fonction "suffisamment régulière" à intégrer dans un domaine au contour "suffisamment régulier", on utilise la méthode de Monte-Carlo pour $n > 4$, parfois pour $n \geqslant 3$;

pour une fonction "régulière par morceau" (possédant des discontinuités et des variations brutales localisées), on utilise la méthode de Monte-Carlo pour $n \geqslant 2$;

pour une fonction très peu régulière, on utilise en général la méthode de Monte-Carlo dès $n = 1$.

Richtmyer (1958) donne des conclusions à peu près analogues.

4.6. - Autre approche des méthodes de Monte-Carlo ; Notion de suite uniformément dense ; Nombres quasi-aléatoires.

Définition d'une suite uniformément dense sur (0,1) : considérons deux segments quelconques de l'intervalle (0,1) de même longueur l ; parmi les N premiers termes de la suite x_i , il y en a a_N dans le premier segment et b_N dans l'autre ; la suite x_i est uniformément dense sur (0,1) si, lorsque $N \to \infty$, les rapports a_N/N et b_N/N tendent vers une limite commune égale à l.

Théorème de Hermann Weyl : pour que la suite x_i soit uniformément dense sur (0,1) il faut et suffit que pour toute fonction intégrale H(x), définie sur (0,1) on ait :

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{i=N} H(x_i) = \int_0^1 H(x)\, dx$$

La notion de suite de nombres uniformément dense est généralisable pour une suite de points uniformément dense dans le n-cube (1 x 1 x .. x1)

Le théorème de H. Weyl fournit donc, sans aucun appel au calcul des probabilités une adaptation de la première méthode de Monte-Carlo (du § 4.1.1). On démontre (cf. [4]), qu'il conduit aussi à une adaptation de la deuxième méthode de Monte-Carlo.

On peut donc utiliser les méthodes de Monte-Carlo, avec des suites de nombres ou de points uniformément denses. Le caractère probabiliste de la convergence (§ 4.5) disparaît alors comme le montre l'énoncé du théorème de Weyl.

Un premier exemple de suite de nombres uniformément dense sur (0,1) est fourni par les suites de Van der Corput (1935) :

Si le nombre entier naturel i est représenté dans la base r par :

$$i = a_m\, a_{m-1} \ldots a_1\, a_0 \qquad 0 \leqslant a_s \text{ entier } \leqslant r - 1,$$

alors, dans la base r, le i ème terme de la suite est représenté par

$$p_r(i) = 0,\, a_0\, a_1 \ldots a_{m-1}\, a_m.$$

Par exemple (base r = 2) :

i = 1(décimal)	= 1(binaire)	p_r(i)=0,1(binaire)	= 0,5(décimal)
2	10	0,01	0,25
3	11	0,11	0,75
4	100	0,001	0,125
5	101	0,101	0,625
6	110	0,011	0,375
7	111	0,111	0,875
8	1000	0,0001	0,0625

Ce type de suites a été généralisé par Halton (1960) pour le cas à n dimensions : on obtient une suite uniformément dense sur le n cube (1 x 1 .. 1) en prenant pour chaque coordonnée des points une suite de Van der Corput, les nombres, bases de suites, étant premiers entre eux.

Ces nombres (ou ces points) apparaissent ainsi comme des nombres quasi-aléatoires utilisables pour l'intégration par les méthodes de Monte-Carlo.

A ce stade la distinction apparaît peut-être difficile à faire entre nombres pseudo-aléatoires et nombres quasi-aléatoires ; en effet les méthodes d'obtention des uns et des autres sont bien semblables.

Pour les nombres pseudo-aléatoires on essaye de satisfaire le plus grand nombre de tests statistiques de façon à ce que ces nombres puissent servir à de multiples usages. On veut pouvoir utiliser ces nombres comme de véritables nombres aléatoires. Par contre pour les nombres quasi-aléatoires on ne s'intéresse qu'à la propriété utile pour le problème traité : ici cette propriété est la régularité de l'uniformité de la distribution, sur lequel le caractère uniformément dense de la suite donne des renseignements précis.

Ainsi, si l'on considère une suite de nombres pseudo-aléatoires produits par la formule (1), et que l'on écrit ces nombres en numération binaire, un des tests que l'on peut vérifier (ceci n'est en fait pas vérifié pour les dernières positions binaires) est que la suite de 1 et de 0 dans la i ème position binaire du nombre, est <u>aléatoire</u> avec les probabilités 1/2 et 1/2 ; il est évident sur le tableau (17) qu'il n'est pas question de vérifier cette propriété avec les nombres de Van der Corput (suite de 1 et de 0 régulière dans chaque position).

D'autres exemples de suites uniformément denses ont été donnés par Richtmyer (1958) et Bass et Guilloud (1958) [4].

Dans [4] on utilise en particulier le théorème :

Si $\phi(t) = a_0 + a_1 t + .. + a_\nu t^\nu$ est un polynome de degré $\nu \geqslant 2$, et si a_ν est irrationnel, la suite

$$x_n = \underbrace{\phi(n)} \quad , \quad y_n = \underbrace{\phi(n + p)}, \quad p \text{ entier} \geqslant 1,$$

où $\smile$ désigne la partie décimale, est uniformément dense sur le carré (1 x 1). Ce résultat est généralisable à k dimensions, auquel cas on doit avoir $\nu \geqslant k$.

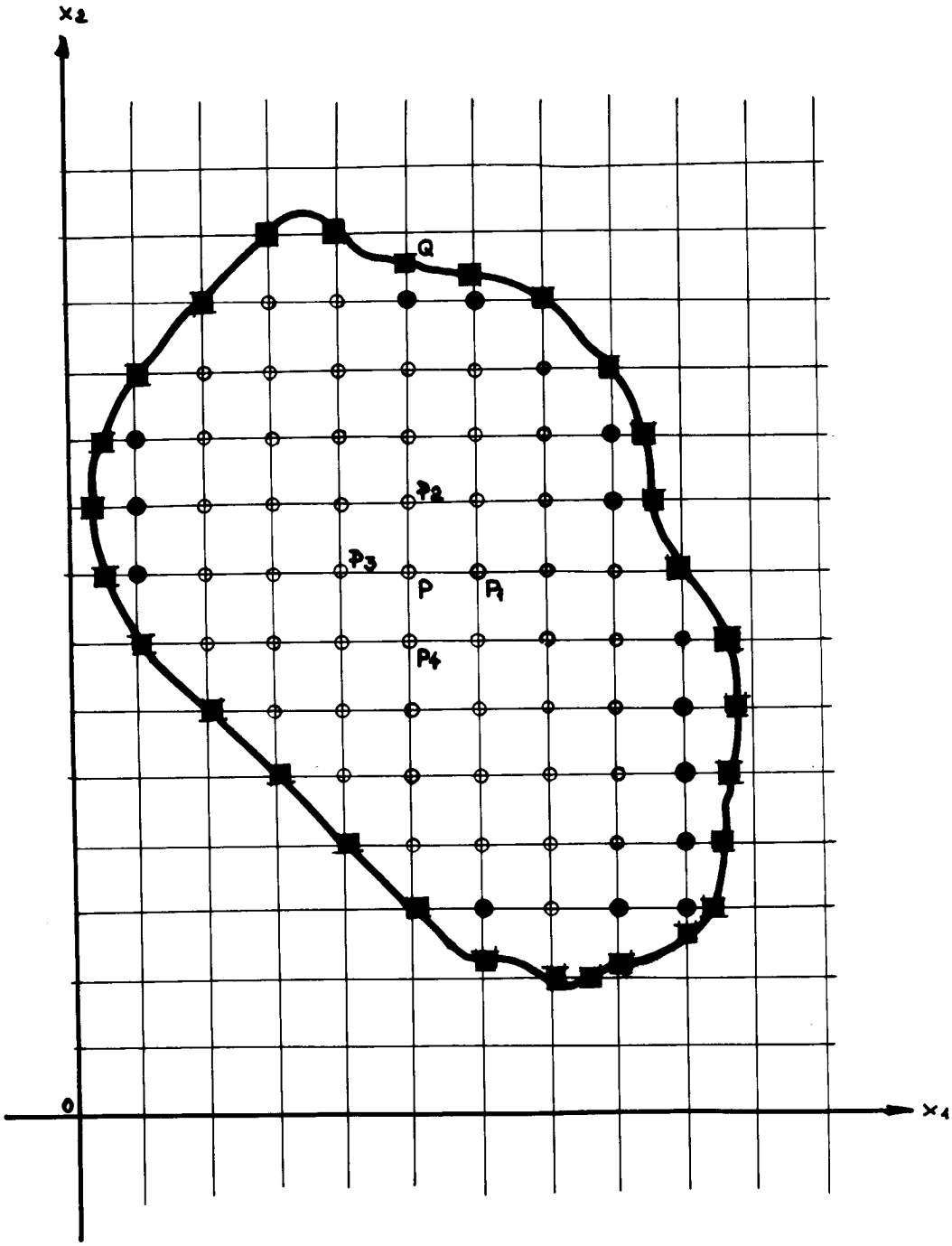

figure 3

Il est évident que le caractère irrationnel de a_ν est fondamental pour la démonstration du théorème ; or les nombres irrationnels ne pouvant être exactement représentés dans un calculateur on peut se demander quelle est la conséquence pratique de l'erreur que l'on commet ainsi nécessairement : si l'on utilise N nombres de la suite, cette erreur peut être négligée à condition que le nombre de chiffres significatifs soit suffisamment grand (cf. [2] p. 45).

Nous avons reproduit sur la figure 2, les résultats donnés dans [4] pour le calcul de l'intégrale $\int_0^1 \sqrt{1 - x^2}\cos x \, dx$, avec le polynome $\phi(t) = \pi t^2$ (trait conitnu épais) par la deuxième méthode de Monte-Carlo. Il est à noter que si l'on utilise la première méthode, l'emploi de variable de contrôle est toujours possible et améliore dans ce cas aussi la rapidité de la convergence.

On doit retenir de cette approche des méthodes de Monte-Carlo en termes de théorie des nombres, que l'utilisation de suites arithmétiques dont les propriétés sont bien connues, peut se révéler meilleure que celle de suites pseudo-aléatoires qui relèvent toujours d'un certain empirisme. Elle permet en particulier une meilleure estimation de l'erreur commise.

5. - AUTRES UTILISATIONS DES METHODES DE MONTE-CARLO.

Il n'est pas question ici d'exposer toutes les utilisations des méthodes de Monte-Carlo ; on pourra pour cela se reporter à [1] et [2]. Citons parmi les applications possibles : la résolution de systèmes d'équations algébriques linéaires, la recherche des valeurs propres et des fonctions propres d'opérateurs différentiels linéaires, la résolution d'équations intégrales de Fredholm etc..., qui font appel à la simulation de processus Markoviens.

Nous étudierons un exemple : la résolution du problème de Dirichlet.

5.1. - Résolution du problème de Dirichlet : Méthode du chemin de Pólya.

Nous exposons la méthode dans le cas du problème à deux dimensions, pour la commodité des représentations graphiques.

On cherche une fonction $u \in C^2(D) \cap C^0(\bar{D})$, satisfaisant $\Delta u = 0$ sur D et $u = f$ donné sur ∂D.

On commence par se ramener au problème aux différences finies correspondant, et c'est ce problème que l'on cherche à résoudre par une méthode de Monte-Carlo.

Un maillage carré de même pas h dans les deux directions est mis en place. On considère les noeuds P de ce maillage qui appartiennent à D, et les points d'intersections Q (noeuds ou non) des lignes du réseau avec ∂D. (fig. 3).

La discrétisation de l'opérateur différentiel donne les relations :

en un point P dont les quatre voisins P_i sont à égale distance h de lui :

$$(18) \qquad u(P) = \frac{1}{4}\left(u(P_1) + u(P_2) + u(P_3) + u(P_4)\right)$$

en un point P dont les voisins sont à des distances différents h_i de lui (point P dont un ou plusieurs voisins sont des points Q appartenant à ∂D) :

$$(19) \qquad u(P) = u(P_1)\,\frac{h_2\,h_3\,h_4}{(h_1+h_3)(h_1h_3+h_2h_4)} + u(P_2)\,\frac{h_1\,h_3\,h_4}{(h_2+h_4)(h_1h_3+h_2h_4)} + \ldots$$

ou encore

$$(20) \qquad u(P) = \sum_{i=1}^{i=4} p_i\, u(P_i) \text{ , on remarque que } \sum_{i=1}^{i=4} p_i = 1.$$

Les valeurs $u(Q)$ sont connues et il s'agit de déterminer la valeur de $u(P)$ en chaque point P de D.

Soit un point P où on cherche à déterminer $u(P)$ solution du problème aux différences finies approximation de la solution du problème de Dirichlet.

On considère un chemin aléatoire qui part de P et qui procède par étapes en chaque noeud du réseau, chaque fois aléatoirement vers l'un des quatre points voisins, la probabilité de passage d'un point étape P vers son voisin P_i étant p_i de la formule (20) ; jusqu'à ce qu'il atteigne la frontière en un point Q où le chemin s'arrête ; on lui associe alors la valeur $f(Q)$.

Désignons par $v(P,Q)$ la probabilité pour qu'un chemin aléatoire partant de P aboutisse au point Q. On a évidemment :

$$(21) \qquad v(P,Q) = \sum_{i=1}^{i=4} p_i v(P_i, Q)$$

et

$$(22) \qquad v(Q,Q) = 1$$

car si le point est un point du contour, le chemin part et finit en ce point.

Si l'on considère tous les chemins aléatoires issus de P, l'espérance mathématique des valeurs attribuées à ces chemins est :

$$(23) \qquad w(P) = \sum v(P, Q_j).f(Q_j) \ ,$$

où les Q_j sont les j points Q du contour ∂D. Associant (21), (22) et (23) il est clair que $w(P)$ vérifie :

$$(24) \qquad w(P) = \sum_{i=1}^{i=4} p_i \, w(P_i) \quad \text{et} \quad w(Q) = f(Q).$$

<u>$w(P)$ est donc la valeur de la fonction cherchée, $u(P)$, solution du problème aux différences finies.</u>

Ainsi pour obtenir $u(P)$ on fait un grand nombre d'expériences de tracer de chemins aléatoires issus du point P et on prend la moyenne des valeurs de f au point final de chacun de ces chemins.

Le tracer d'un chemin aléatoire issu de P n'est autre que la simulation d'une chaîne de Markov à états absorbants.

Il existe des variantes de cette méthode.

5.2. - Résolution du problème de Dirichlet : Méthode de Muller

La méthode présentée au § précédent converge très lentement aussi utilise-t-on plutôt la méthode de Muller.

Soit à résoudre le problème de Dirichlet: trouver $u \in C^2(D) \cap C^0(\bar{D})$ satisfaisant

$$\Delta u = f \text{ sur } D \qquad \text{et} \qquad u = f \text{ sur } \partial D.$$

La réciproque du théorème de la moyenne (cf. Courant-Hilbert, methods of mathematical physics tII p. 279) permet d'énoncer que :

résoudre le problème de Dirichlet est équivalent à chercher une fonction u telle que sa valeur en un point quelconque de D soit la moyenne de ses valeurs sur une sphère (au moins) de rayon non nul centrée en ce point et intérieure à D, et qui soit égale à f sur ∂D.

La méthode de Muller est basée sur cette propriété.

Désignons par $S(P)$ la plus grande sphère de centre P intérieure à D. On considère une suite de points : $P_0 \equiv P$, P_1, P_2, .. où P_{m+1} est un point aléatoire à la surface de $S(P_m)$ jusqu'à ce que l'on atteigne un point P_k suffisamment proche de ∂D. La suite se termine alors et on lui associe la valeur $f(Q)$ de f et le point du contour le plus proche de P_k.

Soit $v(P,Q)$ la densité de probabilité pour que la suite aléatoire de points issue du point P se voit attribuer la valeur $f(Q)$. Il est clair que, P_1 étant choisi au hasard sur $S(P)$ on a :

$$(25) \qquad v(P,Q) = \left(\int_{S(P)} v(P_1, Q)\, dS\right) / \left(\int_{S(P)} dS\right).$$

et $v(Q,Q) = \delta(Q)$.

Si l'on considère toutes les suites de points issues de P, du type indiqué plus haut, l'espérance mathématique des valeurs qui leur sont attribuées est :

$$(26) \qquad w(P) = \int_{\partial D} v(P, Q)\, f(Q)\, dS \;;$$

et d'après (25) :

$$w(P) = \frac{\int_{S_p} w(P_1)\, dS}{\int_{S_p} dS}$$

(26) et $w(Q) = f(Q)$.

w(P) est donc la valeur en P de la solution du problème de Dirichlet.

Pour estimer w(P) , il suffit d'effectuer un grand nombre d'expériences de constructions de suites $P_0 \equiv P, P_1, \ldots$ et de prendre la moyenne des valeurs f(Q) correspondantes. Il s'agit là encore de la simulation d'un processus Markovien.

Il est à remarquer qu'elle apparaît sous une forme totalement différente de la précédente : sachant que la fonction cherchée satisfait la réciproque du théorème de la moyenne, on cherche à simuler cette propriété. C'est un exemple de simulation de loi limite.

Signalons enfin que des méthodes analogues peuvent être employées pour la résolution de problème de Neumann, de l'équation de Poisson etc...

BIBLIOGRAPHIE

[3] J. Bass (1967) Eléments de calcul des probabilités, 2ème éd. Masson, Paris.

[4] J. Bass et J. Guilloud (1958) Chiffres 1, 4, 149-155.

Buffon (1777) Essai d'arithmétique morale.

G.E. Forsythe (1951) Nat. Bur. Stand., Appl. Math. Séries 12, 34-35.

M. Greenberger (1961) J. Assoc. Comp. Mach. 8, 163-167.

J.M. Hammersley et K.W. Morton (1956) Proc. Camb. Phil. Soc. 52, 449-475.

J.M. Hammersley et J.G. Mauldon (1956) Proc. Camb. Phil. Soc. 52, 476-481.

[2] J.M. Hammersley et D.C. Handscomb (1967) Les méthodes de Monte Carlo, Dunod, Paris.

D.C. Handscomb (1958) Proc. Camb. Phil. Soc. 54, 300-301.

J.H. Halton (1960) Nümerische Math. 2, 84-90 et 196.

D.H. Lehmer (1951) Ann. Comput. Lab. Harvard Univ. 26, 141-146.

G. Marsaglia (1968) Proc. US Nat. Ac. Sc. 61, 1.

Rand Corporation (1955) A million random digits with 100 000 normal deviates, Glencoe, Illinois, Free Press.

R.D. Richtmyer (1958) Inst. Math. Sc. New-York Univ. Report NYO-8764 Physics.

[1] Yu.A. Shreider (1966) The Monte Carlo Method, Pergamon Press.

J.C. Van der Corput (1935) Proc. Kon. Akad. Wet. Amsterdam, 38, 813-821 et 1058-1066.