

Recherche opérationnelle - 1 À propos de la méthode de Monte-Carlo

Claude Tricot

Volume 38, numéro 3, octobre–décembre 1962

URI : <https://id.erudit.org/iderudit/1001851ar>

DOI : <https://doi.org/10.7202/1001851ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)

1710-3991 (numérique)

[Découvrir la revue](#)

Citer cet article

Tricot, C. (1962). Recherche opérationnelle - 1 : à propos de la méthode de Monte-Carlo. *L'Actualité économique*, 38(3), 425–444.
<https://doi.org/10.7202/1001851ar>

Analyses

Recherche opérationnelle - 1

À propos de la méthode de Monte-Carlo

Les techniques de la recherche opérationnelle intéressent un nombre croissant de domaines scientifiques. Elles ont, en particulier, sensiblement réduit la marge qui sépare encore l'administration de la théorie économique. Par ailleurs, il s'agit d'un domaine où les problèmes méthodologiques abondent, donnant lieu à une littérature qui ne peut être suivie au jour le jour que par un groupe relativement restreint de spécialistes. Pour permettre aux lecteurs de L'Actualité Économique de se tenir au fait de l'évolution de la technique de la recherche opérationnelle, le professeur Claude Tricot, du Service de la Recherche opérationnelle de l'École des Hautes Études commerciales fera périodiquement une revue de la littérature la plus récente sur ce sujet et discutera des problèmes qui lui paraîtront les plus significatifs.

On connaît la méthode dite de Monte-Carlo qui consiste, en recherche opérationnelle, à simuler un processus aléatoire en utilisant des nombres «au hasard» fabriqués artificiellement.

Prenons un exemple couramment proposé :

Un grand magasin se charge de la livraison à domicile des achats de ses clients, autant que possible dans les 24 heures; il s'agit de déterminer l'importance de la flotte de camions de livraison.

Le nombre de colis à livrer dans la journée du mercredi, par exemple, est aléatoire, mais étant donné la multitude de causes susceptibles de modifier légèrement ce nombre, et leur indépendance, on peut le considérer comme distribué suivant une loi normale.

Même réflexion pour le nombre de colis livrés par un camion.

Les seules données qu'on demandera à l'expérience sont: moyenne et écart-type de ces deux nombres aléatoires; elles permettront, à l'aide d'un échantillon de nombres issus d'une population parente normale, de fabriquer deux suites de nombres représentant les colis à livrer et ceux qui seront livrés par un camion au cours de mercredis, non réels, mais possibles.

Pour fixer les idées, considérons les colis à livrer et supposons qu'il y en ait 6,000 en moyenne, l'écart-type étant de 1,000 colis.

On peut fabriquer un échantillon de nombres issus d'une population normale ou prendre cet échantillon dans des tables comme celles qui sont publiées par la Rand Corporation: ces nombres ont pour moyenne 0 et pour écart-type 1.

Supposons obtenue la suite: -1.361

-1.799

-0.672

1.830

0.197

elle nous permettra de construire des mercredis possibles:

$$6,000 + 1,000 (-1.361) = 4,639 \text{ colis}$$

$$6,000 + 1,000 (-1.799) = 4,201 \text{ "}$$

$$6,000 + 1,000 (-0.672) = 5,328 \text{ "}$$

$$6,000 + 1,000 (1.830) = 7,830 \text{ "}$$

$$6,000 + 1,000 (0.197) = 6,197 \text{ "}$$

Même manœuvre pour les livraisons, après quoi l'analyse portera sur ces échantillons artificiels.

Mon propos d'aujourd'hui n'est pas le dénombrement des applications fort nombreuses de ces méthodes à fabriquer du hasard (rappelons cependant leur usage dans les techniques de sondages ou même dans la résolution d'équations aux dérivées partielles, l'inversion de matrices, etc.), mais de critiquer leur base même: l'obtention de nombres «au hasard».

Cette étude sera divisée en deux parties plus ou moins indépendantes:

- a) Comment fabrique-t-on ces nombres;
- b) Comment s'assurer que cette fabrication est légitime.

*
* *

Le premier problème qui se présente porte sur la construction de nombres appartenant à un certain intervalle, disons l'intervalle $(0, 1)$, et distribués suivant une loi uniforme, ce qui signifie qu'aucun parmi les nombres compris entre 0 et 1 n'est privilégié, ou, plus précisément, que la probabilité d'obtenir un nombre appartenant à un intervalle (a, b) lui-même compris dans $(0, 1)$ ne dépend que de l'étendue de (a, b) et nullement de sa position dans l'intervalle $(0, 1)$.

La première idée qui vient à l'esprit suggère d'obtenir de tels nombres comme résultat d'une épreuve où le hasard intervient effectivement. Par exemple, on notera les numéros sortis à la roulette de Monte-Carlo, moyen peu commode mais qui a fourni son nom à la méthode. Ainsi, le vendredi 24 février 1939, le 186ième coup et les suivants ont été: 21, 28, 20, 21, 35, 19, 7, 0, 6, 30, 2, etc.

Les nombres 0.21, 0.28, 0.20, 0.21, etc. sont alors des nombres compris dans l'intervalle $(0, 0.36)$ et la surveillance de l'État quant à la bonne marche de la roulette garantit l'uniformité de la distribution.

On peut encore prendre des numéros dans l'annuaire des téléphones, en utilisant sans la formuler l'hypothèse que l'ordre alphabétique garantit cette manière d'étalement uniforme des numéros que l'on cherche.

On s'est préoccupé cependant d'une fabrication en série, par des moyens purement mathématiques de ces nombres «au hasard» en utilisant des procédés itératifs, ce qui doit étonner de prime abord, car dans un tel procédé, un nombre est déduit du précédent, donc pas de hasard, au sens ordinaire, dans une telle suite; cependant, la distribution des nombres obtenus peut être uniforme, mais il conviendra de tester soigneusement ces suites.

Je prends deux exemples dus, semble-t-il, le premier à Von Neumann et le second à Lehmer; le lecteur pourra consulter à ce sujet l'article de Olga Taussky et John Todd¹.

¹. *Symposium on Monte-Carlo Methods*, John Wiley & Sons, New-York, 1957, p. 15.

Prenons un nombre de quatre chiffres, soit par exemple 4,032, que nous élevons au carré, ce qui donne 16,257,024; les quatre chiffres du milieu (2570) formeront le second nombre. On recommence ensuite l'opération: le carré de 2,570 est 06,604,900 (un 0 étant placé en première position pour obtenir un nombre de 8 chiffres), d'où le 3ième nombre, obtenu de la même façon que le précédent, sera donc 6049, etc. . . La suite obtenue: 0.4032, 0.2570, 0.6049, . . . etc. est censée constituer une suite de nombres dont la distribution est uniforme.

Ce genre de calcul itératif est obtenu simplement et rapidement à l'aide d'un calculateur électronique et c'est là son grand intérêt.

On peut être choqué du caractère apparemment arbitraire d'une telle méthode, mais, à bien y réfléchir, ce n'est pas plus choquant qu'une roulette et l'intérêt mathématique se retrouvera dans la critique des suites obtenues.

Deuxième procédé, que j'utilise personnellement sur l'ordinateur I.B.M. 1620 de l'École des Hautes Études commerciales. On utilise la formule de récurrence

$$x_{n+1} = kx_n (M).$$

Les données initiales sont: k , M et x_1 , x_1 étant un nombre de 8 chiffres par exemple. On multiplie x_1 par k et l'on retranche M autant de fois qu'il le faut pour obtenir un nouveau nombre de 8 chiffres (M étant bien entendu supérieur à 10^8); k et M sont choisis de manière à obtenir pour ces nombres une période la plus grande possible. On pourra prendre (valeurs proposées)

$$\begin{aligned} k &= 23 \\ M &= 10^8 + 1 \end{aligned}$$

Pour la production de nombres aléatoires distribués suivant une loi autre qu'uniforme, il existe une méthode générale et des méthodes particulières mieux adaptées à l'ouvrage des calculatrices mais qui n'ont pas reçu, que je sache, de théorie vraiment générale.

Soit $F(x)$ et $f(x)$, respectivement la fonction de répartition et la fonction de densité de la loi L considérée que nous supposons continue.

Y étant une variable aléatoire distribuée uniformément sur $(0, 1)$, on effectue le changement de variable $Y = F(X)$; la loi de probabilité élémentaire de Y étant dy , celle de X est donc la différentielle de y , soit $f(x)dx$.

X a donc pour densité $f(x)$: c'est une variable aléatoire distribuée suivant la loi L . Il en résulte qu'une suite de nombres distribués uniformément étant donnée, soit

$$y_1, y_2, \dots, y_i, \dots, y_n$$

on obtiendra une suite correspondante de nombres distribués suivant la loi L en résolvant, pour chaque indice i , l'équation

$$F(x_i) = y_i.$$

Par exemple, supposons que la loi L soit la loi exponentielle, dont on connaît l'importance dans les problèmes de files d'attente, on a ici

$$F(x) = 1 - e^{-cx},$$

l'équation précédente devient

$$1 - e^{-cx_i} = y_i$$

d'où

$$x_i = \frac{-1}{c} \text{Log}(1 - y_i)$$

Le problème de la fabrication d'une suite de nombres distribués suivant une loi L donnée est donc, théoriquement, résolu pourvu qu'on possède, initialement, des nombres répartis uniformément sur $(0, 1)$; on a vu plus haut comment on se procurait ceux-ci.

On pourrait dire que ces nombres répartis uniformément introduisent l'élément «hasard», la technique suivante n'étant que transformation ou éparpillement différent de ces nombres, ce qui laisse pressentir que dans la fabrication des nombres «au hasard» interviennent deux notions distinctes: celle de répartition, au sens du calcul des probabilités, et celle de hasard, au sens des séries chronologiques; j'exploiterai cette distinction dans la deuxième partie de cet article.

Donnons une autre méthode qui se prête mieux au calcul, due à Von Neumann, et concernant de nouveau la loi exponentielle¹.

1. *Symposium...*, *op. cit.*, article de J.-W. Butler, p. 249.

1) *Passage d'une distribution uniforme à une distribution de densité nx^{n-1} .*

Tout d'abord, prenons deux nombres distribués uniformément sur $(0, 1)$ et rejetons le plus petit; en recommençant cette opération un certain nombre de fois, on obtient une suite de nombres dont la distribution n'est évidemment plus uniforme puisqu'on a favorisé les plus grands. Plus précisément, soit X_1 et X_2 les deux nombres, et soit X le plus grand des deux.

$$\begin{aligned} \text{Prob} \{x < X < x + dx\} &= \text{Prob} \{X_1 < x\} \times \text{Prob} \{x < X_2 < x + dx\} \\ &+ \text{Prob} \{X_2 < x\} \times \text{Prob} \{x < X_1 < x + dx\} \\ &= \quad \quad \quad xdx + xdx \quad \quad = 2xdx. \end{aligned}$$

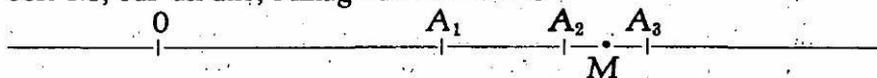
La densité est donc $2x$. De proche en proche, on montrerait que le plus grand des éléments d'un groupe de n nombres distribués uniformément sur $(0, 1)$ a pour densité nx^{n-1} .

Ce qui est une première façon, par simple choix, d'obtenir une distribution différente de la distribution initiale; notons que les calculatrices raffolent de ces classements dont elles s'acquittent à merveille.

2) *Distribution de densité $\frac{e^x}{e-1}$.*

Soit R uniformément réparti sur $(0, 1)$. La variable aléatoire $U = (e-1)R$ est alors uniformément répartie sur $(0, e-1)$ et sa densité est $\frac{1}{e-1}$.

Soit M , sur un axe, l'image du nombre U



et plaçons sur cet axe les points

A_1 d'abscisse 1

A_2 " $1 + \frac{1}{2!}$

A_3 " $1 + \frac{1}{2!} + \frac{1}{3!}$

$$A_i \quad \text{''} \quad \sum_{k=1}^i \frac{1}{k!}$$

Les points A_i ont un point limite dont l'abscisse est

$$\sum_{k=1}^{\infty} \frac{1}{k!} = e - 1.$$

Le point M se trouvera donc placé sur l'un des segments $A_{n-1} A_n$, et l'on fait apparaître un nouveau nombre aléatoire \mathcal{N} , qui est le rang du segment sur lequel se trouve M ; il est défini par la relation

$$\sum_{k=1}^{N-1} \frac{1}{k!} < U \leq \sum_{k=1}^N \frac{1}{k!}$$

La distribution de U étant uniforme, la probabilité pour que M appartienne au segment $A_{n-1} A_n$ est donc $\frac{1}{n!} \times \frac{1}{e-1}$, puisque $\frac{1}{n!}$ est la mesure de $A_{n-1} A_n$; c'est aussi la probabilité pour que $\mathcal{N} = n$.

Désignons alors par S le plus grand parmi \mathcal{N} nombres tirés au hasard sur $(0, 1)$, S et \mathcal{N} étant deux variables aléatoires. La loi conditionnelle de S liée par n est, d'après ce qui précède:

$$\text{Prob} \left\{ s < S < s + ds \mid \mathcal{N} = n \right\} = ns^{n-1} ds.$$

La loi du couple S et \mathcal{N} est, par suite,

$$\begin{aligned} \text{Prob} \{ s < S < s + ds, \mathcal{N} = n \} &= \frac{1}{e-1} \times \frac{1}{n!} ns^{n-1} ds \\ &= \frac{1}{e-1} \times \frac{s^{n-1}}{(n-1)!} ds \end{aligned}$$

La loi marginale de S est donc:

$$\frac{1}{e-1} \sum_{n=1}^{\infty} \frac{s^{n-1}}{(n-1)!} ds = \boxed{\frac{e^s ds}{e-1}}$$

Donc, pour obtenir une suite de nombres distribués suivant cette loi on peut:

- a) prendre un nombre r_1 au hasard sur $(0, 1)$;
- b) le multiplier par $e - 1$, ce qui donne u_1 ;
- c) placer u_1 dans la suite des nombres de la forme

$$\sum_{k=1}^{\infty} \frac{1}{k!}$$

- ce qui détermine le nombre n_1 ;
- d) choisir le plus grand parmi n_1 nombres pris au hasard sur $(0, 1)$: soit s_1 .

Puis on recommence. Toutes ces opérations s'effectuent simplement et sans erreurs dues aux calculs qui sont, ici, peu nombreux.

3) *Distribution de densité* $(e - 1) e^{-k}$ (k entier).

Soit X_i , une variable aléatoire distribuée uniformément sur $(0, 1)$, R_{i1} , R_{i2} , R_{i3} , etc., d'autres variables ayant la même distribution, prises dans cet ordre.

$$\begin{aligned} \text{Posons } Y_i &= R_{i1} + R_{i2} + \dots + R_{i(j-1)} + R_{ij} \\ \text{et } Z_i &= R_{i1} + R_{i2} + \dots + R_{i(j-1)} \end{aligned}$$

j est une valeur prise par l'entier aléatoire J , déterminé par les relations:

$$Z_i < X_i \leq Y_i$$

On forme la suite

$$Y_1, Y_2, \dots, Y_i, \dots, Y_k,$$

L'indice k est lui-même une valeur prise par l'entier aléatoire K déterminé par l'épreuve suivante: K est le rang du premier terme de la suite $\{Y_i\}$ formée au moyen d'un nombre impair de R_{ij} .

En d'autres termes:

$$Y_k = R_{k1} + R_{k2} + \dots + R_{kj} \quad (j \text{ est impair})$$

et, pour tout $i < k$

$$Y_i = R_{i1} + R_{i2} + \dots + R_{ij} \quad (j' \text{ est pair}).$$

On cherche à déterminer la loi de probabilité du nombre K .

Pour cela, considérons tout d'abord celle de J .

Supposons X_i fixé, égal à x .

Les deux événements:

- a) $J = 1$
- b) $R_{i1} \geq x$ sont équivalents,

on a donc:

$$\text{Prob} \left\{ J = 1 \mid \mathcal{X}_i = x \right\} = \text{Prob} \left\{ R_{ij} \geq x \right\} = \int_x^1 dr = 1 - x$$

De même, les deux événements

- a) $J = 2$
- b) $R_{i1} < x$
 $R_{i2} \geq x - R_{i1}$ sont équivalents,

d'où:

$$\text{Prob} \left\{ J = 2 \mid \mathcal{X}_i = x \right\} = \int_0^x dr_1 \int_{x-r_1}^1 dr_2 = x - \frac{x^2}{2}$$

D'une façon générale, les deux événements,

- a) $J = j$
- b) $Z_i < x$

$R_{ij} \geq x - Z_i$ sont équivalents; en remarquant que Z_i est la somme de $j - 1$ termes indépendants distribués uniformément sur $(0, 1)$, et que, d'autre part, Z_i doit être inférieur à x , donc à

1, on conclut que sa densité de probabilité est: $\frac{z^{j-2}}{(j-2)!}$.

Par suite:

$$\begin{aligned} \text{Prob} \left\{ J = j \mid \mathcal{X}_i = x \right\} &= \int_0^x \frac{z^{j-2}}{(j-2)!} \int_{x-z}^1 dr_i \\ &= \frac{1}{(j-2)!} \int_0^x z^{j-2} (1-x+z) dz \\ &= \frac{1}{(j-2)!} \left[(1-x) \frac{x^{j-1}}{j-1} + \frac{x^j}{j} \right] \\ &= \frac{1}{(j-2)!} \left[\frac{x^{j-1}}{j-1} - \frac{x^j}{j(j-1)} \right] \\ &= \frac{x^{j-1}}{(j-1)!} - \frac{x^j}{j!} \end{aligned}$$

$$\begin{aligned} \text{Donc: Prob} \left\{ J \text{ pair} \mid \mathcal{X}_i = x \right\} &= \sum_{j \text{ pair}} \left(\frac{x^{j-1}}{(j-1)!} - \frac{x^j}{j!} \right) \\ &= x - \frac{x^2}{2} + \frac{x^3}{3!} - \frac{x^4}{4!} + \dots \\ &= 1 - e^{-x} \end{aligned}$$

et, d'autre part,

$$\begin{aligned} \text{Prob} \left\{ J \text{ impair} \mid \mathcal{X}_i = x \right\} &= \sum_{j \text{ impair}} \left(\frac{x^{j-1}}{(j-1)!} - \frac{x^j}{j!} \right) \\ &= 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \dots \\ &= e^{-x} \end{aligned}$$

En remarquant que \mathcal{X}_i est une variable distribuée uniformément sur $(0, 1)$, donc, dont la loi de probabilité élémentaire est dx :

$$\text{Prob} \left\{ J \text{ pair} \right\} = \int_0^1 (1 - e^{-x}) dx = e^{-1}$$

$$\text{Prob} \left\{ J \text{ impair} \right\} = \int_0^1 e^{-x} dx = 1 - e^{-1}$$

On en déduit la loi de probabilité de K :

K est égal à k si les $k - 1$ premiers termes de la suite $\{\mathcal{Y}_i\}$ sont formés d'un nombre pair de termes, le $k^{\text{ième}}$ étant formé d'un nombre impair de termes.

D'où:

$$\begin{aligned} \text{Prob} \left\{ K = k \right\} &= (e^{-1})^{k-1} (1 - e^{-1}) \\ &= e^{-k} (e - 1) \end{aligned}$$

4) Distribution exponentielle.

Des résultats obtenus à la 2^e et 3^e étape, on déduit que la variable aléatoire:

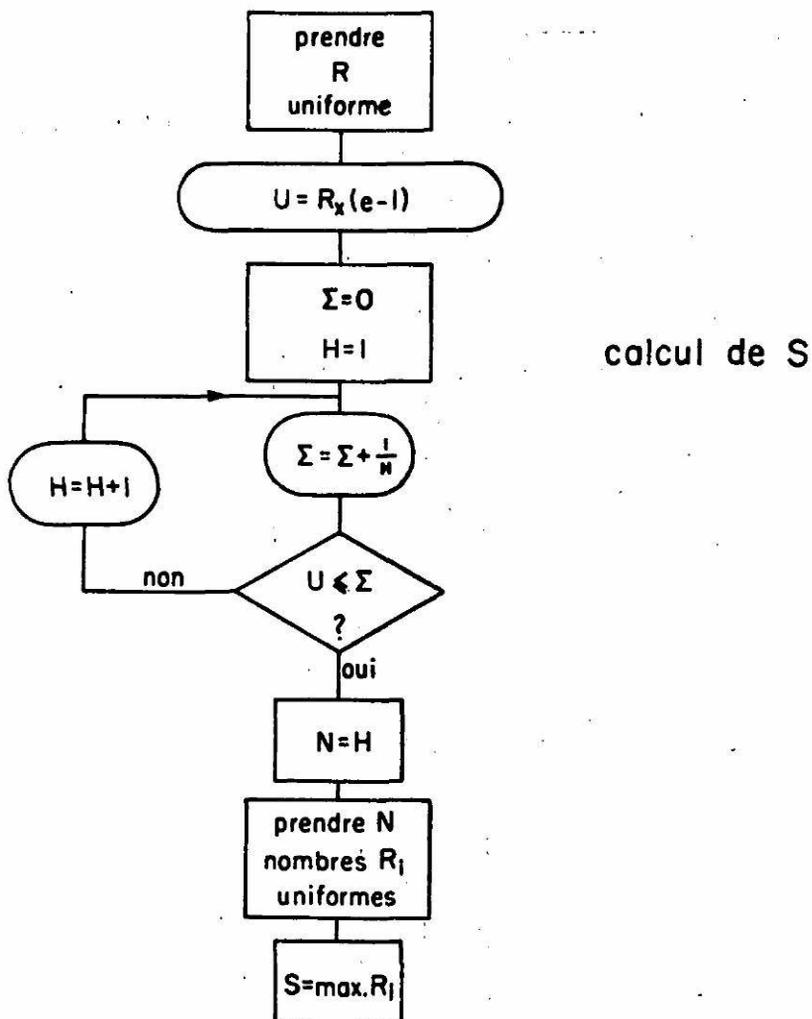
$$V = K - S$$

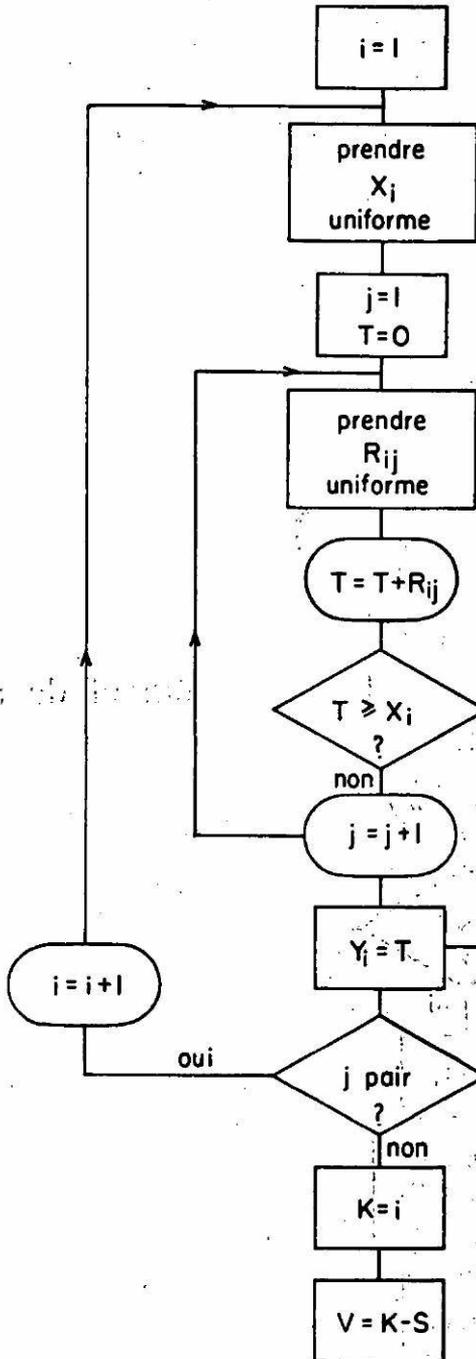
est distribuée suivant une loi exponentielle.

À PROPOS DE LA MÉTHODE DE MONTE-CARLO

On a, en effet, k étant entier et supérieur à v , $k - 1$ étant inférieur à v :

$$\begin{aligned} & \text{Prob} \{v < V < v + dv\} \\ &= \text{Prob} \{K = k\} \times \text{Prob} \{k - v < S < k - v + dv\} \\ &= (e - 1) e^{-k} \times \frac{e^{k-v}}{e - 1} dv \\ &= e^{-v} dv, v \text{ appartenant à l'intervalle } (0, + \infty). \end{aligned}$$





calcul de K

On en déduit un procédé très rapide pour obtenir une suite de nombres issus d'une population parente distribuée exponentiellement. L'organigramme ci-dessus suppose que l'on peut appeler à un instant quelconque des nombres extraits d'une population parente distribuée uniformément.

* * *

Lorsqu'on veut tester l'hypothèse suivante: une suite de nombres est issue d'une population parente distribuée uniformément, on est conduit classiquement à utiliser les lois d'échantillonnage: loi du χ^2 , du t de Student, etc. Ces lois, pour être appliquées, supposent le hasard dans l'obtention des nombres de la suite; elles ne testent pas ce hasard, mais seulement la distribution parente. On ne tient pas compte, alors, de l'ordre dans lequel apparaissent les nombres de la suite, mais de la fréquence de ces nombres dans un intervalle donné. Autrement dit, on analyse un échantillon et non pas une suite.

Or, si la roulette de Monte-Carlo trouvait, un beau jour, intelligent de donner les 37 nombres dont elle dispose selon l'ordre dans lequel ils se présentent: 0, 1, 2, 3 . . ., un test d'échantillon donnerait l'idée d'une distribution du genre uniforme; la roulette n'en serait pas moins détraquée et la suite obtenue manquerait de hasard, je veux dire, d'imprévu.

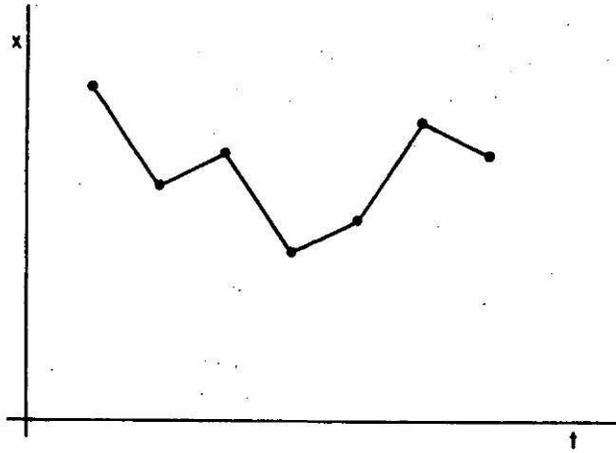
C'est bien un phénomène de ce genre que l'on doit craindre lorsqu'on utilise des itérations dans la fabrication de nombres «au hasard».

Ce problème a été étudié et une solution utilisée dans d'autres champs d'application par M. Kiveliiovitch et J. Vialar¹. Leur ouvrage s'intitule *Les séries chronologiques et la théorie du hasard* et son importance me paraît grande par sa façon originale de prendre le problème du hasard.

Les auteurs désignent sous le nom de série chronologique un échantillon ordonné: c'est le temps qui ordonne.

Ce qui est considéré, c'est la forme du graphique obtenue en portant en abscisse le temps et en ordonnée les valeurs des éléments de l'échantillon.

1. M. Kiveliiovitch et J. Vialar, *Les séries chronologiques et la théorie du hasard*, Publications scientifiques du Ministère de l'Air, Paris.



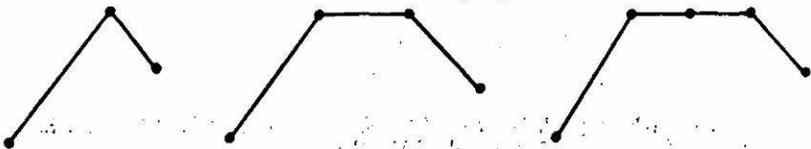
Un tel graphique présente des maxima relatifs, des minima, des phases croissantes et décroissantes.

La possibilité pour trois points donnés de former un maximum, dépend évidemment de l'ordre de ces trois points; le nombre de maxima dépendra donc de cet ordre. Réflexions analogues pour les minima et les phases.

On cherchera donc quelle est la fréquence théorique des maxima dans une suite « purement aléatoire », c'est-à-dire une suite $\{x_i\}$, échantillon ordonné de la variable aléatoire X , dans lequel la probabilité d'apparition de la valeur x_{i+1} est indépendante de la valeur x_i . Je pense qu'il faudrait ajouter que cette probabilité doit être indépendante des valeurs qui ont précédé x_i .

Probabilité d'un maximum. Donnons un exemple simple (*op. cit.* p.7). Nous plaçant dans le cas d'une suite purement aléatoire, supposons d'autre part que la variable aléatoire X prenne chacune des valeurs $0, 1, 2, \dots, n$ avec la probabilité $\frac{1}{n+1}$.

Un maximum peut être formé de trois points, ou 4, ou 5, etc.



À PROPOS DE LA MÉTHODE DE MONTE-CARLO

Dans le 1^{er} cas, supposons que ce maximum ait pour ordonnée i , et désignons par X_1, X_2, X_3 , les variables aléatoires correspondant aux trois points, ce qui donnera :

$\text{Prob} \left\{ X_1 = x_1, X_2 = i, X_3 = x_3 \right\} = \frac{1}{(n+1)^3}$ puisqu'il y a indépendance —

$$\text{Prob} \left\{ X_1 < i, X_2 = i, X_3 < i \right\} = \sum_{x_1=0}^{i-1} \sum_{x_3=0}^{i-1} \frac{1}{(n+1)^3} = \frac{i^2}{(n+1)^3}$$

c'est la probabilité d'un tel maximum.

De même, la probabilité d'un maximum égal à i , en 4 points, est égale à :

$$\frac{i^2}{(n+1)^4}, \text{ etc. } \dots$$

La probabilité d'un maximum égal à i est donc :

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{i^2}{(n+1)^{3+k}} &= \frac{i^2}{(n+1)^3} \sum_{k=0}^{\infty} \frac{1}{(n+1)^k} = \frac{i^2}{(n+1)^3} \times \frac{1}{1 - \frac{1}{n+1}} \\ &= \frac{i^2}{n(n+1)^2} \end{aligned}$$

Il en résulte que la probabilité d'un maximum égal à une valeur quelconque (0 est exclu) est :

$$\begin{aligned} \sum_{i=1}^n \frac{i^2}{n(n+1)^2} &= \frac{1}{n(n+1)^2} \sum_{i=1}^n i^2 \\ &= \frac{1}{n(n+1)^2} \times \frac{n(n+1)(2n+1)}{6} \\ &= \frac{1}{3} - \frac{1}{6(n+1)} = P_u \end{aligned}$$

Si N est alors la taille de l'échantillon (N doit être grand puisqu'on a supposé l'existence d'un maximum formé d'un nombre infini de points), on aura, disent nos auteurs (p. 17), un nombre théorique de maxima égal à :

$$\left(\frac{1}{3} - \frac{1}{6(n+1)} \right) (N-2) = P_u (N-2)$$

$N - 2$ étant le nombre de groupes de trois éléments successifs dans une suite de N éléments (il faut 3 éléments pour faire un maximum). À vrai dire, les maxima considérés n'ont pas tous trois éléments; cependant, une telle remarque ne peut amener de grands changements dans les résultats, car le nombre de maxima égaux à i et formés de trois points est:

$$\frac{i^2}{(n+1)^3}$$

et le nombre de maxima égaux à i est:

$$\frac{i^2}{n(n+1)^2}$$

nombre voisins dès que n est suffisamment grand.

Probabilité d'une phase croissante. Il est clair qu'à une unité près, le nombre total de phases, croissantes ou décroissantes, est égal au nombre d'extrema. De plus, on a autant de phases croissantes que de phases décroissantes.

On en déduit que la probabilité d'une phase croissante est également P_u .

Or, un calcul analogue au précédent permettrait de calculer la probabilité d'une phase croissante ne comprenant qu'un seul membre (cas d'un minimum et d'un maximum consécutifs); dans le cas présent, cette probabilité est

$$p_1 = \frac{5n^2 + 5n + 2}{24(n+1)^2}$$

La probabilité conditionnelle d'une telle phase, en prenant l'ensemble des phases croissantes comme ensemble fondamental, est donc:

$$\frac{p_1}{P_u} = q_1$$

Test des phases croissantes à un membre. Supposons donc que dans une suite obtenue on ait observé P_c phases croissantes au total et n_1 telles phases ne comprenant qu'un seul membre.

L'espérance mathématique de ce dernier nombre est: $q_1 P_c$

et son écart-type: $\sqrt{P_c q_1 (1 - q_1)}$

À PROPOS DE LA MÉTHODE DE MONTE-CARLO

Si P_c est suffisamment grand, on devra avoir, avec une probabilité de 0.95:

$$q_1 P_c - 2\sqrt{P_c q_1(1 - q_1)} < n_1 < q_1 P_c + 2\sqrt{P_c q_1(1 - q_1)}$$

relation qui fournit un test de l'hypothèse: «la suite proposée est purement aléatoire».

Plusieurs tests sont ainsi donnés dans l'ouvrage cité: test des phases à un, deux membres ou plus; test des maxima; test des indices de phase; test des valeurs égales; test des croisements.

Ces tests sont fondés sur un principe analogue à celui que j'ai présenté et utilisant une variable discrète quelconque ou une variable continue de fonction de répartition donnée.

* * *

Donnons quelques résultats d'application des tests précédents.

I — *Roulette*. On peut lire à la page 24 de l'ouvrage précédent plusieurs résultats de tests indiqués concernant la roulette de Monte-Carlo.

Par exemple, en ce qui concerne le test des phases croissantes à un membre, sur une suite de 571 coups, on a observé 125 telles phases.

De plus, la valeur observée du nombre de phases croissantes est $P_c = 192$. Enfin, $n = 37$.

D'où, avec les notations précédentes:

$$p_1 = \frac{5n^2 + 5n + 2}{24(n + 1)^2} = 0.2028$$

$$P_u = \frac{1}{3} - \frac{1}{6 \times 37} = 0.3288$$

d'où,
$$q_1 = \frac{p_1}{P_u} = 0.6166.$$

On en déduit:

$$104.9 < n_1 < 131.9$$

ce qu'on vérifie, puisque $n_1 = 125$.

La roulette réagit convenablement à ce test, ainsi qu'aux autres.

II — Nombres au hasard. Nous avons éprouvé, à l'aide de ces tests, les deux méthodes de fabrication de nombres au hasard exposés dans la première partie de cet article.

a) En ce qui concerne les nombres établis à l'aide de la formule de récurrence

$$x_{n+1} = kx_n (M)$$

en prenant

$$k = 23$$

$$M = 10^8 + 1.$$

Sur un échantillon de $\mathcal{N} = 200$ nombres,

on a obtenu $P = 124$ phases

réparties en: $n_1 = 72$ phases de 1 élément

$n_2 = 34$ phases de 2 éléments

$n_3 = 18$ phases de 3 éléments ou plus.

On doit, ici, avoir:

$$\frac{5}{8}P - 0.968 \sqrt{P} < n_1 < \frac{5}{8}P + 0.968 \sqrt{P}$$

$$\frac{11}{40}P - 0.893 \sqrt{P} < n_2 < \frac{11}{40}P + 0.893 \sqrt{P}$$

$$\frac{1}{10}P - 0.600 \sqrt{P} < n_3 < \frac{1}{10}P + 0.600 \sqrt{P}$$

d'où: n_1 doit appartenir à l'intervalle (67, 88)

n_2 " " " " (24, 44.2)

n_3 " " " " (5.8, 19)

ce qu'on vérifie.

Test des maxima ou minima:

il y a P phases observées,

donc, $P + 1 = 125$ extrema,

et, par suite, $\frac{P+1}{2}$ maxima, par exemple.

Leur fréquence relative est donc:

$$r = \frac{P+1}{2(\mathcal{N}-2)} = \frac{125}{396}$$

À PROPOS DE LA MÉTHODE DE MONTE-CARLO

Le test des maxima établit que cette valeur doit être comprise, la suite étant purement aléatoire, entre deux valeurs extrêmes :

$$\tau \text{ min.} = \frac{P + 1}{2(\mathcal{N}_{\text{max.}} - 2)}$$

et

$$\tau \text{ max.} = \frac{P + 1}{2(\mathcal{N}_{\text{min.}} - 2)}$$

avec

$$\mathcal{N}_{\text{min.}} = \frac{3}{2}P - 1.718 \sqrt{P} + n_0 + 1$$

$$\mathcal{N}_{\text{max.}} = \frac{3}{2}P + 1.718 \sqrt{P} + n_0 + 1$$

n_0 étant le nombre de membres de la suite situés avant et après les extrema, débutant ou finissant la suite, soit ici 0.

On en déduit :

$$\frac{125}{412} < \tau < \frac{125}{336}$$

ce qu'on vérifie.

La méthode simule donc le hasard de façon satisfaisante.

b) *Méthode des carrés.* J'ai utilisé des nombres de 8 chiffres, les nombres de 4 chiffres (exemple proposé dans l'article cité) ayant l'inconvénient suivant : la probabilité d'obtenir, par cette méthode, un nombre de 4 chiffres dont les deux derniers sont deux zéros est assez forte — tous les nombres suivants se terminent alors par deux zéros.

Un échantillon de $\mathcal{N} = 200$ nombres de 8 chiffres m'a donné :

$$P = 130 \text{ phases}$$

réparties en : $n_1 = 75$ phases de 1 élément

$$n_2 = 46 \quad \text{“} \quad \text{de 2 éléments}$$

$$n_3 = 9 \quad \text{“} \quad \text{de 3 éléments ou plus.}$$

Un calcul analogue au précédent montre que :

n_1 doit appartenir à l'intervalle (70.22, 92.28)

n_2 “ “ “ (25.57, 45.93)

n_3 “ “ “ (6.16, 19.84)

Le deuxième de ces deux tests conduirait à rejeter l'hypothèse: «la suite est purement aléatoire».

Ce résultat n'est pas confirmé par les deux autres, ni par le test des maxima qui nous donne:

$$\tau = 0.331$$

ce nombre devant appartenir à l'intervalle (0.3024, 0.3692).

Il reste que la suite considérée semble, au seuil de 5 p.c., présenter une légère organisation, en tant que série chronologique, alors que le test de χ^2 qui teste l'hypothèse: «la distribution est uniforme», donne un résultat satisfaisant.

Je dois ajouter que deux autres essais entrepris sur deux nouvelles suites, de deux cents nombres chacune, obtenus par le même procédé, n'ont pas confirmé ce résultat.

Concluons donc, non pas à la mauvaise qualité de la méthode de Von Neumann, mais à la nécessité, lorsque la chose est possible, de tester les suites de nombres «au hasard» que l'on utilise.

Claude TRICOT,
*professeur à l'École des
Hautes Études commerciales (Montréal)*