

Recherche opérationnelle - 5 Méthode de Monte-Carlo : problèmes d'utilisation

A. Haurie and C.-R. Duguay

Volume 40, Number 3, October–December 1964

URI: <https://id.erudit.org/iderudit/1002876ar>

DOI: <https://doi.org/10.7202/1002876ar>

[See table of contents](#)

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)

1710-3991 (digital)

[Explore this journal](#)

Cite this article

Haurie, A. & Duguay, C.-R. (1964). Recherche opérationnelle - 5 : méthode de Monte-Carlo : problèmes d'utilisation. *L'Actualité économique*, 40(3), 577–593. <https://doi.org/10.7202/1002876ar>

Analyse

Recherche opérationnelle - 5

Méthode de Monte-Carlo: problèmes d'utilisation

Les phénomènes aléatoires qui ne sont pas trop complexes peuvent être représentés analytiquement par un certain nombre de variables soumises à des lois de probabilité. Par exemple, une entreprise ayant k points de vente peut représenter la demande en chacun de ces points par k variables $X_1 \dots X_k$ et leurs lois de probabilité respectives $f_1(x) \dots f_k(x)$.

Un problème faisant intervenir un phénomène aléatoire peut être étudié de façon analytique, mais si la solution se révèle trop compliquée on tente d'éviter ces difficultés en procédant à une série d'observations de situations réelles ou en employant la méthode de Monte-Carlo. Souvent les conditions du problème ne permettent pas de disposer d'une série d'observations de situations réelles statistiquement suffisante. En fait, il est nécessaire de considérer un certain nombre de situations réelles pour estimer les paramètres des diverses lois de probabilité mais une bonne estimation n'en demande pas un trop grand nombre, tandis qu'une bonne estimation de la solution du problème en exige souvent beaucoup plus. La méthode de Monte-Carlo met à notre disposition un nombre aussi grand qu'on veut de situations simulées correspondant aux conditions du problème.

La façon d'obtenir des situations simulées a été exposée dans un article précédent de cette revue ¹.

1. Tricot, Claude, « À propos de la méthode de Monte-Carlo », *L'Actualité Économique*, oct.-déc. 1962.

Les premières applications de la méthode de Monte-Carlo à une simulation de phénomènes aléatoires ont été faites lors de l'étude de problèmes de diffusion en physique nucléaire ².

Les problèmes de recherche opérationnelle aussi font souvent intervenir des liaisons entre les variables représentant des phénomènes aléatoires beaucoup trop complexes pour qu'on puisse arriver à une solution par une méthode analytique. Par contre, les machines à calculer modernes rendent souvent les simulations possibles.

On peut toujours poser le problème de façon que l'ensemble de la méthode puisse être présenté selon ce schéma :

Graphique 1



- L'entrée représente la « décision » que l'on a à prendre. On a le choix entre un ensemble de décisions possibles. Le problème est de prendre la meilleure décision. Une fois la décision choisie on a une « valeur » d'entrée qu'on fournit au simulateur.
- Le simulateur donne aux variables représentant le phénomène aléatoire des valeurs correspondant à une réalisation vraisemblable du phénomène et calcule la « valeur » de sortie à partir de la « valeur » d'entrée.
- La sortie est donc aléatoire pour une entrée donnée. La comparaison entre les différentes valeurs de sortie pour des valeurs d'entrée différentes peut nous indiquer la meilleure décision. Le but est de trouver l'entrée qui donne la sortie optimale.

Il est commode d'avoir comme sortie une variable à une seule dimension pour pouvoir reconnaître aisément la valeur optimale.

2. Pour plus de détails sur l'historique de la méthode et l'application à la physique, voir *Symposium on Monte Carlo Methods*, John Wiley and Sons, N.Y., 1957.

On sait, par exemple, qu'on ne peut à la fois obtenir le maximum de satisfaction et la dépense minimale ; par contre, pour une dépense fixée on peut maximiser la satisfaction. En recherche opérationnelle, la sortie est souvent un coût ou un profit à optimiser.

L'entrée peut, par contre, être constituée d'une ou plusieurs variables, ou même d'une fonction comme dans le problème de redistribution de stock, traité par E.B. Berman³, que nous allons exposer ci-dessous.

Il s'agit d'un stock de pièces uniformes à répartir entre deux « bases ». Le niveau global du stock est fixe ; les pièces ne sont utilisées qu'une seule fois et leur vie utile est limitée. Cela pourrait être des pièces de rechange d'un moteur dont la conception devient désuète au bout d'un certain temps. On décompose cette durée de vie utile en périodes t_i ($i = 1, 2, \dots, n$), au début de chacune d'elles une décision de redistribution du stock est prise. Au départ le stock est réparti également entre les deux bases. Les phénomènes aléatoires sont ici les demandes en chaque base ; dans l'exemple utilisé par M. Berman elles sont poissonniennes.

Pour une période t , il y a dans chaque base une espérance mathématique du coût de pénurie, fonction de l'état du stock au début de cette période.

$$C_{t1}(S_1) = p_1 \sum_{x=S_1}^{x=\infty} (x - S_1) f_{t1}(x)$$

où :

- S_1 est l'état du stock dans la base 1.
- $C_{t1}(S_1)$ est l'espérance mathématique du coût de pénurie pour la base 1.
- p_1 est le coût unitaire de pénurie à la base 1. Il est constitué du coût d'attente de l'équipement en panne et du coût de transport d'urgence des pièces demandées.
- $f_{t1}(x)$ est la probabilité d'une demande de x objets dans la base 1 à la période t .

3. E.B. Berman, « Monte Carlo Determination of Stock Redistribution », *Operations Research*, juillet-août 1962.

Pour la base 2 on a une espérance mathématique du coût de pénurie :

$$C_{t_2}(S_2) = p_2 \sum_{x=S_2}^{x=\infty} (x - S_2) f_{t_2}(x)$$

avec des notations évidentes.

Si on prend une décision de redistribution, on transporte, par exemple, R objets de la base 2 à la base 1. On a, à la fin de la période, un « système de coût » constitué des coûts probables de pénurie en fonction des nouveaux stocks en chaque base et du coût du transport effectué.

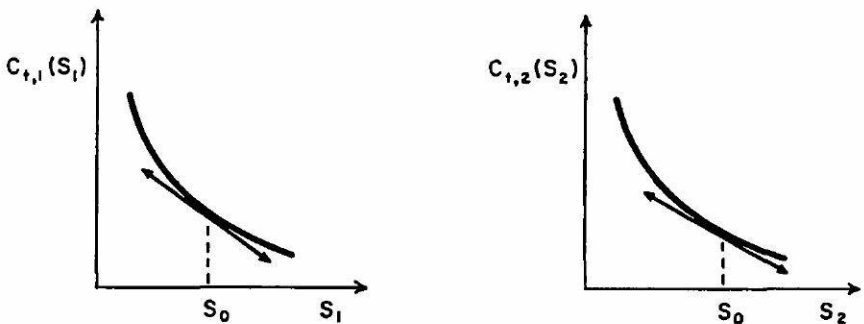
$$C_t(R_t) = C_{t_1}(S_1 + R) + C_{t_2}(S_2 - R) + \beta(t) k R + K$$

où :

- k est le coût unitaire de transport d'une base à l'autre.
- K représente les frais fixes de papeterie et d'administration dès qu'il y a un transfert.
- Enfin $\beta(t)$ est une fonction de t . Cette fonction $\beta(t)$ indique la proportion du coût de transport qu'on impute à la période t . Son rôle sera expliqué plus tard.

Détermination de la quantité à redistribuer. Représentons sur deux graphiques les variations de $C_{t_1}(S_1)$ avec S_1 , et celles de $C_{t_2}(S_2)$ avec S_2 .

Graphique 2



Si la pente de la 1^{ère} courbe pour $S_1 = S_0$ (état initial) est plus forte que celle de la 2^{ième} courbe pour $S_2 = S_0$, il peut être intéressant de diminuer le stock de la base 2 pour augmenter celui de la base 1, puisque le « coût espéré » de pénurie en 1 décroît plus rapidement pour une augmentation de stock que le « coût espéré » en 2.

Pour préciser, introduisons les quantités suivantes :

$$\Delta C_t(R) = \Delta C_{t1}(S_0 + R) + \Delta C_{t2}(S_0 - R) + \beta(t) k$$

avec :

$$\Delta C_{t1}(S_0 + R) = C_{t1}(S_0 + R) - C_{t1}(S_0 + R - 1)$$

$$\Delta C_{t2}(S_0 - R) = C_{t2}(S_0 - R) - C_{t2}[S_0 - (R - 1)].$$

D'après la forme des fonctions $C_{t1}(S_1)$ et $C_{t2}(S_2)$, on voit que :

$$\Delta C_{t1}(S_1 + R)$$

est négatif et que

$$\Delta C_{t2}(S_2 - R)$$

est positif.

Tant que

$$\Delta C_{t2}(S_2 - R) + \beta(t) k,$$

qui représente le coût de la distribution d'une unité supplémentaire, est inférieur à la valeur absolue de

$$\Delta C_{t1}(S_1 + R),$$

qui représente le gain espéré du fait de cette redistribution d'une unité, on aura intérêt à réaliser ce transfert.

On peut donc dire que le « coût marginal espéré » de la redistribution est la somme du coût marginal de pénurie en 3 et de la proportion du coût de transport d'une unité affectée à la période. Le « profit marginal espéré » est la diminution du « coût espéré » de pénurie de la base 1.

Tant que le profit est supérieur au coût ($\Delta C_t(R) < 0$), on a intérêt à poursuivre la redistribution. La décision de redistribution sera R^* , plus grande valeur entière que prend R , telle que $\Delta C_t(R)$ soit négatif.

Signification de la fonction $\beta(t)$. Les décisions de redistribution sont prises en utilisant un modèle faisant intervenir les coûts de pénurie calculés sur une période. On a ainsi une décision de redistribution à chaque période. Cependant si on effectue une redis-

tribution, celle-ci devra tenir compte des pénuries possibles dans les périodes suivantes. Si la redistribution pour la période t est calculée en ne tenant compte que des situations qui peuvent s'y produire, c'est-à-dire en considérant la fin de la période comme horizon, on peut décider de redistribuer R^* unités (en les transportant de 2 vers 1, par exemple). Comme on l'a vu, la rentabilité d'une redistribution d'une unité supplémentaire détermine R^* . Mais en procédant ainsi il y a une certaine tendance à ce que, lors de la période suivante, il soit opportun de faire une redistribution encore dans le même sens.

On peut alors tenir compte dès la période t des tendances pour les périodes futures et faire une redistribution qui influera sur plusieurs périodes. On affecte pour cela le coût unitaire de transport d'un coefficient $\beta(t)$, dépendant de la période, et représentant, selon M. Berman : « La valeur de la redistribution présente pour les périodes suivantes ».

Pour préciser la signification de $\beta(t)$, supposons que l'on affecte à une période la totalité des coûts de transports [$\beta(t) = 1$]. La décision de redistribution R^* sera telle que

$$0 = \Delta C_1 (\bar{S}_1 + R^*, t) + \Delta C_2 (\bar{S}_2 - R^*, t) + k$$

qu'on écrit :

$$\Delta C' = k$$

où $\Delta C'$ est cependant supérieur à k d'après la détermination de R

$$\Delta C' = - [\Delta C_1 (\bar{S}_1 + R^*, t) + \Delta C_2 (\bar{S}_2 - R^*, t)]$$

cette quantité $\Delta C'$ peut être explicitée :

$$\begin{aligned} \Delta C' = & C_1 [S_1 + (R^* - 1), t] - C_1 [S_1 + R^*, t] \\ & + C_2 [S_2 - (R^* - 1), t] - C_2 [S_2 - R^*, t] \end{aligned}$$

Elle est positive. Elle exprime que le « profit » marginal effectué en augmentant le stock 1 de une unité est plus important que le coût marginal dû à la perte d'une unité pour le stock 2. $\Delta C'$ exprime le déséquilibre que laisse entre les deux stocks, la redistribution R^* . Du fait de ce déséquilibre il suffira d'une faible modification des conditions (de demande par exemple) pour les périodes suivantes pour qu'une distribution « dans le même sens » soit encore souhaitable. Une distribution de 2 vers 1 est donc privilégiée pour les

périodes suivantes : on peut corriger cette situation en distribuant plus d'objets dès maintenant.

Mais alors une partie du coût de transport ne devra pas être comptée dans ce système de coûts relatif à la période puisque la redistribution tient compte d'éventualités futures. On affecte alors au coût k de transport un coefficient $\beta(t)$ dépendant de la période et inférieur à 1. On a alors :

$$\Delta C' = \beta(t) k.$$

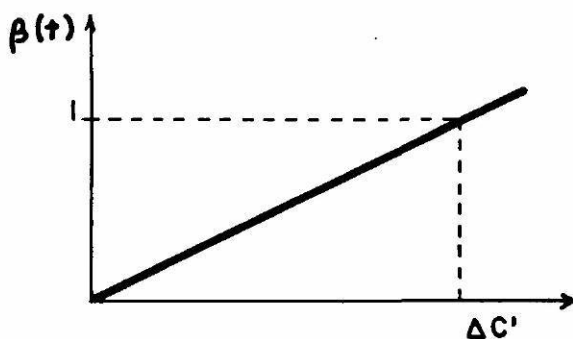
La valeur de $\beta(t)$ est difficile à déterminer. On voit par la relation précédente qu'un choix de $\beta(t)$ très faible entraîne un faible déséquilibre entre les états des stocks, donc des redistributions pouvant se faire aussi bien de 1 vers 2 que de 2 vers 1, ce qui infirme le choix de $\beta(t)$ faible qui voudrait exprimer qu'une grande partie du coût de transport est affectée aux périodes qui suivent : donc qu'une redistribution de même sens est très probable.

On peut imaginer deux $\beta(t)$. Un que l'on choisit arbitrairement et qui entraîne un déséquilibre $\Delta C'$ à la fin de la période :

$$\Delta C' = \beta(t) k$$

Donc on a un graphe, où $\beta(t)$ est arbitraire.

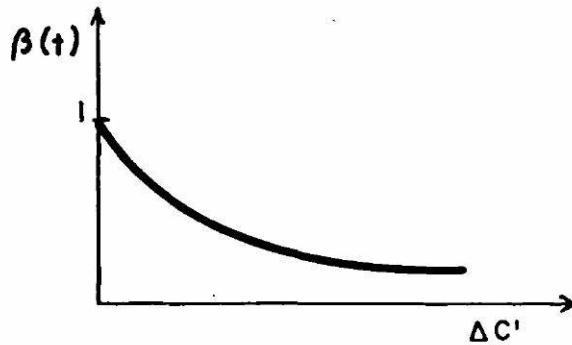
Graphique 3



Un autre qui est déterminé justement par l'état de déséquilibre qui subsiste entre les deux stocks et qui est d'autant plus voisin de 1

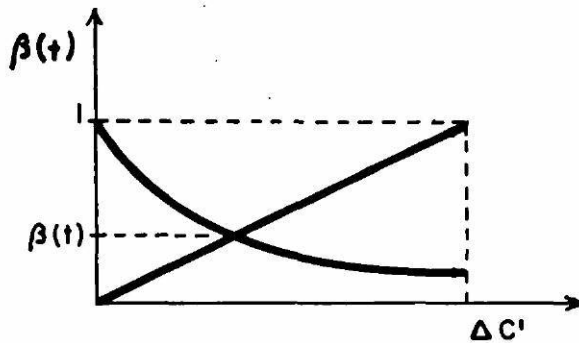
que $\Delta C'$ est faible. M. Berman l'appelle $\beta(t)$ « implicite ». On a le graphe :

Graphique 4



Les deux liaisons doivent être satisfaites simultanément, la valeur de $\beta(t)$ à trouver est celle qui est l'ordonnée de l'intersection de ces deux courbes.

Graphique 5



Le coefficient $\beta(t)$ tient compte des périodes à venir ; le nombre de ces périodes diminue quand t croît ; il est naturel, d'après la signification « implicite » de $\beta(t)$, de supposer que la fonction

$\beta(t)$ tend vers 1 quand on se rapproche de la dernière période, elle vaut 1 pour la dernière période car alors la fin de la période est aussi la fin de la durée de vie considérée.

Détermination de $\beta(t)$. Si on est convaincu de la nécessité de la fonction $\beta(t)$, il paraît difficile de l'évaluer analytiquement. Par contre, si on prend comme hypothèse que cette fonction est celle qui minimise les coûts totaux à la fin de la « durée de vie » considérée, une détermination par la méthode de Monte-Carlo est possible.

Si on choisit une fonction $\beta(t)$, comme on connaît les lois de probabilité des demandes pour les diverses périodes, on peut décider, selon la règle de décision exposée, des redistributions à faire. Si cette fonction a été mal choisie, elle donne une redistribution soit trop forte soit trop faible. Ces deux éventualités peuvent entraîner des coûts élevés (les transports trop importants ou les coûts dus à des pénuries systématiques dans une des deux bases).

Si on dispose d'un simulateur donnant des demandes vraisemblables suivant les périodes, on peut simuler un déroulement complet des opérations. Pour un choix de la fonction $\beta(t)$, la règle de décision est appliquée ; après réalisation des demandes pour toute la durée de vie on calcule le coût total de toutes les opérations *ex post*. Ce coût est le coût effectif pour une réalisation ; il sera donc calculé à partir des pénuries effectives et non « espérées », et à partir de tous les frais de transport $kR^* + K$ et non de la proportion $\beta(t)k$ qui intervient dans la règle de décision.

Avec de nombreuses simulations on peut essayer diverses formes de fonction $\beta(t)$, en retenant celle qui donne un « coût total moyen *ex post* » minimal. Ces diverses fonctions devront cependant toutes être croissantes et tendre vers 1 quand t tend vers la fin de la durée de vie.

Remarques. Il peut sembler arbitraire d'avoir choisi le coefficient $\beta(t)$ inférieur à 1. S'il a pour unique rôle de tenir compte des périodes suivant celles où la décision est prise, rien n'implique qu'il soit inférieur à 1. Seule une certaine stabilité que l'on suppose lorsqu'on parle de « tendances à avoir des redistributions de même sens dans les périodes suivantes » justifie cela.

L'utilité de ce facteur est montrée par l'application de la méthode de Monte-Carlo. Si ce $\beta(t)$ se révélait inutile, les simulations successives donneraient des coûts *ex post* minimums lors du choix de $\beta(t) = 1$, quel que soit t .

Lors d'une simulation on dispose des valeurs d'entrée, mais la sortie est aléatoire. Dans l'exemple précédent, on cherche à minimiser la valeur moyenne de la sortie. Celle-ci est une estimation de l'espérance mathématique de la variable de sortie, liée par une valeur donnée de l'entrée.

L'entrée est souvent une variable à plusieurs dimensions (X_1, X_2, \dots, X_k). Ainsi, dans l'exemple précédent on peut fixer la forme générale de $\beta(t)$ et n'avoir à déterminer qu'une suite de coefficients.

L'espérance mathématique de la valeur de sortie Y est liée strictement à l'entrée :

$$E(Y / X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = f(x_1, x_2, \dots, x_k)$$

ceci constitue par définition la régression⁴ de Y par rapport à X_1, X_2, \dots, X_k .

Pour une entrée fixée, il faudra faire un assez grand nombre de simulations pour avoir une bonne estimation de l'espérance mathématique liée de la variable de sortie. Pour obtenir une bonne précision, il est aussi nécessaire de considérer des valeurs d'entrée prises en grand nombre dans un vaste domaine de variation. Ces opérations étant longues et coûteuses, on cherche des techniques réduisant le nombre de simulations nécessaires à une bonne précision dans la détermination des valeurs d'entrée optimales.

Une de ces méthodes est d'estimer en une première approximation la fonction de régression au moyen de quelques valeurs (x_1, x_2, \dots, x_k) prises dans un vaste domaine de variation. On fait alors une estimation plus précise au moyen d'autres valeurs (x_1, x_2, \dots, x_k) située dans un domaine plus restreint entourant la valeur optimale de la fonction de régression obtenue en première approximation. On répète ainsi de suite l'opération jusqu'au degré de précision voulu. À chacune de ces étapes on a à ajuster une courbe de régression à l'échantillon obtenu. La forme la plus simple que l'on puisse imposer à la fonction de régression est une *forme* linéaire,

4. Pour les questions de régression, voir H. Cramer, *Mathematical Methods of Statistics*, chapitres 21, 22, 23.

mais cette linéarité n'est justifiée que dans le cas d'une population distribuée suivant une loi normale ; dans les autres cas elle entraîne des erreurs systématiques importantes en plus des erreurs aléatoires ou résidus. Une forme de fonction pouvant mieux s'ajuster à l'échantillon entraîne des difficultés de détermination.

J.F. Walsh propose un compromis⁵. Il utilise une fonction implicite en $x_1, x_2, \dots, x_k, \bar{y}$ où :

$$\bar{y} = E(Y \mid X_1 = x_1 \dots X_k = x_k)$$

Cette fonction aura la forme :

$$\bar{y} + A_1 g_1(\bar{y}) + A_2 g_2(\bar{y}) + \dots + A_s g_s(\bar{y}) = \\ A_{s+1} + A_{s+2} g_{s+2}(x_1 \dots x_k) + \dots + A_t g_t(x_1 \dots x_k)$$

Les fonctions g_1, g_2, \dots, g_t sont toutes complètement précisées. On impose au premier membre :

$$\bar{y} + A_1 g_1(\bar{y}) + \dots + A_s g_s(\bar{y})$$

d'être une fonction monotone de y de façon qu'à une valeur de x_1, x_2, \dots, x_k il ne corresponde qu'une seule valeur de \bar{y} .

Aucune des fonctions g_1, \dots, g_t n'est linéaire. Elles sont choisies pour avoir un bon ajustement et des calculs simples.

Les inconnues sont les coefficients A_1, \dots, A_t qui interviennent de façon linéaire. Ces coefficients ne sont pas tous indépendants puisque, comme on l'a vu, le premier membre doit être fonction monotone de \bar{y} . Cette monotonie entraîne qu'un certain nombre $r - 1$ des coefficients A_1, \dots, A_s sont liés, tous les autres coefficients étant alors indépendants.

La linéarité par rapport aux coefficients permet une certaine simplicité dans leur estimation, tandis que la non-linéarité des fonctions g_1, \dots, g_t et la forme implicite de la fonction de régression entraînent une grande souplesse d'emploi.

Deux méthodes d'estimation des coefficients sont proposées par J.E. Walsh. Elles ne s'appliquent qu'aux coefficients indépendants. L'une est un ajustement par les moindres carrés, l'autre est une méthode probabiliste.

L'ajustement par les *moindres carrés* est simple. Il faut d'abord éliminer les $r - 1$ coefficients dépendants. On sait que pour éli-

5. J.E. Walsh, « Use of Linearized Nonlinear Regression for Simulations Involving Monte Carlo », *Operations Research*, mars 1963, pp. 228-235.

miner $r - 1$ inconnues, il suffit de disposer de r équations où elles interviennent. On obtiendra r relations indépendantes entre tous les coefficients $A_1 \dots A_t$ en faisant correspondre à r valeurs d'entrée $(x_1, x_2 \dots x_k)$ les moyennes correspondantes de la valeur de sortie. Ces moyennes sont des estimations de \bar{y} , l'espérance mathématique conditionnelle de Y quand $X_1 = x_1 \dots X_k = x_k$. On note encore ces moyennes \bar{y} . En portant ces valeurs dans l'équation en $x_1 \dots x_k$, \bar{y} définissant la fonction de régression on a r relations entre $A_1 \dots A_t$.

Si on dispose de n relations, $n > k$, obtenues de cette façon, entre $A_1 \dots A_t$ on peut constituer au moins n groupes de r relations entre lesquelles on élimine les $r - 1$ coefficients dépendants que l'on suppose être : $A_1 \dots A_{r-1}$.

On a alors n relations entre A_r, \dots, A_t :

$$K_{0i} + K_{ri} A_r + \dots + K_{si} A_s = K_{(s+1)i} A_{s+1} + \dots + K_{ti} A_t$$

où $i = 1$ à n . Les K_{ji} sont calculés à partir des valeurs d'entrée et des moyennes des valeurs de sortie ; ce sont donc des quantités bien déterminées. Les moindres carrés sont utilisés pour l'estimation de

$$A_r, A_{r+1}, \dots A_t.$$

Les estimations sont les valeurs qui minimisent :

$$\sum_{i=1}^{i=n} (K_{0i} + K_{ri} A_r + \dots + K_{si} A_s - K_{(s+1)i} A_{s+1} \dots - K_{ti} A_t)^2$$

Le défaut de cette méthode est de ne pas fournir d'indication sur la validité des résultats obtenus. Les estimations de $A_r \dots A_t$ sont aléatoires, elles diffèrent de la valeur vraie de ces coefficients d'une quantité aléatoire due au fait que lors d'une simulation la valeur de sortie est aléatoire et que la moyenne \bar{y} des valeurs de sortie correspondant à une même valeur d'entrée $(x_1 \dots x_k)$ est elle-même aléatoire et n'est qu'une estimation de l'espérance mathématique $E(Y | X_1 = x_1, \dots X_k = x_k)$ qui intervient dans la régression. Il peut donc se faire que les résultats obtenus, c'est-à-dire les estimations de A_r, \dots, A_t , soient très mauvaises ; il serait intéressant de savoir quelle confiance accorder à ces résultats.

La méthode probabiliste se propose d'établir une estimation des coefficients et des intervalles de confiance, c'est-à-dire de calculer les probabilités que la vraie valeur du coefficient soit située entre deux valeurs calculées à partir de l'échantillon. (L'échantillon consiste ici en un ensemble d'estimations du coefficient.)

Dans la première méthode, on utilise r relations entre $A_1 \dots A_t$ pour éliminer $A_1 \dots A_{r-1}$; mais on pourrait obtenir une estimation de A_1, \dots, A_t par résolution de t équations en $A_1 \dots A_t$.

Ainsi, en séparant les n relations obtenues entre $A_1 \dots A_t$, en groupes de t relations entre $A_1 \dots A_t$, on a à résoudre des systèmes linéaires d'ordre t . Pour un n assez grand on obtient aussi un certain nombre d'estimations de A_v ($v = 1 \dots t$).

Les manipulations sur les estimations qui suivent peuvent paraître arbitraires, mais elles auront pour but de donner une valeur approchée de la valeur médiane des estimations et de permettre surtout l'établissement d'intervalles de confiance.

On groupe les estimations de A_v en U classes qui contiennent à peu près le même nombre d'éléments. Cette répartition se fait au hasard pour éviter des erreurs systématiques. Dans chaque classe u , $L(u; v)$ est la moyenne arithmétique des estimations de A_v qui y figurent. On a donc :

$$L(u; v) = A_v + e(u; v) \quad \begin{array}{l} v = 1 \dots t \\ u = 1 \dots U \end{array}$$

où A_v est la valeur exacte cherchée. $e(u; v)$ est une erreur aléatoire qui dépend de l'indice v du coefficient et aussi de l'indice u de la classe considérée. Si :

$$p(u; v) = P[L(u; v) < A_v]$$

la moyenne des $p(u; v)$ pour les différentes valeurs u est voisine de $1/2$, car il n'y a pas de raison que les $L(u; v)$ deviennent toujours trop fortement dans un sens de A_v .

Le nombre U de classes est assez faible mais supérieur à 5 et est impair pour un calcul pratique de la médiane.

On range les $L(u; v)$ par valeurs croissantes :

$$L_v[1] < L_v[2] < L_v[3] \dots < L_v[U]$$

La valeur $L_v \left[\frac{U+1}{2} \right]$ est une approximation de la valeur médiane des estimations de A_v .

Mais, comme on a dit plus haut, la création de ces variables aléatoires $L(u; v)$ est surtout utile par la définition d'intervalles de confiance.

On va chercher à rendre minimum :

$$P \{ L_v[u_1] \leq A_v < L_v[u_2] \}$$

où :

$$u_1 < \frac{U}{2} \quad u_2 > \frac{U}{2}$$

on sait que pour tout u

$$p(u; v) = P L(u; v) < A_v$$

en moyenne est voisin de $\frac{1}{2}$.

On a réalisation de l'événement

$$L_v [u_1] \leq A_v < L_v [u_2]$$

si on a :

$$L_v [1] < A_v$$

$$L_v [2] < A_v$$

.

.

.

$$L_v [u_1] \leq A_v$$

$$L_v [u_2 + 1] > A_v$$

.

.

$$L_v [U] > A_v$$

ou bien :

$$L_v [1] < A_v$$

$$L_v [2] < A_v$$

.

.

.

$$L_v [u_1] < A_v$$

$$L_v [u_1 + 1] \leq A_v$$

$$L_v [u_1 + 2] > A_v$$

.

.

$$L_v [U] > A_v$$

.

.

.

jusqu'à :

$$L_v [1] < A_v$$

$$L_v [2] < A_v$$

.

.

.

$$L_v [u_2] > A_v$$

.

.

.

.

$$L_v [U] > A_v$$

Or chacun des événements a une probabilité voisine de :

$$C_U^u \left(\frac{1}{2}\right)^U$$

où u prend les valeurs $u_1, u_1 + 1 \dots u_2 - 1$. En effet, il suffit que u_1 classes parmi les U donnent des $L(u_1, v)$ inférieurs ou égaux à A_v pour qu'il y ait une réalisation du premier type.

Ceci peut se produire de $C_U^{u_1}$ façons différentes, soit le nombre de combinaisons de U classes, u_1 à u_1 .

La probabilité d'avoir $L_v [1] < A_v$ etc., est à chaque fois voisine de $\frac{1}{2}$. D'autre part, chaque inégalité constitue un événement indépendant.

La probabilité d'avoir les U inégalités marquées est :

$$p(1; v) \cdot p(2; v) \cdot \dots \cdot p(U; v) = \left(\frac{1}{2}\right)^U$$

puisque les $p(u; v)$ sont voisins de $\frac{1}{2}$ et

$$\sum_u p(u, v) = \frac{U}{2}$$

On a donc :

$$P \{L_v [u_1] < A_v < L_v [u_2]\} \approx \left(\frac{1}{2}\right)^U \sum_{u=u_1}^{u_2-1} C_U^u$$

Ce qui permet pour une probabilité donnée de trouver l'intervalle dans lequel A_v a cette probabilité de se trouver.

En réalité on peut montrer que :

$$P \{L_v [u_1] < A_v < L_v [u_2]\} \geq \left(\frac{1}{2}\right)^U \sum_{u=u_1}^{u_2-1} C_U^u$$

si les $p(1, v)$, $p(2, v)$..., $p(U, v)$ sont tels que leur moyenne vaille $\frac{1}{2}$.

On peut prendre pour A_v la valeur médiane $L_v \frac{U+1}{2}$ obtenue plus haut, l'inégalité ci-dessus permettant d'évaluer la probabilité d'erreur.

Une autre méthode pour réduire le nombre de simulations nécessaires lors de l'usage d'une méthode de Monte-Carlo, est l'échantillonnage d'importance. Nous allons expliquer rapidement cette méthode, analogue à l'échantillonnage stratifié, en nous inspirant de l'exposé clair et simple de M. Charles Clark ⁶.

Dans la méthode de Monte-Carlo, la sortie est une variable aléatoire dont on veut généralement estimer un caractère à partir d'un échantillon de taille n obtenu au moyen de n simulations. L'échantillonnage simple consiste à tirer au hasard n éléments de la population et à prendre pour estimation le caractère moyen calculé à partir de l'échantillon. L'échantillonnage d'importance consiste aussi à tirer au hasard n éléments mais d'une population différente et telle que la précision de l'estimation sera plus grande pour une taille de l'échantillon tiré.

Reprenons l'exemple simple de M. Clark pour illustrer la méthode. Soit une population définie par sa fonction de densité

$$p(x) = 0.01 \exp(-0.01)x \quad 0 < x < \infty.$$

6. « Importance Sampling in Monte Carlo Analyses », *Operations Research*, septembre-octobre 1961, pp. 603-620.

Nous voulons connaître la probabilité P qu'à une valeur tirée de cette population d'être inférieure à 1. On peut la calculer facilement à partir de la densité. On a :

$$P = \int_0^1 (x) dx = 1 - e^{-0.01} = 0.00995.$$

Il en est ainsi à cause de la simplicité de la fonction de densité. Supposons qu'on veuille estimer cette probabilité P au moyen de la méthode de Monte-Carlo : on produirait un échantillon de taille n , et P serait le rapport du nombre des valeurs inférieures à 1 au nombre total N . Étant donné la faible probabilité d'avoir $x < 1$, il faudrait prendre n très grand pour obtenir un minimum de précision.

Pour obtenir une plus forte proportion des valeurs tirées comprises dans l'intervalle $(0, 1)$, nous allons modifier le mode d'échantillonnage. Nous allons tirer au hasard des éléments à partir d'une nouvelle population dont la fonction de densité est :

$$p'(x) = \exp - x \quad 0 < x < \infty.$$

Cette population n'a aucun rapport avec le problème, mais elle permet d'obtenir un échantillon où le nombre relatif de valeurs appartenant à l'intervalle « important » pour nous $(0, 1)$ est plus élevé.

L'espérance mathématique de ce nombre doit être $P' = \int_0^1 e^{-x} dx = 0.63$. Le résultat trouvé à partir de ce nouvel échantillon est biaisé, mais nous pouvons corriger ce biais au moyen du rapport des probabilités :

$$\omega(x) = \frac{p(x)}{p'(x)}$$

Pour chacune des valeurs, on se trouve ainsi à pondérer les résultats de l'échantillon. Par exemple, la probabilité d'avoir $X = 2$ est $p(2) = 0.009802$ dans la première population et $p'(2) = 0.13534$ dans la deuxième population, d'où :

$$\omega(x) = \frac{p(2)}{p'(2)} = 0.07.$$

Ceci signifie que dans un échantillon tiré de la première population, la valeur 2 est rencontrée en moyenne 14.2 fois moins souvent que dans un échantillon issu de la deuxième population. Si nous voulons nous faire une idée de la première population à partir d'un échan-

tillon de la deuxième nous devons donc multiplier les résultats observés par le rapport $\frac{p}{p'}$, c'est-à-dire pondérer les résultats de façon à corriger le biais introduit par le mode d'échantillonnage.

En résumé, en tirant un échantillon de la population de densité $p(x)$, nous aurions eu environ une valeur sur 100 inférieure à 1. En tirant de la seconde population, nous en obtenons 63 sur 100, ce qui rend les fluctuations moins dangereuses. Cependant, ce résultat doit être rectifié au moyen du rapport des probabilités $\omega(x)$ pour le rendre utile dans l'étude de la première population. Mais pour calculer ce facteur $\omega(x)$, il faut connaître $p(x)$ et $p'(x)$, or on cherche à évaluer $p(x)$ par la méthode de Monte-Carlo justement parce qu'on ne peut le faire plus simplement par le calcul direct⁷.

D'autre part, dans un processus aléatoire où plusieurs lois de probabilités se combinent, le choix des populations auxiliaires devient beaucoup plus délicat et peut entraîner des erreurs grossières. Il est bon de montrer une grande prudence quand le problème a un peu d'ampleur, car il n'y a pas de règle fixe permettant de trouver les populations donnant la meilleure précision. Chaque problème est un cas particulier.

A. HAURIE,
professeur à
l'École des Hautes Études commerciales
(Montréal).

et
 C.-R. DUGUAY,
licencié en sciences commerciales
(Montréal).

7. Le cercle vicieux que l'on pourrait voir est dû en partie au fait que la méthode est exposée sur un exemple volontairement trop simple. Il n'en reste pas moins qu'il y a ici une difficulté.