JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

G. MATHERON

Présentation des variables régionalisées

Journal de la société statistique de Paris, tome 107 (1966), p. 263-275
<http://www.numdam.org/item?id=JSFS_1966__107__263_0>

© Société de statistique de Paris, 1966, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

PRÉSENTATION DES VARIABLES RÉGIONALISÉES [1]

VARIABLES RÉGIONALISÉES

Un grand nombre de phénomènes naturels se présentent à l'homme sous forme régionalisée : ils se déploient, ou se distribuent dans l'espace. De tels phénomènes peuvent se caractériser, localement, par certaines grandeurs qui varient dans l'espace, et constituent, par conséquent, des fonctions numériques (ordinaires). Ce sont de telles fonctions numériques que nous appelons des variables régionalisées : il s'agit là d'un terme neutre, purement descriptif, antérieur, en particulier à toute interprétation probabiliste.

1. Conférence prononcée le 15 juin 1966 devant la Société de Statistique de Paris. Cette conférence est une présentation de la thèse de l'auteur : *Les Variables régionalisées et leur estimation.* — Masson, Paris 1965, soutenue sous la direction de M. le Professeur R. FORTET et sous la présidence de M. le Professeur L. SCHWARTZ.

Comme exemples simples de variables régionalisées, citons entre autres :

— la teneur, dans un gisement minier,

— le rendement à l'hectare, dans une région céréalière,

— la densité de population, à l'intérieur d'un pays.

Dans ce qui suit, j'emprunterai surtout mes exemples aux sciences de la terre et à l'estimation des gisements miniers, domaines où j'ai surtout travaillé, mais le champ d'application est évidemment plus général.

Caractère mixte aléatoire — Structuré

Le plus souvent, les variables régionalisées présentent un haut degré d'irrégularité. Elles ne sont généralement pas dérivables, ni même continues. On ne peut les représenter graphiquement que d'une manière grossière et approximative, sous forme de courbes en « dents de scie ». L'étude directe de courbes ce genre serait prohibitive, et ne présenterait sans doute qu'un intérêt limité, en raison de leur complication même. Il y a dans le comportement des variables régionalisées un *aspect aléatoire*, qui suggère presque irrésistiblement le recours à une interprétation probabiliste.

Cet aspect aléatoire permet de comprendre l'insuffisance des méthodes traditionnelles d'estimation des gisements miniers. Les mineurs admettent que chaque échantillon prélevé représente la teneur réelle de sa zone d'influence. Ils obtiennent bien ainsi une estimation globale utilisable, mais ne peuvent évidemment pas l'assortir d'un calcul d'erreur (d'une variance d'estimation). Il est clair qu'entre l'échantillon et sa zone d'influence il n'y a pas identité, mais seulement corrélation, corrélation d'autant meilleure que la zone d'influence est plus petite et que la minéralisation est elle-même plus continue.

De même encore les méthodes d'interpolation fonctionnelle surestiment, en général, de façon inadmissible le degré de continuité du phénomène représenté. Par quatorze points expérimentaux, on peut toujours faire passer un polynome du treizième degré. Mais, en général, ce polynome ne reflète pas le moins du monde l'évolution réelle du phénomène entre les points expérimentaux.

Et cependant, sous la complication et l'irrégularité extrême que présente une régionalisation dans sa variation spatiale se dissimule, en général, *la structure d'un phénomène naturel.* C'est de cette structure saptiale qu'à leur tour ne peuvent pas rendre compte les méthodes purement statistiques. Quand on classe les échantillons sous forme d'histogramme, on fait, par là même, abstraction de l'endroit où ils ont été prélevés : on détruit les structures spatiales, qui constituent justement l'aspect le plus important du phénomène. Dans un gisement minier, par exemple, il ne suffit pas de savoir avec quelle fréquence se répète une teneur donnée. On a besoin, aussi, d'apprécier le degré de continuité de la minéralisation, de connaître la taille et la position des zones exploitables, etc.

Parmi les caractères qualitatifs ou structurels que l'histogramme statistique est incapable d'exprimer, et qui doivent obligatoirement être pris en charge par une théorie des variables régionalisées, nous mentionnerons les suivants :

Localisation

Une variable régionalisée ne prend pas ses valeurs n'importe où, mais dans un domaine bien défini de l'espace, que l'on appelle son *champ géométrique*. Pour une teneur, par exemple, le champ géométrique V sera la formation minéralisée elle-même, ou éventuellement une portion de celle-ci. De même, la variable elle-même est parfois définie comme une fonction de point $f(x)$ [dans un gisement, $f(x) = 0$ ou 1 selon que le point x appartient à un grain de stérile ou de minerai]. Le plus souvent, on s'intéresse plutôt aux valeurs moyennes de la variable dans un petit domaine v ou *support géométrique*. Pour une teneur, le support sera le volume de l'échantillon prélevé : le support doit être défini de manière précise, avec sa forme, ses dimensions et son orientation. Si l'on change le support v ou le champ V, on modifie les caractéristiques apparentes de la régionalisation. Une teneur, par exemple, apparaîtra comme d'autant plus dispersée qu'elle sera définie sur un support plus petit ou dans un champ plus grand.

Une des tâches majeures de la théorie des variables régionalisées consistera à prévoir les caractéristiques de la variable définie sur un support v dans un champ V, connaissant par exemple celles de la variable ponctuelle dans un champ différent V'.

Continuité

Un deuxième caractère essentiel est le degré de plus ou moins grande continuité de la variable régionalisée dans sa variation spatiale. Une estimation effectuée à partir d'un même réseau de prélèvements sera d'autant plus précise que la variable sera plus continue. Dans certains cas (par exemple pour des variables à caractère très géométrique, comme la puissance, ou épaisseur, d'une formation géologique) on observera la continuité stricte, celle qui se définit avec des ε et des η. Plus souvent, on n'observera qu'une continuité plus lâche, dite continuité en moyenne, et parfois même, la continuité en moyenne ne sera même pas vérifiée : dans ce dernier cas, qui correspond à des variables dont le comportement est particulièrement erratique, on dit qu'il y a *effet de pépite* (exemple des gisements d'or pépitique).

Anisotropie

En troisième lieu, les différentes directions de l'espace ne sont pas toujours équivalentes. Il peut arriver qu'une variable régionalisée se modifie lentement ou régulièrement le long d'une direction et présente une variation beaucoup plus rapide ou irrégulière le long d'une direction perpendiculaire. Ce genre de phénomènes est lié à l'existence de structures zonales (runs, rubannements, etc.).

Phénomènes de transition

On voit aussi apparaître certains types de structure liés à la présence d'un réseau de discontinuité dans le champ géométrique de la régionalisation. On parle alors de phénomènes de transition. Dans un gisement sédimentaire, par exemple, la stratification peut donner lieu à un effet de ce genre. Il peut arriver que les teneurs, constantes ou presque constantes tant que l'on reste à l'intérieur d'un même banc, subissent des variations brusques lorsque l'on passe d'un banc à l'autre.

Le langage des fonctions aléatoires

Il fallait donc adopter un mode de formulation capable de prendre en charge ces deux aspects contradictoires, aléatoire et structuré, permettant également de poser et de résoudre des problèmes essentiellement pratiques, comme celui de l'estimation d'une variable régionalisée à partir d'un échantillonnage fragmentaire.

Ce langage adéquat, probabiliste et capable d'exprimer les structures spatiales, c'est évidemment la théorie des fonctions aléatoires qui va le fournir. Nous ferons donc l'hypothése que notre variable régionalisée peut être considérée comme une *réalisation d'une fonction aléatoire.* [On sait qu'une fonction aléatoire est définie par la donnée d'une loi de probabilité sur un espace abstrait de fonctions. Toute fonction particulière obtenue à l'issue d'un tirage au sort effectué selon cette loi de probabilité et sur cet espace de fonctions s'appelle une réalisation de la fonction aléatoire. On peut aussi considérer la fonction aléatoire $f(x)$ comme une variable aléatoire vectorielle à une infinité de composantes. Ces composantes, en nombre infini, sont les valeurs numériques prises par $f(x)$ en chacun des points x de l'espace. Ainsi, il y a le même rapport entre une fonction aléatoire et une de ses réalisations qu'entre une variable aléatoire ordinaire et une valeur numérique particulière obtenue à l'issue d'un tirage au sort effectué selon la loi de probabilité de cette variable.]

Or il y a, dans une telle hypothèse, un aspect presque *platonicien.* La seule réalité digne de ce nom devient alors, en effet, la fonction aléatoire elle-même, définie idéalement par une loi de probabilité donnée sur un espace abstrait. Le phénomène réel, régionalisé, n'est plus considéré que comme une réalisation, c'est-à-dire un reflet lointain et déformé, une imitation maladroite de ce modèle idéal.

Malheureusement, nous n'avons pas accès directement au ciel des idées, et nous devons examiner si une telle hypothése est susceptible de recevoir un sens. Or il apparaît tout de suite une objection très grave, liée à l'unicité des phénomènes naturels et à l'impossibilité de l'inférence statistique. Il n'est pas possible, en effet, de reconstruire la loi d'une fonction aléatoire à partir d'une réalisation unique, pas plus qu'il n'est possible de reconstituer la loi d'une variable aléatoire ordinaire à partir du résultat numérique d'une épreuve unique. En général, donc, le recours à une interprétation probabiliste ne sera pas fondé, et on devra se contenter d'interroger la variable régionalisée elle-même, c'est-à-dire le phénomène naturel donné qui constitue la seule réalité physique. Tel a été effectivement le point de vue adopté dans un premier groupe de méthodes (les méthodes *transitives*), où l'on parvient à des solutions utilisables sans introduire d'hypothèse probabiliste. Je ne vous parlerai pas de ces méthodes, d'une part parce que ce serait un peu long, de l'autre parce que — par une convergence remarquable — les résultats ultimes auxquels conduisent les méthodes transitives ne se distinguent pratiquement pas de ceux que l'on obtient dans la théorie probabiliste des schémas intrinsèques.

Il y a cependant un cas où inférence statistique est fondée, et où par suite le recours au langage probabiliste est justifié. C'est le cas *stationnaire.* Une fonction aléatoire est dite stationnaire si sa loi spatiale est invariante par translation, c'est-à-dire si les valeurs $f(x_1)\ldots f(x_k)$ qu'elle prend en k points d'appui $x_1, x_2\ldots x_k$ ont la même loi de probabilité que les valeurs $f(x^1 + h)\ldots f(x_k + h)$ et celà, quels que soient les k points d'appui, en nombre quelconque, et quel que soit le vecteur de translation h. En termes plus intuitifs, cela signifie que le phénomène se répète indéfiniment lui-même dans l'espace entier et, du fait de cette répétition,

on conçoit bien que l'inférence statistique devienne possible même à partir d'une réalisation unique. C'est à ce cas stationnaire que nous nous limitons dans ce qui suit.

Variances a priori infinies

Dans la théorie des fonctions aléatoires stationnaires d'ordre deux l'outil de travail est la fonction

$$K(h) = E[f(x) f(x+h)]$$

qui représente la covariance des valeurs prises par la fonction aléatoire $f(x)$ en deux points distants de h. Cette covariance n'existe que si la variance à priori existe également, c'est-à-dire si l'on a

$$K(0) < \infty$$

Or, d'assez nombreux phénomènes présentent une capacité de dispersion illimitée et ne peuvent pas être décrits correctement si on leur attribue une variance à priori finie.

Ici d'ailleurs la nature nous tend une sorte de piège. Lorsque l'on prélève des échantillons v dans un champ géométrique V on obtient un histogramme à partir duquel on peut toujours calculer numériquement une variance qui prend ainsi une valeur parfaitement définie. Mais cette variance est, en réalité, une fonction $\sigma^2(v \mid V)$ du support v et du champ V. Elle augmente, en particulier, lorsque le champ V augmente. Si les échantillons de taille v possèdent une variance à priori, celle-ci doit apparaître comme la limite pour V infini de la variance expérimentale $\sigma^2(v \mid V)$. Il y a une dizaine d'années, à propos du grand gisement d'or du Rand, des auteurs d'Afrique du Sud (D.G. Krige, H.S. Sichel, de Wijs, etc...), ont calculé, — à partir de centaine de milliers d'échantillons — la variance de ces échantillons dans des panneaux de plus en plus grands, puis dans une concession entière, puis dans le gisement du Rand dans son ensemble. Ils ont ainsi obtenu, expérimentalement, une relation du type :

$$\sigma^2(v \mid V) = \alpha \log \frac{V}{v}$$

La croissance de la variance observée se poursuit sans défaillance selon cette loi logarithmique jusqu'au dernier point expérimental pour lequel V/v est de l'ordre de plusieurs milliards. On peut conclure en toute certitude qu'il n'existe pas ici de variance à priori finie.

LES SCHÉMAS INTRINSÈQUES

Cependant, même lorsque la variance a priori est infinie, il arrive souvent que les accroissements $f(x+h) - f(x)$ conservent une variance finie. On appellera schéma intrinsèque une fonction aléatoire à accroissements stationnaires d'ordre deux. Étudier une fonction aléatoire en tant que schéma intrinsèque revient à l'étudier par l'intermédiaire de ses accroissements, c'est-à-dire à une constante près [c'est ainsi d'ailleurs que l'on étudie le processus de Wiener — Levy ou les processus poissoniens]. L'outil de base qui remplace la covariance $K(h)$ est le *demi-variogramme*, défini comme l'espérance du carré des accroissements :

$$\gamma(h) = \frac{1}{2} E\left[(f(x+h) - f(x))^2\right]$$

Si la covariance existe, on a

$$\begin{cases} \gamma(h) = K(0) - K(h) \\ K(h) = \gamma(\infty) - \gamma(h) \end{cases}$$

de sorte que le variogramme et la covariance constituent un seul et même outil. Il n'en est ainsi que si $\gamma(\infty)$ existe. Si au contraire $\gamma(h)$ tend vers l'infini avec $|h|$, la variance a priori K (0) est infinie et K (h) n'existe pas. Mais le variogramme $\gamma(h)$ reste défini et permet d'étudier la fonction aléatoire [ainsi, les processus classiques de Poisson ou de Wiener—Levy ont un variogamme linéaire $\gamma(h) = a\,h$ qui tend vers l'infini].

La variance $\sigma^2(v \mid V)$ d'un échantillon v dans un champ V peut s'exprimer à l'aide du variogramme. On trouve

$$\sigma^2(v \mid V) = \frac{1}{V^2}\int_V\int_V \gamma(x - x')\,dxdx' - \frac{1}{v^2}\int_v\int_v \gamma(x - x')\,dxdx' \tag{1}$$

ou, si l'on veut :

$$\sigma^2(v \mid V) = F(V) - F(v) \tag{2}$$

la fonctionnelle F (V) représentant la valeur moyenne de $\gamma(h)$ lorsque les deux extrémités du vecteur h décrivent le volume V.

On voit que, si $\gamma(h)$ tend vers l'infini avec h, F (V) tendra vers l'infini avec V et par conséquent aussi la variance de v dans V. Avec un variogramme du type

$$\gamma(h) = 3\,\alpha \log |h|$$

(et en supposant v et V géométriquement semblables), on obtient

$$\sigma^2(v \mid V) = \alpha \log \frac{V}{v}$$

c'est-à-dire justement la formule obtenue expérimentalement par les auteurs d'Afrique du Sud.

De l'expression (2) résulte une *relation d'additivité.*

$$\sigma^2(v \mid V) = \sigma^2(v \mid V') + \sigma^2(V' \mid V)$$

dont le sens intuitif est évident : la variance d'un échantillon v dans le gisement V est la somme de la variance de v dans un panneau V′ et de la variance du panneau V′ dans le gisement V.

On peut définir également la covariance de deux échantillons v et v' (situés à une distance fixe l'un de l'autre). On obtient une formule analogue à (2), F (v) étant remplacée par la valeur moyenne de $\gamma(h)$ lorsque les deux extrémités de h décrivent respectivement les volumes v et v'. Si $\gamma(h)$ tend vers l'infini avec h, variances et covariances sont des infiniment grands équivalents à F (V), de sorte que le *coefficient de corrélation* expérimental tend toujours vers l'unité : ce coefficient n'est plus un instrument de travail adéquat, et c'est en définitive le variogramme lui-même qui exprime au mieux le réseau des corrélations et la structure spatiale du phénomène.

1) le variogramme donne un contenu précis à la notion traditionnelle de *zone d'influence* d'un échantillon. Sa croissance plus ou moins rapide reflète la manière plus ou moins rapide dont se détériore l'influence d'un échantillon sur des zones de plus en plus lointaines du gisement.

2) *les anisotropies* se manifestent par le comportement différentiel du variogramme dans les différentes directions de l'espace.

3) Les phénomènes *de transition* se traduisent sur le variogramme par des paliers et des seuils dont l'analyse permet souvent de mettre en évidence la superposition de plusieurs structures d'échelles différentes.

4) *La continuité* et la régularité d'une régionalisation, enfin, sont très bien exprimées par le comportement du variogramme au voisinage de l'origine. Par ordre de régularité décroissante, on peut distinguer quatre types :

— Comportement parabolique au voisinage de l'origine :

$\gamma(h)$ deux fois dérivable en $h = 0$.
Ce type correspond à une variable régionalisée dérivable en moyenne quadratique, donc à haut degré de régularité.

— Tangente oblique à l'origine :

$\gamma(h)$ est continu à l'origine, mais n'est pas deux fois dérivable.
Ce type correspond à une variable régionalisée continue mais non dérivable en moyenne quadratique, donc déjà moins régulière.

— Effet de pépite :

il se caractérise par une discontinuité à l'origine, $\gamma(h)$ ne tendant pas vers 0 avec h. Il correspond à une variable régionalisée qui n'est pas continue en moyenne quadratique, donc extrêmement irrégulière.

— Variogramme plat, ou effet de pépite à l'état pur.

Ce cas limite correspond à une variable régionalisée purement aléatoire (mesure aléatoire à accroissements indépendants).

Partie irrégulière

Lorsque $\gamma(h)$ est de la forme $\gamma(r)$, c'est-à-dire ne dépend que du rayon vecteur $r = |h|$, on peut caractériser son comportement au voisinage de l'origine par un développement limité de la forme

$$\gamma(r) = \Sigma\, a_{2n}\, r^{2n} + \Sigma\, C_\lambda\, r^\lambda$$

Sur un tel développement on distingue une partie régulière et une partie irrégulière. La première est constituée de termes de degrés entiers pairs. Si elle existait seule, $\gamma(r)$ serait indéfiniment dérivable, donc aussi la variable régionalisée elle-même qui présenterait ainsi

le plus haut degré de régularité. C'est donc la partie irrégulière, constituée de termes du type r^λ où λ est différent d'un entier pair (éventuellement aussi de termes logarithmiques en $r^{2n} \log r$) qui définit le type d'irrégularité de la variable régionalisée, et, dans cette partie irrégulière, c'est le terme du plus bas degré qui joue le rôle principal.

On est ainsi amené à définir le degré de régularité d'une variable régionalisée comme l'ordre λ du terme irrégulier r^λ du plus bas degré de son variogramme.

Régularisation et montée

Lorsque l'on remplace une variable régionalisée $f(x)$ par la valeur moyenne de $f(x)$ dans un échantillon v prélevé au point x (ou par une moyenne pondérée plus générale) on obtient une nouvelle variable régionalisée dont le variogramme se déduit du $\gamma(h)$ initial par des opérations de convolution, sur lesquelles nous n'insisterons pas ici, mais dont l'effet est toujours régularisant. Un exemple particulièrement instructif, et important dans les applications, de cette régularisation est la *montée* : La montée, opération permettant de passer d'une variable régionalisée définie dans l'espace à n dimensions à des variables régionalisées définies dans les espaces à $n - 1$, $n - 2$... dimensions est la transposition de la technique minière qui consiste à forer des sondages verticaux, tracer des niveaux horizontaux, etc. Par exemple, si $f(x, y, z)$ est la teneur au point de coordonnées $(x\, y\, z)$ à l'intérieur d'une formation stratiforme horizontale, un sondage implanté au point de coordonnées x, y de la surface topographique contient la teneur (exprimée en quantité de métal au mètre carré)

$$\int f(x\, y\, z)\, dz$$

Cette intégrale représente une nouvelle variable régionalisée, définie dans l'espace à 2 dimensions (la surface topographique), déduite de l'ancienne par une montée d'ordre 1.

La régularisation d'une variable régionalisée à la montée se manifeste par l'effet suivant : si le terme irrégulier de plus bas degré du variogramme de la variable initiale est en r^λ, celui de la variable qui s'en déduit par montée d'ordre 1 est en $r^{\lambda+1}$. Le degré de régularité augmente d'une unité (lorsque λ est un entier impair on a la séquence logarithmique : $\log r \to r \to r^2 \log r \to \ldots$ où alternent les termes impairs r^{2k-1} et les termes logarithmiques $r^{2k} \log r$).

Variance d'estimation

Montrons maintenant comment le variogramme permet de résoudre un problème essentiellement pratique comme celui de l'estimation d'une variable régionalisée à partir d'un échantillonnage fragmentaire. Soit par exemple à estimer la teneur moyenne z d'un domaine V :

$$z = \frac{1}{V} \int_V f(x)\, dx$$

à partir de prélèvements (ici supposés ponctuels) effectués en n points $x_1 \ldots x_n$. On forme l'estimateur

$$Y = \frac{1}{n} \sum f(x_i)$$

et, pour apprécier l'ordre de grandeur de l'erreur possible, on introduit la variance $D^2(Y - z)$ de la différence $Y - z$. Si les variances a priori existent, on doit avoir :

$$D^2(Y - z) = \sigma_z^2 + \sigma_Y^2 - 2\sigma_{zY} \tag{3}$$

Par contre, si $\gamma(h)$ devient infini avec h, ces variances et cette covariance a priori n'existent plus. Mais la variance d'estimation $D^2(Y - z)$ reste définie, car la différence $Y - z$ est une combinaison linéaire d'accroissements de $f(x)$. On montre que cette variance est :

$$D^2(Y - z) = \frac{2}{nV}\sum_i \int_V \gamma(x - x_i)\,dx - \frac{1}{n^2}\sum_{ij} \gamma(x_i - x_j) - \frac{1}{V^2}\int_V\int_V \gamma(x - x')\,dx\,dx' \tag{4}$$

Cette formule a la même signification que (3). A la covariance σ_{Yz} correspond le terme mixte, où figurent une intégrale et une sommation discrète, tandis que σ_Y^2 et σ_z^2 ont pour homologues la somme double et l'intégrale double. Mais (4) est plus générale que (3), et subsiste même si les variances a priori n'existent pas.

On notera la structure remarquable de la formule (4), où alternent des expressions exactes et approchées de mêmes intégrales. Le deuxième terme apparaît comme une valeur approchée du premier, lui-même approximation du dernier. La théorie du calcul numérique approché des intégrales nous fait donc pressentir que la variance d'estimation sera d'autant plus faible :

— que le réseau de prélèvements sera plus serré et plus représentatif de la géométrie du domaine V à estimer,

— que le variogramme $\gamma(h)$ sera lui plus régulier analytiquement, et donc que la variable régionalisée sera elle-même plus régulière et continue dans sa variation spatiale.

Ces conclusions sont bien conformes à ce que suggère l'intuition. Toutefois, la formule (4) conduirait à des calculs assez longs, lorsque les prélèvements sont un peu nombreux, et on est amené à utiliser des principes d'approximation.

1) Dans l'espace *à une seule dimension*, et pour des prélèvements à maille régulière a, l'estimation du champ de longueur $L = n\,a$ constitué de la juxtaposition des zones d'influence de longueur a au centre desquelles sont implantés chacun des n échantillon admet la variance.

$$D^2(Y - z) = \frac{1}{n}\sigma_E^2$$

σ_E^2, *variance d'extension* élémentaire, est fonction de a et du variogramme $\gamma(h)$. Si r^λ est le terme irrégulier de plus bas degré, la partie principale de la variance d'extension est proportionnelle à a^λ

$$\sigma_E^2 = A(\lambda)\,a^\lambda$$

Comme $a = \frac{L}{n}$, on voit que la variance d'estimation est de la forme :

$$D^2(Y - Z) = \frac{C}{n^{1+\lambda}}$$

Elle est en $\frac{1}{n^{1+\lambda}}$, et non plus en $\frac{1}{n}$ comme dans la statistique ordinaire. Le cas $\lambda = 0$, qui correspond à un effet de pépite, redonne bien la formule usuelle en $\frac{1}{n}$; cela est naturel, puisque l'effet de pépite à l'état pur correspond à des échantillons indépendants. Mais pour une valeur usuelle comme $\lambda = 1$, la variance est en $\frac{1}{n^2}$ et décroît beaucoup plus vite : dans les applications, ce phénomène permet des économies substantielles.

2) Dans l'espace *à plusieurs dimensions*, on applique un principe d'approximation connu sous le nom de composition des termes de tranches et des termes de lignes. Exposons-le sur un exemple. Soit un gisement filonien reconnu par traçages (galeries horizontales tracées dans le plan du filon) à des niveaux équidistants. Chaque traçage est lui-même échantillonné par des prélèvements équidistants (on suppose que la maille a de ces prélèvements est inférieure à la relevée h entre traçages). L'erreur d'estimation apparaît comme la somme de deux erreurs :

— celle que l'on commet en étendant aux traçages la teneur moyenne des prélèvements,

— celle que l'on commet en étendant au gisement lui-même la teneur des traçages supposée parfaitement connue.

La première erreur admet une variance $\frac{1}{n}\sigma_1^2(a)$, où n est le nombre total de prélèvements et $\sigma_1^2(a)$ la variance d'extension d'un échantillon dans sa zone d'influence, à une dimension, de longueur égale à la maille a. On la calcule comme ci-dessus.

La deuxième erreur admet la variance $\frac{1}{N}\sigma_2^2(h)$, N étant le nombre des niveaux tracés, et $\sigma_2^2(h)$ la variance d'extension d'un traçage dans sa tranche d'influence ($\sigma_2^2(h)$ se calcule à partir du variogramme déduit de $\gamma(h)$ par montée d'ordre 1).

Si a est inférieur à h, on peut montrer que ces deux erreurs sont à peu près indépendantes, de sorte que la variance d'estimation est :

$$D^2(Y - Z) = \frac{1}{n}\sigma_1^2(a) + \frac{1}{N}\sigma_2^2(h)$$

En général, les prélèvements reviennent moins cher que le mètre de galerie. Néanmoins, il est inutile de les multiplier indéfiniment : la variance d'estimation ne descendra pas en dessous de $\frac{1}{N}\sigma_2^2(h)$, limite assez vite atteinte dès que n est grand. Il arrive un moment où on ne peut plus améliorer l'estimation sans travaux miniers supplémentaires.

Connaissant le prix de revient du prélèvement et du mètre de traçage, et l'expression de $\sigma_1^2(a)$ et $\sigma_2^2(h)$ en fonction de a et de h un calcul d'optimisation élémentaire permet de déterminer le nombre d'échantillons que l'on doit prélever au mètre de traçage pour minimiser la variance d'estimation à dépense donnée.

D'une manière générale, la théorie des variables régionalisées, en donnant, sous la forme d'une variance d'estimation, une mesure précise de l'information disponible, fournit une base solide pour la recherche d'optima économiques en matière de reconnaissance

minière. Je ne peux malheureusement pas développer ce point ici, non plus que bien d'autres applications intéressantes, telles que : le krigeage (ou recherche de l'estimateur optimal), la simulation numérique des gisements miniers, la théorie des milieux poreux, etc... pour lesquelles on me pardonnera de renvoyer à la bibliographie sommaire suivante :

G. MATHERON

Ingénieur au Corps des Mines,

Docteur ès Sciences, École Nationale Supérieure des Mines de Paris

LISTE BIBLIOGRAPHIQUE

Traité de géostatistique appliquée. Tome I (1962), Tome II (1963). Ed. Technip. Paris.
Les variables régionalisées et leur estimation (Thèse) — Masson, Paris 1965.
Granulométries en place et milieux aléatoires — (Thèse complémentaire).
Recherche d'optimum économique dans la reconnaissance et la mise en exploitation des gisements miniers — En collaboration avec Ph. FORMERY. Annales des Mines, Mai et Juin 1963.
Principles of geostatistics — Economic Geology — Déc. 1963.
Structure et composition des perméabilités — Revue de l'I. F. P. Avril 1966.
Comparaison entre les échantillonnages à poids constant et à effectif constant. « Revue de l'Industrie minérale » (à paraître).

DISCUSSION

M. FOURASTIÉ fait remarquer l'analogie qui existe entre les formulations aléatoires de M. Matheron et celles que M. Vendryès utilise en biologie et pour la description des « articulations mentales ». Il faut accorder une grande importance à ces analogies qui s'étendent ainsi des sciences humaines aux sciences physiques. A ce point de vue comme à plusieurs autres, l'étude de M. Matheron est remarquable.

— Le conférencier indique qu'il a effectivement essayé d'étudier la répartition des densités de population à la surface de la France. Le variogramme obtenu obéit au même schéma logarithmique que les teneurs en or du gisement du Rand. Il faut toutefois éliminer au préalable la région parisienne, qui se comporte comme une énorme pépite un peu aberrante.

M. JOLY demande alors à M. Matheron si, pour cet essai, celui-ci a utilisé les densités moyennes de chaque départements ou s'il s'est servi des densités de circonscriptions territoriales petites. (Le conférencier s'est servi des moyennes départementales).

M. Joly pense, qu'à première vue, les travaux du conférencier devraient recevoir une meilleure application dans ce domaine lorsqu'il s'agit d'un sondage véritable à partir duquel on cherche à reconstituer un ensemble plutôt que dans un travail à posteriori utilisant des valeurs moyennes établies pour des circonscriptions relativement grandes, de dimensions variables, de limites arbitraires et recouvrant le territoire d'une manière exhaustive.

Reprenant l'intervention de M. Fourastié sur l'utilisation des travaux du conférencier en matière de géographie physique, M. Joly cite comme application possible à priori l'étude des climats et micro-climats.

M. Henri GUITTON. — La très intéressante analyse que vous venez de faire sur les variables régionalisées a surgi de votre expérience de géologue. Mais tout en vous écoutant, j'ai compris qu'elle avait une valeur générale, et qu'elle pourrait fort bien être étendue aux problèmes économiques et sociaux.

Je vous avouerai d'abord que j'ai été surpris par votre expression de « régionalisation ». Vous dites que c'est un terme neutre, purement descriptif. Je veux bien vous croire quand on se place dans votre perspective de mineur. Pour nous autres économistes, ce terme a un tout autre sens : il est lié à l'organisation du territoire. Nous risquons ainsi entre nous l'ambiguïté. Je vois cependant très bien l'importance de ce que vous appelez un « champ géométrique ». Je me demande si « spatialisé » ne serait pas meilleur que « régionalisé ». Je n'en suis toutefois pas assuré; car de même qu'il existe une structure des choses dans l'espace, il y a aussi une structure dans le temps. Votre analyse n'a pas été faite dans ce sens. Je pose la question : « ne pourrait-elle pas lui être aussi appliquée? »

En ce qui concerne l'espace, j'ai été très frappé par votre critique du procédé de l'histogramme, qui détruit les structures spatiales. Cette critique doit valoir aussi pour les observations humaines. J'aurais cependant scrupule à abandonner d'un seul coup cette « graphie » si commode. Je n'oublie pas que le mot « histos » veut dire tissu : vous vous appliquez, bien vous aussi, à exprimer un « tissu », une contexture.

Je retiens votre idée majeure : il y a dans le comportement des variables régionalisées un aspect aléatoire. J'ai été tout d'abord surpris par le nouvel outil d'analyse dont vous avez parlé, comme si nous le connaissions, à savoir le *variogramme*. Vous en êtes en somme le créateur, et nous avons à faire connaissance avec lui, à nous familiariser avec son usage, puisque c'est lui désormais qui exprime le mieux le réseau des corrélations et la structure spatiale du phénomène.

Je demande réflexion plus poussée pour être sûr que ce variogramme transformera aussi les analyses économiques, comme il est en train de transformer les analyses minières. Je pense par exemple à la structure d'un revenu national, c'est notre champ, selon les revenus, selon les âges, selon les lieux,... Il y a tellement de types d'espaces à considérer.

En tout cas je rends hommage à votre interprétation platonicienne du monde statistique aléatoire. J'aimerais revenir un jour sur l'idée que vous avez lancée : « la seule réalité digne de ce nom devient la fonction aléatoire elle-même, définie idéalement par une loi de probabilité donnée sur un espace abstrait ».

Je suis de ceux qui pensent en effet, que plus les techniques statistiques s'améliorent et se compliquent, plus elles doivent être éclairées, soutenues, vivifiées par une pensée philosophique.

M. RÉMERY. — Les problèmes évoqués à propos de gisements miniers semblent, à priori, présenter un très grand caractère de généralité.

Ils se retrouvent tant dans le domaine du comportement du consommateur sur le plan géographique ou sur le plan temporel que dans ceux de la variation dans le temps des caractéristiques techniques (rendement, taux d'arrêt des machines, % de rebut...) ou économiques (prix...).

Or, on sait depuis fort longtemps les mécomptes dus à l'hypothèse de l'existence d'une fonction unique, continue... exprimant de telles variations spaciales ou temporelles. Ce qui n'a pas suffi à modifier le courant classique et la littérature sur les chroniques qui feint d'ignorer l'existence possible de plusieurs fonctions de nature différente.

D'où l'intérêt certain de l'introduction de fonctions aléatoires dans l'étude de tous ces phénomènes et spaciaux et temporels. Malheureusement, la difficulté d'observation est grande car l'univers en cause, loin d'être figé se meut constamment sous des impulsions généralement mal connues.

M. Bernard remarque que la distribution des variables « régionalisées » dans les exemples miniers de M. Matheron conduit à une fonction de la forme $3 \alpha \log r$ qui est analogue à la fonction de distribution de revenus, de Pareto. Ce sont donc des problèmes formellement analogues.

— Le conférencier pense, en effet, que la technique du variogramme doit être utilisable dans bien des domaines des sciences humaines ou économiques. N'étant pas économiste lui-même. il a utilisé l'expression « régionalisation » comme un terme neutre n'impliquant rien de plus que l'existence d'une certaine structure spatiale. Les économistes donnent à cette même expression un sens plus précis et plus spécifique, se rattachant néanmoins à la même idée générale. Rien ne s'oppose à ce qu'un même terme soit utilisé sensu lato et sensu stricto respectivement dans deux disciplines assez distinctes pour qu'aucune ambiguité ne puisse en résulter.