N5-47

BUREAU DE RECHERCHES GEOLOGIQUES
ET MINIERES

Département des Réserves

NOTE GEOSTATISTIQUE N° 49

LES LOIS DE PROBABILITES DES REGIONALISATIONS INTRINSEQUES

\- o -

G. MATHERON

Juin 1963

IG/Rés.

NOTE GEOSTATISTIQUE N° 49

LES LOIS DE PROBABILITES DES REGIONALISATIONS INTRINSEQUES

1.- LOIS INDEFINIMENT DIVISIBLES.

Rappelons d'abord la notion de loi indéfiniment divisible. Une variable aléatoire X est dite obéir à une loi indéfiniment divisible s'il est possible, quel que soit l'entier n, de considérer X comme la somme :

$$X = X_1 + \dots + X_n$$

de n variables X_i indépendantes obéissant à une même loi de probabilité. Autrement dit, si $C(u)$ est la fonction caractéristique de la loi de X, $[C(u)]^{\frac{1}{n}}$ est, quel que soit n, fonction caractéristique d'une loi de probabilité, qui n'est autre que la loi des X_i. Dans le cas où X a une variance finie, un théorème très général nous donne une condition nécessaire et suffisante pour que la loi de X soit indéfiniment divisible. Cette condition exprime que le logarithme de la fonction caractéristique doit être de la forme :

$$(1) \quad \log C(u) = ium + \int_{-\infty}^{\infty} \frac{e^{iut} - 1 - iut}{t^2} \, dG(t)$$

G étant une mesure positive sommable. La somme de G est égale à la variance, comme on le voit en dérivant (1) deux fois

$$D^2(X) = \int_{-\infty}^{\infty} dG(t)$$

Si la mesure G est une mesure de Dirac $\sigma^2 \delta(t)$, la relation (1) se réduit à :

$$(2) \quad \log C(u) = ium - \tfrac{1}{2} \sigma^2 u^2$$

et la loi de X n'est autre que la loi normale (m, σ^2). Si l'on prend pour Θ une translatée $\lambda^2 \Theta \delta(t - \theta\lambda)$ d'une mesure de Dirac, il reste :

$$(3) \quad \log C(u) = iu(m - \theta\lambda) + \theta\,(e^{iu\lambda} - 1)$$

Le changement de variable

$$Y = \frac{X - m}{\lambda} + \theta$$

montre que Y est une variable de Poisson, de paramètre θ. La variable X, se déduisant de Y par la relation

$$X = m + \lambda(Y - \theta)$$

sera dite poissonienne au sens large.

Ainsi, la condition (1) exprime que toute variable à loi indéfiniment divisible peut être considérée comme la somme d'une variable normale et d'une infinité de variables poissonniennes indépendantes.

On sait que dans un processus stationnaire à accroissements indépendants les lois de probabilité sont obligatoirement de ce type.

Dans la suite, nous nous intéresserons particulièrement à des variables X (teneurs etc ...) ne pouvant prendre que des valeurs positives. Une loi indéfiniment divisible très générale, pour une variable positive de ce type, s'obtient en composant des variables poissonniennes positives, soit :

$$(4) \quad \log C(u) = \int_0^\infty (e^{iu\lambda} - 1)\, d\theta(\lambda)$$

θ étant une mesure positive sommable sur le demi-axe des $\lambda \geqslant 0$. On notera que la moyenne et la variance

$$\begin{cases} m = \displaystyle\int_0^\infty \lambda\, d\theta(\lambda) \\ \sigma^2 = \displaystyle\int_0^\infty \lambda^2\, d\theta(\lambda) \end{cases}$$

ne sont pas nécessairement finies. La relation (4) s'écrit aussi :

$$(5) \quad \log C(u) = \psi(u) - \psi(0)$$

La fonction $\psi(u)$ est la transformée de Fourier de la mesure Θ (celle-ci étant supposée identiquement nulle sur le demi axe des $\lambda < 0$). Il en résulte deux conséquences intéressantes, qui permettent de préciser la nature analytique de $C(u)$.

En premier lieu, en tant que transformée de Fourier d'une mesure positive sommable, $\psi(u)$ est, d'après le théorème de Bochner, une fonction continue de type positif. Il en est donc de même de $\log C(u)$.

En deuxième lieu, en tant que transformée de Fourier d'une mesure identiquement nulle sur le demi axe des λ négatifs, $\psi(u)$ peut être considérée comme la limite, pour $v \to 0$, d'une fonction $\psi(u + iv)$ de la variable complexe $u + iv$ ne possédant aucun pôle dans le demi plan des v positifs.

Inversement, si ces deux conditions sont remplies, $\log C(u) = \psi(u) - \psi(0)$, se met sous la forme (4) et définit une loi positive indéfiniment divisible.

Comme exemple simple, citons la fonction :

$$\psi(u) = - \alpha \log \left(1 - i \frac{u}{\beta}\right)$$

qui correspond à la variable X obéissant à la loi gamma de paramètre α et β.

En ce qui concerne les lois à plusieurs variables, nous obtiendrons une famille assez vaste (mais non la plus générale possible) de lois positives indéfiniment divisibles en prenant des fonctions caractéristiques $C(u,v \ldots)$ de la forme :

$$(6) \quad \log C(u,v \ldots) = \int \left[e^{iu\lambda + iv\mu + \ldots} - 1\right] \quad d\,\Theta(\lambda, \mu)\ldots]$$

$\Theta(\lambda, \mu \ldots)$ étant une mesure positive sommable de l'espace des λ, μ ..., identiquement nulle en dehors du secteur $\lambda \geqslant 0$, $\mu \geqslant 0$ ... Les variables X Y, Z ... obéissant à une loi simultanée de type (6) peuvent être considérées comme des sommes de variables poissoniennes X_i, Y_i, Z_i, chaque variable d'indice i étant indépendantes des variables d'indice $j \neq i$, et les variables de même indice étant liées par une relation fonctionnelle de la forme :

$$\frac{x}{\lambda} = \frac{y}{\mu} = \ldots$$

2.- LOIS LINEAIREMENT STABLES

Lorsque n variables non indépendantes $X_1 \ldots X_n$ obéissent à une loi normale à n composantes, toute combinaison linéaire

$$(7) \quad Z = \sum \lambda_i \ X_i$$

est une variable normale. On peut se demander si d'autres lois que la loi de Gauss possèdent cette même propriété. Du point de vue pratique, on peut rattacher à ce problème l'énigme de la permanence de la lognormalité dans certains gisements : est-il possible de définir des lois à n composantes telles que toutes les combinaisons linéaires (7) à coefficients λ_i positifs soient individuellement lognormales ? En fait, nous verrons que la réponse est négative, car la loi lognormale n'est pas indéfiniment divisible. La permanence, constatée expérimentalement, ne peut être vérifiée qu'approximativement, et dans certaines limites. Par contre, toutes les lois du type (6) possèdent la stabilité linéaire, car toute combinaison linéaire à coefficient positif vérifie une loi du type (4).

Les lois normale et lognormale ne dépendent que de deux paramètres : une moyenne m et une variance σ^2. On peut se demander à quelle condition doit satisfaire une loi dépendant de deux paramètres seulement pour posséder la stabilité linéaire. . it :

$$\psi(u; m, \sigma^2) = \log C(u; m, \sigma^2)$$

le logarithme de la fonction caractéristique d'une telle loi. Considérons n variables X_j, non indépendantes, obéissant individuellement à la loi ψ, avec :

$$\begin{cases} E(X_j) = m_j \\ D^2(X_j) = \sigma_j^2 \end{cases}$$

et soit, de plus, σ_{ij} la covariance de X_i et X_j. La combinaison

$$Z = \sum u_j X_j$$

doit, par hypothèse, vérifier une loi ψ de paramètres

$$\begin{cases} m = \sum u_j m_j \\ \sigma^2 = \sum_{ij} u_i u_j \sigma_{ij} \end{cases}$$

Par suite, on doit avoir :

$$(8) \quad E(e^{iz}) = E\left[e^{i\sum u_j x_j}\right] = C\left[1; \sum u_j m_j; \sum u_j u_k \sigma_{jk}\right]$$

Si donc la loi ψ est linéairement stable, la fonction caractéristique de la loi de répartition simultanée des X_j est donnée par l'équation (8), qui constitue ainsi une condition nécessaire. Ainsi, seules sont stables linéairement les lois dont la fonction caractéristique $C(u; m, \sigma^2)$ est telle que $C\ 1; \sum u_j m_j, \sum u_j u_k \sigma_{jk}$ soit fonction caractéristique d'une loi de probabilité à plusieurs dimensions. Il est clair, d'ailleurs, que la loi normale vérifie bien cette condition. Prenant :

$$\log C\left[u; m, \sigma^2\right] = ium - \tfrac{1}{2}\sigma^2 u^2$$

on trouve, en effet :

$$\log C\left[1; \sum u_j m_j, \sum u_j u_k \sigma_{jk}\right] = i\sum u_j m_j - \tfrac{1}{2}\sum u_j u_k \sigma_{jk}$$

c'est-à-dire la fonction caractéristique d'une loi normale à plusieurs composantes.

Dans le cas d'une variable à valeur positive (lognormale par exemple), seules les combinaisons linéaires à coefficients $u_j \geqslant 0$ peuvent être assujetties à vérifier une loi du même type (lognormal par exemple); de sorte que (8) ne définit la fonction caractéristique des x_j que dans le secteur $u_j \geqslant 0$. Mais, comme la loi des x_j est elle-même cantonnée dans le quadrant des x_j positifs, on sait que sa fonction caractéristique peut être regardée comme la limite, pour $v_j = 0$, d'une fonction des variables complexes $u_j + iv_j$ dépourvue de pôles dans le domaine $v_j > 0$. Le prolongement analytique est possible, et permet de définir la fonction caractéristique pour des valeurs quelconques des u_j.

Pour une loi gamma, par exemple, on a :

$$\log C(u;m,\sigma^2) = -\frac{m^2}{\sigma^2} \log\left[1 - iu\frac{\sigma^2}{m}\right]$$

Si cette loi est linéairement stable, l'expression

$$(9)\quad \log C(1; m\,\sigma^2) = -\frac{\sum u_j u_k m_j m_k}{\sum u_j u_k \sigma_{jk}} \log\left[1 - i\frac{\sum u_j u_k \sigma_{jk}}{\sum u_j m_j}\right]$$

donne la fonction caractéristique de la loi simultanée des x_j. Mais il reste à vérifier (ce qui n'est pas évident) que l'expression obtenue est effectivement fonction caractéristique d'une loi de probabilité, autrement dit qu'elle est de type positif. En fait, il ne semble pas qu'il en soit ainsi, car, dans le plan $\sum u_j m_j = 0$, la fonction $C(1,m,\sigma^2)$ reste constante et égale à l'unité, alors qu'une fonction de type positif est bornée strictement par sa valeur à l'origine. Ainsi la loi gamma n'est pas linéairement stable.

Le cas de la loi lognormale n'est pas aisé à étudier directement. La fonction caractéristique

$$\begin{cases} C(u; m\,\sigma^2) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{iu\gamma e^{\tau z}}\, e^{-\frac{z^2}{2}}\, dz \\ \log\gamma = \log m - \frac{1}{2}\tau^2 \\ \tau^2 = \log\left[1 + \dfrac{\sigma^2}{m^2}\right] \end{cases}$$

est, en effet, peu maniable. Nous procéderons indirectement, en appliquant (8) au cas de deux variables x et y, et en montrant que la courbe de régression de y en x que l'on déduit de (8) ne peut pas convenir.

Soit $f(xy)$ et $\Phi(uv)$ la loi de xy et sa fonction caractéristique. On a:

$$E(y.x) = \frac{\int y\, f(xy)dy}{\int f(xy)\, dy} = -i\, \frac{\int e^{-iux} \frac{\partial}{\partial v} \Phi(u,o)du}{\int e^{-iux} \Phi(u,o)du}$$

Numérateurs et dénominateurs sont les inverses de Fourier de $-i \frac{\partial}{\partial v} \Phi(u,0)$ et de $\Phi(u,0)$.
Mais, d'après (8), nous avons :

$$\Phi(u,v) = C \left[1;\; m_1 u + m_2 v\; ;\sigma_1^2\, u^2 + 2uv\;\; \sigma_{12} + v^2 \sigma_2^2\right]$$

Par suite, on a immédiatement :

$$\Phi(u,0) = C \left[1; m_1 u; \sigma_1^2\, u^2\right] = C(u;\, m_1,\, \sigma_1^2)$$

et

$$\frac{\partial}{\partial v} \Phi(u,0) = m_2\, C'_m\, (1;\, um_1, u^2\, \sigma_1^2) + 2\, u\, \sigma_{12}\, C'_{\sigma^2}\, (1; um_1,\, u^2\, \sigma_1^2)$$

$$= \frac{m^2}{u} \frac{\partial}{\partial m}\;\; C(u;\, m_1,\, \sigma_1^2) + 2\;\; \frac{\sigma_{12}}{u} \frac{\partial}{\partial_{\sigma}^{2}}\;\; C(u;\, m_1,\, \sigma_1^2)$$

Désignons par $f(x)$ la loi marginale de x

$$f(x) = \int f(xy)dy = \frac{1}{2\pi} \int e^{-iux}\, \Phi(u,0)du$$

On obtient le résultat suivant ;

$$-\frac{i}{2\pi} \int e^{-iux} \frac{\partial}{\partial v} \Phi(u,0)du = \frac{1}{2\pi}\, (m_2 \frac{\partial}{\partial m} + 2\, \sigma_{12} \frac{\partial}{\partial_{\sigma}^{2}}) \int -i \frac{e^{-iux}}{u}\;\; C(u)du$$

$$= -\left[m_2\; \frac{\partial}{\partial m} + 2\, \sigma_{12} \frac{\partial}{\partial_{\sigma}^{2}}\right] \int_{-\infty}^{x} f(x)dx$$

Prenant

$$F(x;\, m_1,\, \sigma_1^2) = \int_{-\infty}^{x} f(x)dx$$

Il vient ainsi

$$(10)\quad E(y.x) = -\frac{1}{f(x)}\left[m_2 \frac{\partial}{\partial m_1} F(x; m_1\, \sigma_1^2) + 2\,\sigma_{12} \frac{\partial}{\partial \sigma^2} F(x, m_1\, \sigma_1^2)\right]$$

Appliquons ce résultat au cas lognormal :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{1}{\tau}\log\frac{x}{\gamma}} e^{-\frac{z^2}{2}}\, dz$$

$$\begin{cases} \tau^2 = \log(1 + \frac{\sigma^2}{m^2}) \\ \log \gamma = \log m - \frac{1}{2}\tau^2 \end{cases}$$

Posant

$$z = \frac{1}{\tau}\log\frac{x}{m} + \frac{1}{2}\,\tau$$

on trouve

$$E(y\,.\,x) = \frac{-1}{\frac{\partial z}{\partial x}}\left[m_2 \frac{\partial z}{\partial m} + 2\,\sigma_{12} \frac{\partial z}{\partial \sigma^2}\right]$$

$$= x\,\frac{m_2}{m_1} + x\left[\frac{1}{2\tau^2}\log\frac{x}{m_1} - \frac{1}{4}\right]\left[\frac{2\,\sigma_{12} - 2\,\sigma_1^2\,\frac{m_2}{m_1}}{m_1^2 + \sigma_1^2}\right]$$

Et cette expression prend nécessairement des valeurs négatives soit pour x petit, soit pour x grand, ce qui n'est pas possible, puisque x et y ne prennent que des valeurs positives : ce résultat montre que $\Phi(u,v)$ ne peut pas être la transformée de Fourier d'une mesure positive, autrement dit que la loi lognormale n'est pas stable linéairement.

Cet échec était prévisible, puisque la loi lognormale n'est pas indéfiniment divisible. Mais la loi gamma, pourtant indéfiniment divisible, n'est pas non plus linéairement stable. Il convient donc de s'intéresser à des familles de lois dépendant de plus de deux paramètres, c'est-à-dire, pratiquement, à des lois du type (6).

3.- CONSTRUCTION D'UNE REGIONALISATION INTRINSEQUE A LOI INDEFINIMENT DIVISIBLE.

Considérons un processus stationnaire à accroissements indépendants, c'est-à-dire une mesure aléatoire T à densité de variance constante. Elle sera complétement définie si l'on se donne la loi de probabilité de la variable $T\varphi$, pour toute fonction de base φ, loi obligatoirement indéfiniment divisible. Dans le cas général, cette loi doit être du type (1). Prenons comme fonction Φ la fonction caractéristique (ensembliste) d'un volume élémentaire dv, et désignons par $m = E(T)$ et D les densités de moyenne et de variance de T, constantes puisque T est stationnaire. On aura :

$$\log C\left[u\varphi\right] = \log E\left[e^{iuT\varphi}\right] = iumdv + dv\int_{-\infty}^{+\infty} D\,\frac{e^{iut} - 1 - iut}{t^2}\,d\,G(t)$$

G étant une mesure positive de somme 1. Soit maintenant une fonction borélienne φ à support borné. On considérera des sommes de type

$$\psi(x) = \sum_i \varphi(x_i)\,\Phi_i(x)$$

où les Φ_i sont les fonctions caractéristiques de volumes élémentaires Δv, centrés aux points x_i et constituant une partition du support de φ. Lorsque l'on fait tendre Δv vers 0 en augmentant indéfiniment le nombre des x_i, $\psi(x)$ converge vers $\varphi(x)$ et, puisque T est une mesure

$$T\psi \longrightarrow T\varphi$$

Mais $T\psi$ est de la forme

$$T\psi = \sum_i \varphi(x_i)\;T\,\Phi_i$$

Comme les ϕ_i sont à support disjoints, les $T\varphi_i$ sont indépendants et obéissent à la même loi

$$\log C(u\varphi_i) = ium\,\Delta v + \Delta v\, D \int \frac{e^{iut} - 1 - iut}{t^2}\, dG(t)$$

Par suite, on a :

$$\log E(e^{it\psi}) = i\,m \sum_i \phi(x_i)\Delta v + D\sum_i \Delta v \int \frac{e^{it\phi(x_i)} - 1 - i\phi(x_i)t}{t^2}\, dG(t)$$

Lorsque $\Delta v \longrightarrow 0$, $\log E(e^{iT\psi})$ converge vers $\log C(\varphi)$, logarithme de la fonctionnelle caractéristique $C(\varphi)$, d'après la condition de continuité posée dans la définition des distributions aléatoires. On obtient ainsi, à la limite :

$$(11)\quad \log C(\varphi) = i\,m \int \phi(x)dx + D\int dx \int \frac{e^{it\varphi(x)} - 1 - it\varphi(x)}{t^2}\, dG(t)$$

Cette formule donne la loi la plus générale d'un processus stationnaire à accroissements indépendants et à variance finie. Si l'on se limite au cas où T est une mesure positive (quantité de métal), on remplace (1) par (4) et (11) devient :

$$(12)\quad \log C(\varphi) = \int dx \int_0^{\infty} \left[e^{i\lambda\varphi(x)} - 1 \right] d\theta(\lambda)$$

La loi simultanée de plusieurs $T\varphi_i$ s'en déduit immédiatement, en prenant :

$$(13)\quad \log C\left[\sum u_i\,\varphi_i\right] = \int dx \int \left[e^{\lambda i \sum u_j \Phi_j} - 1 \right] d\theta(\lambda)$$

Cette loi reste du type général (6)

Supposons maintenant que nous régularisions la mesure aléatoire T par une fonction $k(x)$: on obtient ainsi une variable régionalisée à support ponctuel :

$$(14)\quad f(x) = T * \check{k}$$

qui possède, naturellement, le caractère intrinsèque. Désignons par K(h) le covariogramme transitif associé à la fonction k :

$$(15)\quad K = k * \check{k}$$

La fonction de covariance de f(x) est :

$$D\ \delta * K = D\ K(h)$$

elle ne diffère que par la constante D du covariogramme K(h). Si donc on sait résoudre l'équation (14), ce qui peut se faire en passant aux images de Fourier, on est capable de construire une régionalisation (14) obéissant à un variogramme K(h) donné, et celle-ci est entièrement définie, car sa fonctionnelle caractéristique s'obtient en remplaçant φ par $\varphi * k$ dans (11) ou (12).

Par exemple, si T obéit à une pure loi de Poisson de densité θ, on aura :

$$(16)\quad \log C(\varphi) = \theta \int \left[e^{i\varphi(x) * k(x)} - 1 \right] dx$$

4.- CAS POISSONIEN PUR

Etudions plus en détail ce cas particulier (16). La variable ponctuelle f(x) définie en (14) s'obtient en faisant $\varphi = u\delta(x)$. Sa loi de probabilité est donnée par :

$$\log C\left[u\ \delta(x)\right] = \theta \int \left[e^{iuk(x)} - 1 \right] dx$$

C'est une loi du type (4). Si k(x) est la variable géométrique associée à un volume $V(k(x) = 1$ si $x \in V$, $k(x) = 0$ Si $x \notin (V)$, on obtient très simplement :

$$\log C\ (u\delta) = \theta\ V(e^{iu} - 1)$$

f(x) est, dans ce cas, une variable de Poisson de paramètre θV.

De même, la loi simultanée de plusieurs prélèvements ϕ_j est donnée par

$$\log C \left[\sum u_j \phi_j\right] = \theta \int \left[e^{ik * \sum u_j \Phi_j} - 1\right] dx$$

Si les $\phi_j = \delta(x - a_j)$ sont des prélèvements ponctuels effectués aux points a_j, il vient :

$$(17) \quad \log C = \theta \int \left[e^{i \sum u_j \, k(x-a_j)} - 1\right] dx$$

Si k est la variable géométrique d'un volume V, l'intégrale (17) fait apparaître les intérsections d'ordre 2, 3 ... des volumes V centrés aux points a_j. Par exemple, avec deux variables u et v correspondant à 2 points distants de h, on trouve :

$$\log C = \theta \int \left[e^{i\,u\,k(x) + iv\,(k(x+h)} - 1\right] dx$$

$$= \theta\, K(h) \left[e^{iu + iv} - 1\right] + \theta \left[K(0) - K(h)\right] \left[e^{iu} + e^{iv} - 2\right]$$

Désignons par y et z les deux variables aléatoires obéissant à la loi précédente. Chacune d'elle est individuellement poissonienne, de paramètre $\theta\, K(0) = \theta\, V$, et elles sont reliées par la covariance $\sigma_{yz} = \theta\, K(h)$, de sorte que la fonction caractéristique se met sous la forme

$$(18) \quad \log C = \sigma_{yz} \left[e^{iu + iv} - 1 + (\sigma^2 - \sigma_{yz})(e^{iu} + e^{iv} - 2)\right]$$

Ainsi y et z peuvent être mises sous la forme :

$$\begin{cases} y = y_1 + t \\ z = z_1 + t \end{cases}$$

Les trois variables y_1, z_1 et t étant poissoniennes et indépendantes, de paramètre $\sigma^2 - \sigma_{yz}$ pour y_1 et z_1 et σ_{yz} pour t. Posant maintenant

$$\sigma_{yz} = \rho\, \sigma^2$$

mettons (18) sous la forme

$$\log C = \sigma^2 (e^{iu} - 1) \left[1 + \rho(e^{iv} - 1)\right] + \sigma^2 (e^{iv} - 1)$$

et prenons l'inverse de Fourier de C en u seulement. On obtient une fonction différente de 0 seulement lorsque x est entier. Pour $x = n$, cette fonction est égale à :

$$\frac{\sigma^{2n}}{n!} e^{-\sigma^2} e^{(\sigma^2 - \sigma_{zy})(e^{iv} - 1)} \left[1 + \rho(e^{iv} - 1)\right]^n$$

Après division par la probabilité a priori $\frac{\sigma^{2n}}{n!} e^{-\sigma^2}$ de $y = n$, on obtient la loi de z à y fixé par sa fonction caractéristique :

$$e^{(\sigma^2 - \sigma_{zy})(e^{iv} - 1)} \left[1 + \rho(e^{iv} - 1)\right]^n$$

Ainsi, a $y = n$ fixé, z est la somme de deux variables indépendantes, la première poissonienne de paramètre $\theta = \sigma^2 - \sigma_{yz}$, la deuxième binomiale d'ordre n, avec une probabilité élémentaire $p = \rho$ égale au coefficient de corrélation ρ.

En particulier, à y fixé, valeur probable et variance de z sont données par :

$$(19) \begin{cases} E(z \mid y = n) = \sigma^2 - \sigma_{zy} + \rho n \\ D^2(z \mid y = n) = \sigma^2 - \sigma_{zy} + n\,\rho(1 - \rho) \end{cases}$$

La régression est linéaire, mais la variance liée est une fonction croissante de n.

5.- FORMULATION GENERALE ET APPLICATION A LA LOI GAMMA.

Désignons par $\psi(u)$ la loi élémentaire du processus stationnaire à accroissements indépendants, qui, dans le cas général, est de la forme :

$$(u) = i\,u\,m + \int_{-\infty}^{\infty} \frac{e^{iut} - 1 - iut}{t^2}\, dG(t)$$

et, dans le cas d'une variable positive

$$\psi(u) = \int_0^{\infty} (e^{iu\lambda} - 1)\, d\theta(\lambda)$$

La variable régionalisée, obtenue en régularisant ce processus par une fonction ou une distribution à support borné k, est entièrement définie par sa fonctionnelle caractéristique $C(\varphi)$, qui peut s'écrire :

$$(20) \quad \log C(\varphi) = \int \psi \left[k * \varphi \right] dx$$

En général, on ne saura pas calculer explicitement la loi (20), mais on pourra toujours former ses moments successifs. Introduisons les cumulants, ou semi invariants χ_n de la loi élémentaire ψ

$$\psi(u) = \sum_{n=1} \chi_n \frac{(iu)^n}{n!}$$

(20) peut s'écrire

$$(21) \quad \begin{cases} \log C(\varphi) = \sum\limits_{n=1} C_n(\varphi) \dfrac{i^n}{n!} \\ C_n(\varphi) = \chi_n \int (k * \varphi)^n \, dx \end{cases}$$

Ainsi, la loi de distribution simultanée de p prélèvements $\varphi_1 \ldots \varphi_p$ est déterminée par ses cumulants

$$(22) \quad C_n \left[\sum u_j \varphi_j \right] = \chi_n \int \left[\sum u_j \, k * \varphi_j \right]^n dx$$

A titre d'exemple, supposons que k soit la variable géométrique associée à un volume V et que les φ_j représentent des prélèvements ponctuels effectués aux points a_j. On a ici

$$C_n(u_1 \ldots u_p) = \chi_n \int \left[\sum u_j \, k(x - a_j) \right]^n dx$$
$$= \chi_n \sum \frac{n!}{\alpha_1! \ldots \alpha_p!} u_1^{\alpha_1} \ldots u_p^{\alpha_p} \int \left[k(x-a_1)\right]^{\alpha_1} \ldots \left[k(x-a_p)\right]^{\alpha_p} dx$$

Comme k est une variable géométrique, cette dernière intégrale représente le volume de l'intersection des volumes V centrés aux points a_j pour lesquels $\alpha_j > 0$, et constitue une généralisation du covariogramme transitif au cas des intersections multiples. Examinons le cas de deux prélèvements ponctuels distants de h. D'après (20), la loi de ces deux prélèvements a une fonction caractéristique $C(u,v)$ telle que :

$$\log C(u,v) = \int \psi \left[u\, k(x) + v\, k(x-h) \right] dx$$

Comme k est une variable géométrique de covariogramme transitif $K(h)$, on obtient

$$(23) \quad \log C(u,v) = K(h)\,\psi(u+v) + \left[K(0) - K(h)\right]\left[\psi(u) + \psi(v)\right]$$

Les deux variables aléatoires correspondantes, y et z, peuvent se mettre sous la forme

$$(24) \quad \begin{cases} y = y_1 + t \\ z = z_1 + t \end{cases}$$

les trois variables y_1, z_1 et t étant indépendantes, et obéissant les deux premières à la loi $\left[K(0) - K(h)\right]\psi(u)$ et la dernière à la loi $K(h)\,\psi(u)$. A z fixé, y est la somme de deux variables indépendantes : y_1, obéissant à la loi $\left[K(0) - K(h)\right]\psi(u)$, et t obéissant à la loi de t à z fixé. Si y_1, z_1 et t ont des densités de fréquence a priori $f_1(y_1)$, $f_1(z_1)$ et $f_2(t)$, la loi de t à z fixé a la densité $\dfrac{f_2(t)\, f_1(z-t)}{f(z)}$

A titre d'exemple, examinons le cas où la loi élémentaire est une loi gamma, soit

$$\psi(u) = -\alpha \log\left(1 - i\,\frac{u}{\beta}\right)$$

Les lois de y_1 et t s'en déduisent en remplaçant α par α_1 et α_2 respectivement

$$\begin{cases} \alpha_1 = \left[K(0) - K(h)\right]\alpha = (\sigma^2 - \sigma_{xy})\dfrac{m^2}{\sigma^4} = (1-\rho)\dfrac{m^2}{\sigma^2} \\ \alpha_2 = K(h)\alpha = \sigma_{xy}\dfrac{m^2}{\sigma^4} = \rho\,\dfrac{m^2}{\sigma^2} \end{cases}$$

D'où :

$$\begin{cases} f_1(y) = \dfrac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} \, y^{\alpha_1 - 1} \, e^{-\beta y} \\[2ex] f_2(t) = \dfrac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} \, t^{\alpha_2 - 1} \, e^{-\beta t} \\[2ex] f(z) = \dfrac{\beta^{(\alpha_1 + \alpha_2)}}{\Gamma(\alpha_1 + \alpha_2)} \, z^{\alpha_1 + \alpha_2 - 1} \, e^{-\beta z} \end{cases}$$

A z fixé, y est la somme d'une variable y_1 obéissant à la loi gamma $f_1(y)$ et d'une variable t, indépendante de la précédente, de densité de fréquence :

$$\frac{f_2(t)\, f_1(z-t)}{f(z)} = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_2)\,\Gamma(\alpha_1)} \left(1 - \frac{t}{z}\right)^{\alpha_1 - 1} \left(\frac{t}{z}\right)^{\alpha_2 - 1} \frac{1}{z}$$

C'est-à-dire une variable t telle que $\frac{t}{z}$ obéisse à la loi Béta de paramètres α_1 et α_2. On a ainsi :

$$\begin{cases} E(y.z) = (1 - \rho)m + \rho z \\[2ex] D^2(y.z) = \sigma^2 (1-\rho) \left[1 + \dfrac{\rho\, z^2}{\sigma^2 + m^2} \right] \end{cases}$$

La variance liée augmente rapidement quand z croit (comme dans le cas lognormal), mais la régression reste linéaire. On vérifie facilement, d'ailleurs, que toutes les lois de type (23) conduisent à des courbes de régression linéaire. En dérivant (23) en v, et en faisant $v = 0$, on trouve en effet :

$$\frac{\partial}{\partial v} C(u,0) = C(u,0) \left[K(h)\ \psi'(u) + \left[K(0) - K(h)\right] \psi'(0) \right]$$

$$= \frac{K(h)}{K(0)} \frac{\partial}{\partial u} C(u,0) + \left[1 - \frac{K(h)}{K(0)}\right] \left[\frac{\partial}{\partial u} C(0,0)\right] C(u,0)$$

L'inverse de Fourier, en u, donne $\left[\rho y + (1-\rho)m\right] f(y)$, de sorte que l'on a

$$E(z.y) = \rho y + (1 - \rho)m$$

6.- DILUTION D'UN PHÉNOMÈNE TRANSITIF.

Considérons maintenant une variable régionalisée $f(x)$ à support ponctuel, nulle en dehors d'un certain champ géométrique V, et soit $g(h)$ son covariogramme transitif. Désignons par $S(a)$ l'aire occupée par les points x tels que $f(x) \leqslant a$ et $f(x) \neq 0$. C'est une mesure positive, de somme finie

$$S(\infty) = V$$

La fonction de répartition

$$F(a) = \frac{1}{V} S(a)$$

donne la probabilité pour qu'en un point x, tiré au sort dans le champ V, de telle manière que tous les points de V aient même probabilité a priori d'être choisis, on ait $f(x) \leqslant a$. La dérivée de $F(a)$ donne l'histogramme des $f(x)$. A cette fonction de répartition $F(a)$ est associée sa fonction caractéristique $C(u)$

$$C(u) = \int e^{iua}\, dF(a) = \frac{1}{V}\int e^{iva}\, dS(a)$$

En revenant à la définition de la mesure $S(a)$, on peut écrire, en intégrant cette fois dans l'espace où se manifeste la régionalisation :

$$(26) \qquad C(u) = 1 + \frac{1}{V}\int \left[e^{iuf(x)} - 1\right] dx$$

Et réciproquement, la fonction caractéristique $C(u)$ écrite en (26) définit parfaitement l'histogramme de la régionalisée $f(x)$.

Au lieu de se limiter au champ V, ensemble des points x tels que $f(x) \neq 0$, il peut y avoir intérêt à diluer la régionalisation dans un champ V', contenant V, et à passer à la limite où $V' = \frac{1}{\lambda}$ devient infini. La nouvelle fonction de répartition $F'(a)$ se déduit de l'ancienne par

$$F'(a) = \frac{V\,F(a) + V' - V}{V'}$$

et la nouvelle fonction caractéristique par :

$$C'(u) = 1 + \lambda \int \left[e^{iuf(x)} - 1\right] dx$$

Si $\lambda = \frac{1}{V'}$ tend vers 0, on obtient, à la limite, le logarithme de la fonction caractéristique sous la forme :

$$(27) \quad \log C(u) = \lambda \int \left[e^{iuf(x)} - 1\right] dx$$

Au lieu des valeurs ponctuelles $f(x)$, on peut considérer des prélèvements représentés par des fonctions φ. On sait que cela revient à remplacer f par $f * \check{\varphi}$. L'histogramme des prélèvements $f * \check{\varphi}$ admet comme fonction caractéristique :

$$C(u) = 1 + \frac{1}{V(\varphi)} \int \left[e^{iuf(x)} - 1\right] dx$$

le champ $V(\varphi)$ étant défini comme l'ensemble des points tels que $f * \check{\varphi} \neq 0$, et étant en général plus grand que V. En diluant dans un champ $V' = \frac{1}{\lambda}$, et en faisant tendre λ vers 0, on obtient, comme en (27)

$$(28) \quad \log C(u) = \lambda \int \left[e^{iuf * \check{\varphi}} - 1\right] dx$$

En comparant à (16), on peut interpréter (28) comme le logarithme de la fonctionnelle caractéristique

$$\log C(\varphi) = \lambda \int \left[e^{if * \check{\varphi}} - 1\right] dx$$

d'une régionalisation intrinsèque poissonnienne T_λ. La fonction de répartition F_λ des $T_\lambda \varphi$ comporte une masse finie à l'origine. Lorsque λ tend vers 0, la limite de

$$\frac{F_\lambda(a) - F_\lambda(0)}{1 - F_\lambda(0)}$$

coïncide aves la fonction de répartition des $f * 0 \neq 0$.

L'opération qui consiste à passer de $f(x)$ à T_λ et à faire tendre λ vers 0 peut être appelée dilution. A la régionalisation $f(x)$ est ainsi associée une régionalisation intrinsèque T_λ telle que les prélèvements $f * \check{\varphi}$ et $T\varphi$ aient même histogramme quelle que soit φ lorsque λ tend vers zéro.

G. MATHERON

Juin 1963