Fontainebleau N-195

NOTE GEOSTATISTIQUE N° 110

PROCESSUS STOCHASTIQUES ET STRATEGIE DE LA PROSPECTION

G. MATHERON

Juin 1970

NOTE GEOSTATISTIQUE n° 110

PROCESSUS STOCHASTIQUES ET STRATEGIE DE LA PROSPECTION

Introduction 1

I - Le cas fini 2

1 - Loi de $Y_1, Y_2, .. Y_N$ 3

2 - Interprétation 4

3 - Exemples des lois gamma 6

4 - Cas de $m = \infty$ 8

II - Le cas infini 9

1 - Passage à la limite 9

2 - Interprétation 11

3 - Remarque 14

4 - Exemple (lois gamma et lois stables) 14

5 - Autre exemple (succès et échecs) 16

III - Le processus conditionné 17

1 - La loi conditionnelle de ν_k 17

2 - Calcul de la fonction auxiliaire R_{k-1} 18

3 - Exemple 19

4 - Loi de Y_{k-1} à $Y_1, .. Y_k$ fixés 21

5 - L'estimation des paramètres 21

IV - Version vulgarisée 22

NOTE GEOSTATISTIQUE N° 110

PROCESSUS STOCHASTIQUES ET STRATEGIE DE LA PROSPECTION

INTRODUCTION

Je me propose, dans ce qui suit, de passer en revue quelques types de processus stochastiques qui puissent servir à représenter le déroulement dans le temps de la prospection, disons pétrolière (pour fixer les idées) d'un territoire donné.

Les auteurs américains (cf. article de G.M. Kaufman) utilisent des modèles statistiques fondés sur un usage exclusif de la loi lognormale, et, ce qui est plus grave, sur un mode de raisonnement "par différence" dont on peut contester sérieusement la légitimité.

D'une part, la forme exacte de cette loi statistique du côté des faibles valeurs n'a aucune importance réelle, de l'autre le nombre N total des gisements existant réellement dans la zone à prospecter n'a probablement pas de signification réelle, précisément parce que la définition de ce qui est gisement et de ce qui ne l'est pas devient arbitraire du côté de ces faibles valeurs. La formulation ci-dessous - inspirée de la théorie des processus à accroissements indépendants et positifs - tend à supprimer ces pseudo-problèmes en admettant, en quelque sorte, qu'il existe dans la zone considérée un nombre infini de gisements mais dont la plupart sont infiniment petits, de sorte que seuls sont définis les nombres $N(y)$ des gisements de taille y supérieure à $y > 0$ donné et leurs lois F_y (qui sont des lois tronquées).

Du fait qu'un gisement donné doit avoir, en gros, une probabilité de découverte proportionnelle à sa taille Y, la distribution empirique des $n(t)$ gisements trouvés au temps t est profondément biaisée, et l'on ne pourra reconstituer les lois a priori qu'en résolvant d'assez délicats problèmes de statistique mathématique. Nous ne ferons guère ci-après que poser ces problèmes, dont la solution exigera des moyens assez puissants.

L'idée de base est de représenter la prospection par un processus (non markovien, on le verra), Y_1, $Y_2, \dots$ $Y_n, \dots$ où Y_n est la taille du $n^{ième}$ gisement trouvé. (il est entendu que $Y_n = 0$ ou, plus généralement Y_n inférieur à un y_o donné signifie que le $n^{ième}$ essai est un échec). Le problème est de déterminer la loi conditionnelle de Y_{k+1}, $Y_{k+2} \dots$ à $Y_1 = y_1, \dots$ $Y_k = y_k$ fixés.

On note qu'on ne s'embarasse pas ici de contraintes géométriques. Si les Y_n représentent les surfaces des gisements (ce qu'est l'interprétation la plus plausible lorsque l'on admet que la probabilité de découverte est proportionnelle à Y_n) on devrait avoir une condition du type $\Sigma Y_n \leq S$ où S est la surface totale de la zone. Mais ces contraintes géométriques sont difficiles à exprimer, et nous n'en tiendrons pas compte. Néanmoins, comme nous le verrons, la théorie des processus à accroissements indépendants nous permettra indirectement d'associer une image géométrique à notre formulation abstraite.

I - Le cas fini

Soit N un nombre entier donné, X_1, X_2 .. X_N N variables aléatoires positives indépendantes de même loi F et d'espérance $m = E(X) < \infty$. Nous désignons par :

$$\Phi(\lambda) = \int_0^\infty e^{-\lambda x} F(dx) \quad \text{la transformée de Laplace de la loi F}$$

$$G(dx) = \frac{x}{m} F(dx) \quad \text{la loi de l'autopondérée de X}$$

$$\gamma(\lambda) = \frac{1}{m}\int_0^\infty e^{-\lambda x} G(dx) = \frac{\Phi'(\lambda)}{\Phi'(o)} \quad \text{la transformée de la loi G}$$

A $X_i = x_i$ fixés, on tire au sort l'une des variables X_1, ... X_N en attribuant la probabilité $X_j/\Sigma X_i$ à la variable X_j. On définit ainsi une variable Y_1 de loi :

$$P(Y_1 = X_{i_1}) = \frac{X_{i_1}}{\Sigma X_i}$$

Par récurrence, pour $k \leqslant N$, on tire au sort Y_k parmi les $N-k+1$ variables X_i restantes, selon la loi :

$$(1) \qquad P(Y_k = X_{i_k}) = \frac{X_{i_k}}{\sum_{i=1}^{N} X_i - \sum_{p=1}^{k-1} X_{i_{k-1}}} \qquad (i_k \neq i_1, i_2 .. \, i_{k-1})$$

Autrement dit, le processus (fini) $Y_1, .. \, Y_n$ est une permutation aléatoire de N variables indépendantes $X_1, .. \, X_N$, construite selon la loi (1) qui attribue à chacune des variables subsistant au temps k une probabilité proportionnelle à sa valeur numérique.

1 - Loi de $(Y_1, .. \, Y_N)$

Soit $(i_1, i_2, .. \, i_N)$ une permutation de $(1, 2, .. \, N)$. Pour $k \leqslant N$, et $x_1, x_2, ... \, x_N$ donnés, l'évènement :

$$\{X_i \in (x_i, x_i + dx_i),\ i = 1,2..N \text{ et } Y_1 = X_{i_1}, ... Y_k = X_{i_k}\}$$

à la probabilité :

$$\frac{x_{i_1}}{x_{i_1} + x_{i_2} + .. + x_{i_N}} \; \frac{x_{i_2}}{x_{i_2} + .. + x_{i_N}} \; \frac{x_{i_k}}{x_{i_k} + .. + x_{i_N}} \; F(dx_{i_1}) F(dx_{i_2}) .. F(dx_{i_N})$$

Par suite, l'évènement :

$$\{Y_i \in (y_i, y_i + dy_i), i=1,2,..k \text{ et } Y_i = X_{i_k}, i=1,2..k\}$$

a pour probabilité :

$$y_1 \, F(dy_1) .. y_k \, F(dy_k) \int \frac{F(dy_{k+1}) .. \, F(dy_N)}{(y_1 + .. + y_N) ... (y_k + .. + y_N)}$$

Il suffit de multiplier par le nombre $\frac{N!}{(N-k)!}$ des choix possibles des k premiers indices $i_1, .. i_k$ de la permutation pour obtenir la loi de $(Y_1, ... Y_k)$, soit :

$$(2) \qquad m^k \frac{N!}{(N-k)!} G(dy_1)..G(dy_k)\int \frac{F(dy_{k+1})}{y_1+..+y_N} \cdots\cdots \frac{F(dy_N)}{y_k+..+y_N}$$

Cherchons la transformée de Laplace de la loi 2 :

$$E(e^{-\lambda_1 Y_1 \dots \lambda_k Y_k}) = m^k \frac{N!}{(N-k)!}\int e^{-\lambda_1 y_1} G(dy_1)..e^{-\lambda_k y_k} G(dy_k)\dots$$

$$\dots F(dy_{k+1})\dots F(dy_N)\int e^{-\mu_1(y_1+\dots+y_N)\dots-\mu_k(y_k+..+y_N)}$$

$$d\mu_1 \dots d\mu_k$$

soit, en échangeant l'ordre des intégrations :

$$(3) \quad E(e^{\sum_{i=1}^{k} \lambda_i Y_i}) = m^k \frac{N!}{(N-1)!}\int \gamma(\lambda_1+\mu_1)..\gamma(\lambda_k+\mu_1+..+\mu_k)\Phi^{N-k}(\mu_1+.+\mu_k)d\mu_1..d\mu$$

On notera, en passant, la relation suivante, valable pour s positif quelconque (entier ou non), qui s'établit par récurrence :

$$\int \gamma(\mu_1)\gamma(\mu_1+\mu_2)..\gamma(\mu_1+..+\mu_k)\Phi^s(\mu_1+..+\mu_k)d\mu_1..d\mu_k =$$

$$= \frac{1}{m^k} \frac{\Gamma(1+s)}{\Gamma(1+s+k)}$$

(Γ est la fonction eulerienne habituelle).

Pour $k = N$, on obtient la loi de la séquence complète $(Y_1,..Y_N)$, soit :

$$(2') \qquad m^N N! \prod_{i=1}^{N} \frac{G(dy_i)}{y_i+y_{i+1}+..+y_N}$$

$$(3') \quad E(e^{\sum_{i=1}^{k} \lambda_i Y_i}) = m^N N!\int \gamma(\lambda_1+\mu_1)\dots\gamma(\lambda_N+\mu_1+..+\mu_N)d\mu_1..d\mu_N$$

2 - Interprétation

Nous allons interpréter la loi (3') en considérant $\mu_1,..\mu_N$ comme des paramètres aléatoires. On note que l'expression :

$$\frac{\gamma(\lambda_k+\mu_1+\ldots+\mu_k)}{\gamma(\mu_1+\ldots+\mu_k)}$$

est la transformée de Laplace de la loi

$$(4) \qquad \frac{e^{-(\mu_1+\ldots+\mu_k)x}\, G(dx)}{\gamma(\mu_1+\ldots+\mu_k)}$$

qui se déduit de G par une pondération exponentielle. On peut donc interpréter (3') en disant que les Y_k sont N variables indépendantes admettant respectivement les lois (4) conditionnellement en $M_1 = \mu_1, \ldots M_N = \mu_N$, les M_k désignent N variables aléatoires positives dont la loi admet la densité :

$$(5) \qquad m^N\, N!\ \gamma(\mu_1)\ \gamma(\mu_1)\ \gamma(\mu_1+\mu_2)\ldots\gamma(\mu_1+\ldots+\mu_N)d\mu_1\ldots d\mu_N$$

Si l'on conditionne seulement sur les k premiers M_i, les variables $Y_1, \ldots Y_k$ ont encore les lois (4) conditionnellement en $M_1 = \mu_1, \ldots M_k = \mu_k$, avec pour $(M_1, \ldots M_k)$ la loi :

$$(5') \qquad m^k \frac{N!}{(N-k)!}\ \gamma(\mu_1)\ldots\gamma(\mu_1+\ldots+\mu_k)\Phi^{(N-k)}(\mu_1+\ldots+\mu_k)$$

On le voit en intégrant (5) en $\mu_{k+1}, \ldots \mu_N$, ou, aussi bien, en partant directement de (3).

Faisons maintenant le changement de variable :

$$(6) \qquad \left|\begin{array}{l} \nu_k = \mu_1+\ldots+\mu_k \\ (0 \leq \nu_1 \leq \nu_2 \leq \ldots \leq \nu_N) \end{array}\right.$$

Les Y_k ont les lois conditionnelles :

$$(7) \qquad \frac{e^{-\nu_k x}}{\gamma(\nu_k)}\ G(dx)$$

les ν_k désignant des v.a. positives <u>ordonnées</u> dont la loi a pour densité :

(7')
$$m^N\, N!\;\; \gamma(\nu_1)\gamma(\nu_2)..\gamma(\nu_N)\; d\nu_1..d\nu_N$$

$$(0 \leqslant \nu_1 \leqslant .. \leqslant \nu_N)$$

La loi des ν_1 admet une interprétation simple. Soient, en effet, $Z_1,..Z_N$ N variables <u>indépendantes</u> admettent la même loi $m\gamma(z)ds$, et $i_1,i_2,...i_N$ la permutation (aléatoire) définie par :

$$Z_{i_1} \leqslant Z_{i_2} \leqslant \cdots \leqslant Z_{i_N}$$

(l'égalité éventuelle dans la relation $\leqslant$ a une probabilité nulle, et n'entraine pas d'ambiguité dans la définition p.s. de la permutation $i_1..i_N$). Alors les variables :

$$\nu_k = Z_{i_k}$$

ont justement la loi (7'). D'où le résultat :

<u>Soient</u> $Z_1,...Z_N$ N <u>V.A. indépendantes de densité</u> $m\gamma(z)$, <u>et</u> $\nu_1,...\nu_N$ <u>les mêmes V.A. rangées par ordre croissant. Conditionnellement</u> à $\nu_1,..\nu_k$ <u>fixés</u>, $Y_1,..Y_k$ <u>sont k variables indépendantes admettant respectivement les lois</u> :

$$\frac{e^{-\nu_i y_i}}{\gamma(\nu_i)}\; G(dy_i) \qquad (i = 1,2,..k)$$

3 - <u>Exemple des lois gamma</u>

Prenons comme loi F la loi gamma (a,α), soit

$$\left|\begin{array}{l} F(dx) = \frac{a^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-ax}\, dx\,, \quad \Phi(\lambda) = (\frac{a}{a+\lambda})^\alpha \\[2ex] G(dx) = \frac{a^{\alpha+1}}{\Gamma(\alpha+1)}\, x^\alpha\, e^{-ax}\, dx\,, \quad \gamma(\lambda) = (\frac{a}{a+\lambda})^{\alpha+1} \\[2ex] m = \frac{\alpha}{a} \end{array}\right.$$

Pour avoir des notations condensées, nous désignerons par $\underline{S_\alpha}$ une variable gamma $(\alpha, 1)$, de densité $\frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x}$. Ici, donc, F est la loi de $\frac{1}{a} S_\alpha$ et G celle de $\frac{1}{a} S_{1+\alpha}$. A ν_k fixé, Y_k a la loi :

$$\frac{e^{-\nu_k x}}{\gamma(\nu_k)} G(dx) \to \frac{\gamma(\lambda+\nu_k)}{\gamma(\nu_k)} = \left(\frac{a+\nu_k}{a+\nu_k+\lambda}\right)^{1+\alpha}$$

donc équivaut en loi à $\frac{1}{a+\nu_k} S_{1+\alpha}$. Ainsi :

$$Y_k \equiv \frac{1}{a+\nu_k} S_{1+\alpha}^{(k)} \qquad (k=1,2,..N) \tag{8}$$

avec des $S_{1+\alpha}^{(k)}$ indépendantes, et des ν_k dont la loi est :

$$N!\ \alpha^N \left(\frac{1}{a+\nu_1}\right)^{1+\alpha} \cdots \left(\frac{1}{a+\nu_N}\right)^{1+\alpha} d\nu_1 .. d\nu_N$$

$$(0 \leq \nu_1 \leq \cdots \leq \nu_N)$$

Posons :

$$B_k = \frac{a}{a+\nu_k}$$

Ces B_k ont la densité :

$$N!\ \alpha^N B_1^{\alpha-1} B_2^{\alpha-1} \cdots B_N^{\alpha-1} = N!\ \alpha^N B_1^{\alpha N-1} \frac{B_2^{\alpha(N-1)-1}}{B_1^{\alpha(N-1)}} .. \frac{B_N^{\alpha-1}}{B_{N-1}^{\alpha}}$$

$$(1 \geq B_1 \geq B_2 \geq \cdots \geq B_N \geq 0)$$

Mais, à B_{k-1} fixé, la loi de densité

$$\alpha(N-k+1) \frac{B_k^{\alpha(N-k+1)-1}}{B_{k-1}^{\alpha(N-k+1)}} \qquad (0 \leq B_k \leq B_{k-1})$$

est celle du produit $B_{k-1} \Pi_k$, où Π_k a la densité :

(9) $$\alpha(N-k+1)\,\varpi_k^{\alpha(N-k+1)-1} \qquad (0 \leq \varpi_k \leq 1)$$

Autrement dit, on peut écrire :

$$B_k \equiv \Pi_1 \; \Pi_2 \ldots \Pi_k$$

avec des variables Π_i indépendantes, dont les lois admettent les densités :

$$\alpha(N-i+1)\varpi_i^{\alpha(N-i+1)-1} \qquad (i = 1,2,..k)$$

En résumé, dans le cas où F est la loi gamma (a,α), on a les équivalences :

(10) $$Y_k \equiv \frac{1}{a}\Pi_1 \; \Pi_2 \; .. \; \Pi_k \; S_{1+\alpha}^{(k)} \qquad (k = 1,2,..N)$$

où les Π_k sont N variables indépendantes admettant les lois (9), et les $S_{1+\alpha}^{(k)}$ N variables gamma indépendantes entre elles et indépendantes des Π_k.

Cette relation (10) permet de construire de proche en proche une réalisation du processus sans qu'il soit nécessaire de fixer au préalable les N variables Y_k. Nous verrons que le procédé se généralise au cas infini.

4 - Cas de $m = \infty$

Dans ce qui précède, on a supposé $m = \int_0^\infty x \, F(dx) < \infty$, mais cette hypothèse ne joue au fond aucun rôle. Désignons, en effet, par H la mesure positive :

$$H(dx) = x \, F(dx)$$

et par $\eta(\lambda)$ sa transformée de Laplace :

$$\eta(\lambda) = - \Phi'(\lambda)$$

On a $m = \eta(0) < \infty$ si et seulement si la mesure H est sommable. Il est aisé de reprendre les calculs faits en 1/ en remplaçant mG par H et $m\gamma$ par η, et toutes les conclusions subsistent :

Soient $M_1, .. M_k$ k variables admettant la loi :

$$(11) \quad \frac{N!}{(N-k)!} \eta(\mu_1)\ \eta(\mu_1+\mu_2)..\eta(\mu_1+...+\mu_k)\Phi^{(N-k)}(\mu_1+..+\mu_k)d\mu_1..d\mu_k$$

A $M_1 = \mu_1 .. M_k = \mu_k$ fixés, les k premières variables Y_n sont indépendantes, et admettent les lois :

$$(11') \quad \frac{e^{-(\mu_1+...+\mu_n)y_n}\ H(dy_n)}{\eta(\mu_1+...+\mu_n)}$$

Autrement dit, encore, on a :

$$(11'') \quad E(e^{-\lambda \sum_{n=1}^{k} Y_k}) = \frac{N!}{(N-k)!}\int \eta(\lambda_1+\mu_1)...\eta(\lambda_k+\mu_1+..+\mu_k)$$
$$\Phi^{(N-k)}(\mu_1+..+\mu_k)d\mu_1..d\mu_2$$

II - <u>Le cas infini</u>

1 - <u>Passage à la limite</u>

Nous allons maintenant faire tendre vers l'infini le nombre N des "gisements" et choisir la loi F de manière à ce que la somme $X_1+..+X_N$ reste p.s. finie. Cette loi limite sera nécessairement indéfiniment divisible, et nous pouvons donc supposer dès le départ que la transformée Φ de F est de la forme :

$$\Phi(\lambda) = e^{-t\psi(\lambda)}, \quad \psi(\lambda) = \int_0^\infty \frac{1-e^{-\lambda x}}{x}\ G(dx)$$

où G désigne cette fois une mesure <u>positive</u> sur $(0,\infty)$, et t un nombre positif. On a :

$$\eta(\lambda) = t\ \psi'(\lambda)\ e^{-t\psi(\lambda)}$$

et on remarque que $\psi'(\lambda)$ est la transformée de la mesure G. Donc la mesure G est sommable si et seulement si la loi F admet une espérance finie.

Portons ces expressions de Φ et η dans (11"). On trouve pour la loi des k premières Y_n :

$$\frac{t^k N!}{(N-k)!}\int \psi'(\lambda_1 + \mu_1)..\psi'(\lambda_k + \mu_1 +...+ \mu_k)\, e^{-t\psi(\lambda_1+\mu_1)...-t\psi(\lambda_k+\mu_1+..+\mu_k)}$$

$$e^{-(N-k)t\ \psi(\mu_1+..+\mu_k)} \quad d\mu_1...d\mu_k$$

Si $N \to \infty$, $t \to 0$ avec $N_t \to \theta < \infty$, $t^k \frac{N!}{(N-k)!}$ tend vers θ^k et le théorème de Lebesgue montre que l'on peut passer à la limite sous le signe d'intégration. La loi des k premières variables Y_n du processus infini obtenu par ce passage à la limite est donc :

$$(12) \quad E(e^{-\sum_{n=1}^{k} \lambda_n Y_n}) = \theta^k \int \psi'(\lambda_1+\mu_1)...\psi'(\lambda_k+\mu_1+..+\mu_k)e^{-\theta\psi(\mu_1+..+\mu_k)}\, d\mu_1..d\mu_k$$

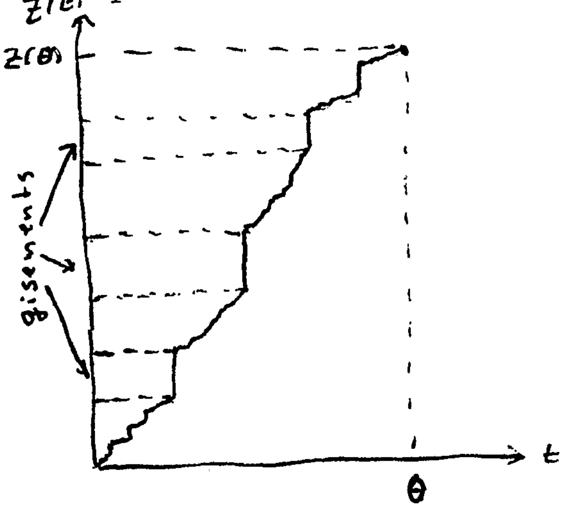
Cette loi décrit entièrement le processus Y_n - c'est-à-dire le déroulement d'une campagne de prospection indéfiniment prolongée. La somme :

$$\sum_{n=1}^{\infty} Y_n$$

est p.s. $< \infty$, et admet la loi $e^{-\theta\ \psi(\lambda)}$. Le <u>nombre</u> N(y) des gisements de taille $Y > y$ est une variable de Poisson de paramètre

$$\theta(y) = \theta \int_y^{\infty} \frac{G(dx)}{x}$$

Si l'intégrale $\int_0^{\infty} \frac{G(dx)}{x}$ est <u>convergente</u>, N(o), c'est-à-dire le nombre des gisements de taille > 0, est une variable de Poisson, donc est p.s. fini. Dans le processus de prospection Y_n, les Y_n sont donc p.s. tous nuls au delà d'un rang fini. Au contraire, si l'intégrale diverge, les Y_n ne s'annulent pas au-delà d'un rang fini (bien que leurs espérances aillent en décroissant). C'est, en principe, ce cas qui nous intéresse.



Géométriquement, on peut désigner par $Z(t)$ un processus à accroissements positifs et indépendants, de loi

$$E(e^{-\lambda Z(t)}) = e^{-t\,\psi(\lambda)}$$

Les "gisements" sont les sauts de ce processus entre 0 et θ. En particulier

$$Z(\theta) = \sum_{n=0}^{\infty} Y_n$$

La zone à prospecter est imagée par le segment $[0, Z(\theta)]$ de l'axe des z, que les sauts du processus découpent selon une partition représentant les gisements. Le processus de prospection consiste à implanter au hasard des points sur $[0, Z(t)]$, en retirant, après chaque essai, le saut Y_n qui vient d'être touché.

2 - Interprétation

Comme dans le paragraphe précédent, on peut interpréter la loi (12) à l'aide d'un processus auxiliaire $M_1, M_2, \ldots$. Conditionnellement à $M_1 = \mu_1, \ldots M_k = \mu_k$ fixés, les $Y_1 \ldots Y_k$ ont la transformée de Laplace :

$$\frac{\gamma(\lambda_1 + \mu_1)}{\gamma(\mu_1)} \cdots \frac{\gamma(\lambda_k + \mu_1 + \ldots + \mu_k)}{\gamma(\mu_1 + \ldots + \mu_k)} \qquad (\gamma(\lambda) = \psi'(\lambda))$$

Ce sont donc k V.A. indépendantes, la loi de Y_i $(1 \leq i \leq k)$ admettant la loi :

$$\frac{e^{-(\mu_1 + \ldots + \mu_i)y}}{\gamma(\mu_1 + \ldots + \mu_i)} \; G(dy)$$

qui se déduit de G par une pondération exponentielle. Le processus auxiliaire $M_1, M_2, \ldots$ est défini par sa loi temporelle qui est la suivante : pour tout entier k, la variable vectorielle $(M_1, \ldots M_k)$ admet la densité :

$$\theta^k \gamma(\mu_1)\gamma(\mu_1 + \mu_2)\ldots \gamma(\mu_1 + \ldots \mu_k)\, e^{-\theta\, \psi(\mu_1+\ldots+\mu_k)}$$

Le changement de variable (6) donne pour les Y'_n les lois conditionnelles :

$$\frac{\theta^{-\nu_n y}}{\gamma(\nu_n)}\; G(dy) \tag{12}$$

avec (pour tout entier k) des $\nu_1,\ldots \nu_k$ aléatoires admettant la densité :

$$\left|\begin{array}{l} \theta^k\, \gamma(\nu_1)\; \gamma(\nu_2)\ldots \gamma(\nu_k)\, e^{-\theta\, \psi(\nu_k)} \\ \\ (0 \leqslant \nu_1 \leqslant \nu_2 \leqslant \ldots) \end{array}\right.$$

Mais $\gamma(\lambda) = \psi'(\lambda)$. Si nous posons :

$$1 - K(x) = e^{-\theta\, \psi(x)}$$

$K(x)$ est la fonction de répartition d'une loi de probabilité :

$$K(dx) = \theta\, \gamma(x)\, e^{-\theta\, \psi(x)}\, dx$$

(puisque $\psi(x) = \int_0^x \gamma(\lambda)d\lambda$ est une fonction croissante).

La loi (12') des k premiers paramètres ν_n se met donc sous la forme :

$$K(d\nu_1)\; \frac{K(d\nu_2)}{1-K(\nu_1)} \cdots \frac{K(d\nu_k)}{1-K(\nu_{k-1})}$$

$$(0 \leqslant \nu_1 \leqslant \ldots \leqslant \nu_k)$$

Ainsi, ν_1 obéit à la loi K ; à ν_1 fixé, ν_2 obéit à la loi $\frac{K(d\nu_2)}{1-K(\nu_1)}$: cette loi est celle d'une variable ν dont la loi a priori serait K, mais qui serait prise conditionnellement en $\nu \geqslant \nu_1$. Autrement dit encore, ψ étant croissante, posons :

$$X_n = \theta\, \psi(\nu_n)$$

On a :

$$P(X_1 \geqslant x_1) = 1-K(\nu_1) = e^{-\theta\, \psi(\nu_1)} = e^{-x_1}$$

Puis :

$$P(X_2 \geqslant x_2 | X_1 = x_1) = \frac{1-K(\nu_2)}{1-K(\nu_1)} = e^{-\theta\phi(\nu_2)+\theta\phi(\nu_1)} = e^{-(x_2-x_1)}$$

Ainsi, les $X_n = \theta\ \phi(\nu_n)$ constituent <u>un processus de Poisson</u> de paramètre 1. On a :

$$X_1 = S_1$$
$$\cdots\cdots$$
$$X_n = S_1 + S_2 + \ldots + S_n$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

les S_n désignant des variables exponentielles indépendantes, de lois e^{-s}. D'où le procédé suivant qui permet d'obtenir <u>tous les processus</u> du type qui nous intéresse ici :

a/ <u>choisir une mesure positive G(dx)</u> sur la demi droite (o,∞) vérifiant $\int_1^\infty \frac{G(dx)}{x} < \infty$, et poser :

$$\gamma(\lambda) = \int_0^\infty e^{-\lambda x}\ G(dx),\quad \phi(\lambda) = \int_0^\lambda \gamma(\mu)d\mu$$

b/ <u>choisir un nombre $\theta \geqslant 0$, et se donner une suite</u> S_n <u>de variables indépendantes admettant la même loi de densité</u> e^{-s}. <u>Définir le processus croissant</u> $0 \leqslant \nu_1 \leqslant \nu_2 \leqslant \ldots$ <u>par la relation</u> :

$$\phi(\nu_n) = \frac{S_1 + \ldots S_n}{\theta}$$

c/ <u>Conditionnellement</u> à $\nu_1, \nu_2, \ldots \nu_k$ fixés, <u>les k premières variables du processus Y_n sont alors indépendantes</u>, et la loi de Y_n est :

$$\frac{e^{-\nu_n y}}{\gamma(\nu_n)}\ G(dy)$$

Ce procédé se prête particulièrement bien à des études par <u>simulation</u>. Rappelons l'interprétation des éléments constitutifs du processus Y_n :

d/ La loi a priori de la somme $\sum_{n=1}^{\infty} Y_n$ admet la transformée $e^{-\theta\phi(\lambda)}$.

Son espérance :

$$E\left(\sum_{n=1}^{\infty} Y_n\right) = \theta\ \psi'(o) = \theta \int_o^{\infty} G(dx)$$

est finie si et seulement si la mesure G est sommable. Pour un y donné, le nombre N(y) des gisements de taille $> y$ est un Poisson de paramètre $\theta \int_y^{\infty} \frac{G(dx)}{x}$.

3 - Remarque

Le processus (Y_n, S_n) est markovien, mais Y_n ne l'est évidemment pas. En vue des applications, l'un des problèmes qu'il faudra résoudre va consister à trouver la loi conditionnelle de $Y_{k+1}, Y_{k+2} \ldots$ à $Y_1, \ldots Y_k$ fixés (puisque c'est de cette loi que dépend la décision de continuer la prospection ou de l'arrêter au temps k). (l'autre problème essentiel sera celui de l'inférence statistique : estimer θ et G à partir des k premières valeurs numériques $y_1, \ldots y_k$).

Lorsque :

$$\psi(\nu_k) = T_k = \frac{S_1 + \ldots + S_k}{\theta}$$

(ou, ce qui revient au même, ν_k lui-même) est connu, le processus défini par $Y'_1 = Y_{k+1}$, $Y'_2 = Y_{k+2}, \ldots$ est du même type que le processus initial, mais avec des lois modifiées. Posons, en effet :

$$\left| \begin{aligned} G_k(dx) &= e^{-\nu_k x}\ G(dx) \\ \psi_k(\lambda) &= \psi(\lambda+\nu_k) - \psi(\nu_k) \end{aligned} \right.$$

Le processus construit à partir de ψ_k est identique en loi au processus $Y'_n = Y_{n+k}$.

On voit que le problème de l'inférence statistique se ramènera essentiellement à l'estimation de ν_k à $y_1, \ldots y_k$ connus.

4 - Exemple (lois gamma et lois stables)

Prenons G la mesure de densité $\frac{x^{\alpha-1}}{\Gamma(\alpha)}\ e^{-ax}$ (avec $a \geqslant 0$).

Le cas le plus intéressant est $\alpha \leq 1$ (mesure $\frac{G}{x}$ non sommable, donc une infinité de gisements potentiels). On a :

$$\gamma(\lambda) = \frac{1}{(a+\lambda)^{\alpha}}$$

$$\psi(\lambda) = \frac{1}{1-\alpha} \left[(a+\lambda)^{1-\alpha} - a^{1-\alpha}\right] \qquad (a \neq 1)$$

$$= \log \frac{a+\lambda}{a} \qquad (\alpha = 1)$$

Pour $a = 0$ et $\alpha < 1$, le processus $e^{-t\psi(\lambda)}$ est un processus à loi stable. Pour $\alpha = 1$ et $a > 0$, c'est un processus à loi gamma. Effectuons le changement de variable :

$$\psi(\nu) = T$$

Soit :

$$\left| \begin{array}{ll} a + \nu = \left[(1-\alpha)T + a^{1-\alpha}\right]^{1/1-\alpha} & (\alpha \neq 1) \\ a + \nu = a\, e^{T} & (\alpha = 1) \end{array} \right.$$

A ν_k fixé, Y_k obéit à la loi de densité :

$$\frac{(a+\nu_k)^{\alpha}}{\Gamma(\alpha)} \; y^{\alpha-1} \; e^{-(a+\nu_k)\, y}$$

Soit :

$$Y_k \equiv \frac{1}{a + \nu_k} \; S_{\alpha}$$

où S_{α} est une variable gamma réduite. Donc, en désignant par $T_1, \dots T_n$ les variables $T_n = \frac{S_1 + \dots + S_n}{\theta}$ associées au processus de Poisson 2/-b, le processus Y_n est défini par :

$$(13) \qquad \left| \begin{array}{ll} Y_k = \dfrac{S_{\alpha}^{(k)}}{((1-\alpha)T_k + a^{1-\alpha})^{1/1-\alpha}} & (\alpha \neq 1) \\ Y_k = \dfrac{1}{a}\, S^{(k)}\, e^{-T_k} & (\alpha = 1) \end{array} \right.$$

où les $S_\alpha^{(k)}$ désignent une suite de variables gamma de paramètre α indépendantes les unes des autres et des T_k.

Dans le cas $\alpha = 1$, on a :

$$e^{-T_k} = e^{-\frac{S_1+\ldots+S_k}{\theta}} = \pi_1 \, \pi_2 \ldots \pi_k$$

et les π_n forment une suite de variables indépendantes admettant la même densité :

$$(14) \qquad \theta \, \varpi^{\theta-1} \qquad (0 \leq \varpi \leq 1)$$

Les Y_k eux-mêmes sont de la forme :

$$(14') \qquad Y_k = \frac{1}{a} \, S^{(k)} \, \pi_1 \, \pi_2 \ldots \pi_k$$

avec des variables $S^{(k)}$ exponentielles indépendantes entre elles et des π_k. Ce cas est le plus simple que l'on puisse envisager.

5 - Autre exemple (succès et échecs)

Prenons une mesure G sommable de la forme :

$$\left| \begin{array}{l} G = q \, \delta + p \, H \\ p+q = 1, \displaystyle\int_0^\infty H(dx) = 1 \end{array} \right.$$

Cette loi comporte des zéros avec des probabilités non nulles, et correspond peut-être mieux au déroulement d'une prospection réelle. On a ici :

$$\left| \begin{array}{l} \gamma(\lambda) = q + p \, \eta(\lambda) \\ \phi(\lambda) = q\lambda + p \displaystyle\int_0^\lambda \eta(\mu) d\mu \end{array} \right.$$

A ν_k fixé, Y_k obéit à la loi :

$$\frac{e^{-\nu_k y}\ (q\ \delta + p\ H)}{q + p\ \eta(\nu_k)}$$

Autrement dit, on a :

$Y_k = 0$ avec la probabilité $\dfrac{q}{q + p\ \eta(\nu_k)}$

et, avec la probabilité $\dfrac{p\ \eta(\nu_k)}{q + p\ \eta(\nu_k)}$, Y_k est une variable de loi :

$$\frac{e^{-\nu_k\ y}}{\eta(\nu_k)}\ H(dy)$$

La suite des ν_n se déduit du processus de Poisson :

$$T_n = \frac{S_1 + \dots + S_n}{\theta}$$

par le changement de variable

$$q\ \nu_n + \int_o^{\nu_n} \eta(\mu)\ d\mu = T_n$$

III - Le processus conditionné

D'après la remarque II-3/, le processus $Y'_n = Y_{n+k}$ conditionné par $Y_1 = y_1, \dots\ Y_k = y_k$ ne dépend que de ν_k. Ainsi, la prédiction des résultats de la prospection postérieure à l'essai N° k ne fait intervenir que la loi conditionnelle de ν_k à $Y_1, \dots\ Y_k$ fixés.

1 - La loi conditionnelle de ν_k à $Y_1, \dots\ Y_k$ fixés

La loi a priori de $Y_1, \dots\ Y_k$, $\nu_1, \dots\ \nu_k$ est, comme on l'a vu :

$$\theta^k\ \frac{e^{-\nu_1 y_1}\ G(dy_1)}{\gamma(\nu_1)} \dots \frac{e^{-\nu_k y_k}\ G(dy_k)}{\gamma(\nu_k)}\ \gamma(\nu_1)\dots\gamma(\nu_k)\ e^{-\theta\phi(\nu_k)}\ d\nu_1 \dots d\nu_k$$

$$(\nu_1 \leq \nu_2 \leq \dots \leq \nu_k)$$

Ainsi, la loi conditionnelle de ν_k à $Y_1 = y_1, \ldots Y_k = y_k$ admet la densité :

$$\frac{e^{-\nu_k y_k} \quad e^{-\theta\phi(\nu_k)} \int_0^{\nu_k} e^{-\nu_{k-1} y_{k-1}} \, d\nu_{k-1} \int_0^{\nu_{k-1}} \cdots \int_0^{\nu_2} e^{-\nu_1 y_1} d\nu_1}{\int_0^{\infty} e^{-\nu_k y_k} \quad e^{-\theta\phi(\nu_k)} \, d\nu_k \int_0^{\nu_k} \cdots \int_0^{\nu_2} e^{-\nu_1 y_1} \, d\nu_1}$$

Introduisons la fonction auxiliaire :

$$\text{(III,1)} \qquad R_{k-1}(\nu; y_1, \ldots y_{k-1}) = \int_0^{\nu} e^{-\nu_{k-1} y_{k-1}} \, d\nu_{k-1} \int_0^{\nu_{k-1}} \cdot\cdot \int_0^{\nu_2} e^{-\nu_1 y_1} \, d\nu_1$$

(qui ne dépend pas de la loi du processus). On voit que la loi conditionnelle de ν_k admet la densité :

$$\text{(III,2)} \qquad \frac{e^{-\nu_k y_k - \theta\phi(\nu_k)} \, R_{k-1}(\nu_k; \, y_1, \, \ldots \, y_{k-1})}{\int e^{-\nu_k y_k - \theta\phi(\nu_k)} \, R_{k-1}(\nu_k; \, y_1, \, \ldots \, y_{k-1}) \, d\nu_k}$$

2 - Calcul de la fonction auxiliaire R_{k-1}

Cherchons la transformée de Laplace (en ν) de la fonction R_{k-1}, soit :

$$\int_0^{\infty} e^{-\lambda\nu} \, d\nu \int_0^{\nu} e^{-\nu_{k-1} y_{k-1}} \, d\nu_{k-1} \int_0^{\nu_{k-1}} \cdot\cdot \int_0^{\nu_2} e^{-\nu_1 y_1} \, d\nu_1 =$$

$$\int_0^{\infty} e^{-\lambda(\mu_1 + \ldots + \mu_k)} \, d\mu_k \int_0^{\infty} e^{-y_{k-1}(\mu_1 + \ldots + \mu_{k-1})} d\mu_{k-1} \cdots \int_0^{\infty} e^{-\mu_1 y_1} \, d\mu_1 =$$

$$= \int e^{-(y_1 + \ldots + y_{k-1} + \lambda) \, \mu_1 - (y_2 + \ldots + y_{k-1} + \lambda) \, \mu_2 \ldots (y_{k-1} + \lambda) \, \mu_{k-1} + \lambda\mu_k} \, d\mu_1 \, d\mu_2 \ldots d\mu_k =$$

$$= \frac{1}{\lambda} \prod_{i=1}^{k-1} \frac{1}{y_i + y_{i+1} + \ldots + y_{k-1} + \lambda}$$

Cette expression se décompose en fractions rationnelles :

$$\frac{1}{\lambda}\prod_{i=1}^{k-1}\frac{1}{y_i+\dots+y_{k-1}+\lambda}=\sum_{i=1}^{k-1}\frac{A_i}{y_i+\dots+y_{k-1}+\lambda}+\frac{A_k}{\lambda}$$

avec :

$$\left|\begin{aligned} A_i &= (-1)^{k-i}\frac{1}{y_1+\dots+y_{i-1}}\,\frac{1}{y_2+\dots+y_{i-1}}\cdots\frac{1}{y_{i-1}}\,\frac{1}{y_i}\,\frac{1}{y_i+y_{i+1}}\cdots\frac{1}{y_i+\dots+y_{k-1}} \\ A_k &= \frac{1}{y_1+\dots+y_{k-1}}\cdots\cdots\frac{1}{y_{k-2}+y_{k-1}}\,\frac{1}{y_{k-1}} \end{aligned}\right.$$

On trouve, par suite :

$$\text{(III-3)}\quad R_{k-1}(\nu;y_1,\dots y_{k-1})=\sum_{i=1}^{k-1}A_i\,e^{-(y_i+\dots+y_{k-1})\nu}+A_k$$

D'où l'expression de la densité conditionnelle (III-2) :

$$\text{(III-4)}\quad\left\{\begin{aligned} &\frac{e^{-\nu_k y_k-\theta\phi(\nu_k)}\,R_{k-1}(\nu_k;y_1,\dots y_{k-1})}{\sum_{i=1}^{k}A_i\,Q\,(y_i+\dots+y_k)} \\ &Q(y)=\int_0^\infty e^{-y\nu-\theta\phi(\nu)}\,d\nu \end{aligned}\right.$$

La transformée de Laplace de cette loi conditionnelle est alors :

$$E(e^{-\lambda\nu_k}|y_1,\dots y_k)=\frac{\sum_{i=1}^{k}A_i\,Q(y_i+\dots+y_k+\lambda)}{\sum_{i=1}^{k}A_i\,Q(y_i+\dots+y_k)}$$

3 - Exemples

Plaçons nous dans le cas du processus multiplicatif (14'), avec :

$$\frac{a}{a+\nu_k}=\pi_1\,\pi_2\dots\pi_k$$

L'espérance de Y_{k+1} à $Y_1,\dots Y_k$ fixés est :

$$E(Y_{k+1}|y_1,\dots y_k) = \frac{1}{a}\ \frac{\theta}{\theta+1}\ E(\frac{a}{a+\nu_k}\ |y_1,\dots y_k)$$

et tout se ramène au calcul de l'espérance conditionnelle de $\frac{a}{a+\nu_k}$ qui est :

$$E(\frac{a}{a+\nu_k}\ |y_1\dots y_k) = a\int_0^\infty e^{-a\lambda}\ E(e^{-\lambda\nu_k}|y_1,\dots y_k)d\lambda$$

Mais on a ici :

$$e^{-\theta\phi(z)} = (\frac{a}{a+z})^\theta$$

d'où

$$\int_0^\infty e^{-a\lambda}\ Q(y+\lambda)d\lambda = \iint e^{-a\lambda-(y+\lambda)z-\theta\phi(z)}\ dz =$$

$$= \int_0^\infty \frac{1}{a+z}\ e^{-yz-\theta\phi(z)}\ dz = \frac{1}{a}\int_0^\infty e^{-yz}\ (\frac{a}{a+z})^{\theta+1}\ dz$$

D'où l'espérance conditionnelle :

$$E(\frac{a}{a+\nu_k}\ ;\ y_1,\dots y_k) = \frac{\sum_1^k A_i\ U_{1+\theta}\ (y_i+\dots+y_k)}{\sum_1^k A_i\ U_\theta\ (y_i+\dots+y_k)}$$

$$U_\theta(y) = \int_0^\infty e^{-yz}\ (\frac{a}{a+z})^\theta\ dz = \frac{a^\theta}{\Gamma(\theta)}\int_0^\infty \frac{x^{\theta-1}\ e^{-ax}}{x+y}\ dy$$

De la même manière, on trouve pour le moment conditionnel d'ordre n :

$$E(\frac{a}{a+\nu_k}\ ;\ y_1,\dots y_k) = \frac{\sum A_i\ U_{n+\theta}(y_i+\dots+y_k)}{\sum A_i\ U_\theta\ (y_i+\dots+y_k)}$$

Il suffit donc de tabuler les fonction $U_{\theta+n}(y)$ et de calculer numériquement les A_i pour obtenir les moments successifs du facteur multiplicatif $\frac{a}{a+\nu_k}$ donc, en particulier, les moments de Y_{k+1} conditionnés par $y_1 = y_1,\dots Y_k = y_k$.

4 - Loi de Y_{k+1} à $Y_1,..Y_k$ fixés

Donnons encore, dans le cas général, l'expression de la loi conditionnelle de Y_{k+1} à $Y_1 = y_1,...Y_k = y_k$ fixés. A ν_k fixé, la loi de ν_{k+1} est :

$$\frac{\theta\phi'(\nu_{k+1})\, e^{-\theta\phi(\nu_{k+1})}}{e^{-\theta\phi(\nu_k)}}\, d\nu_{k+1} \qquad (\nu_{k+1} \geqslant \nu_k)$$

et celle de Y_{k+1} admet la transformée de Laplace :

$$\theta\int_{\nu_k}^{\infty} \frac{\phi'(\lambda+\nu_{k+1})}{\phi'(\nu_{k+1})}\ \phi'(\nu_{k+1})\, e^{\theta\phi(\nu_k)-\theta\phi(\nu_{k+1})}\, d\nu_{k+1}$$

Compte tenu de (III-4), et en remarquant :

$$\int_0^{\nu_{k+1}} e^{-\nu_k y_k}\, R_{k-1}(\nu_k;y_1..y_{k-1}) = R_k(\nu_{k+1};y_1,..y_k)$$

on voit que la loi conditionnelle de Y_{k+1} à $Y_1 = y_1,..\ Y_k = y_k$ admet la transformée :

$$\text{(III-5)} \qquad \frac{\theta\displaystyle\int_0^{\infty} \phi'(\lambda+\nu_{k+1})\, e^{-\theta\phi(\nu_{k+1})}\, R_k(\nu_{k+1}\ ;\ y_1,..y_k)d\nu_{k+1}}{\displaystyle\sum_{i=1}^{k} A_i\ Q(y_1+..+y_k)}$$

On en déduit sans peine l'espérance et la variance conditionnelle de Y_{k+1}.

5 - L'estimation des paramètres

Pour estimer les paramètres (θ, et les paramètre de ϕ), on peut songer à la méthode du maximum de vraisemblance. Pour celà, formons d'abord l'expression de la loi a priori de $Y_1,..Y_k$. En supposant que la mesure G admette la densité g, cette loi a pour densité :

$$g(y_1)\ldots g(y_k)\int_0^\infty e^{-\theta\phi(\nu_k)-\nu_k y_k}d\nu_k\int_0^{\nu_k} e^{-\nu_{k-1}y_{k-1}}d\nu_{k-1}\int_0^{\nu_{k-1}}\cdots\int_0^{\nu_2} e^{-\nu_1 y_1}dy_1 =$$

$$= g(y_1)\ldots g(y_k)\sum_{i=1}^{k} A_i\ Q(y_i+\ldots+y_k)$$

avec la fonction Q définie en (III-4).

La loi g dépend d'un certain nombre de paramètres, soient $a_1, a_2, \ldots a_p$. Q dépend de ces mêmes paramètres, et aussi de θ. L'estimateur du maximum de vraisemblance, pour θ et les a_i, est donc la solution du système :

$$\sum_i A_i \frac{\partial}{\partial\theta} Q(y_i+\ldots+y_k) = 0$$

$$\sum_{r=1}^{k}\frac{1}{g(y_r)}\ \frac{\partial}{\partial a_j}\ g(y_r) + \frac{\sum\limits_{i=1}^{k} A_i\ \frac{\partial Q}{\partial a_j}\ (y_i+\ldots+y_k)}{\sum A_i\ Q(y_i+\ldots+y_k)} = 0 \quad (j=1,2,\ldots p)$$

Il est probable que, dans les applications, on utilisera des estimateurs moins bons mais plus faciles à former.

IV - Version vulgarisée

Bien qu'ils ne soient pas insolubles, les problèmes posés par l'inférence statistique et le passage au processus conditionné sont relativement ardus et conduisent à des calculs numériques assez longs. Dans les applications pratiques, on sera peut-être conduit à remplacer les processus précédents par des processus plus faciles à manipuler (mais moins satisfaisants sur le plan théorique). Par exemple, en démarquant (14'), on pourrait poser à priori :

(IV-1) $$Y_k = Z_k\ \Pi_1\ \Pi_2 \ldots \Pi_k$$

où les Z_k et les Π_k seraient indépendantes, avec pour les Z_k une loi lognormale de médiane γ et de variance σ^2, et pour les Π_k une autre loi lognormale de médiane γ' et de variance σ'^2, avec :

$$E(\Pi) = \gamma'\ e^{\sigma^2/2} < 1$$

(puisque les Y_k doivent décroitre en moyenne). L'avantage évident de ce processus est qu'en passant aux logarithmes on n'a plus affaire qu'à des combinaisons linéaires gaussiennes. Ce point pourra faire l'objet de développements ultérieurs.

ADDITIF A LA NOTE GEOSTATISTIQUE N° 110

I - Schémas des échecs et succès.

Considérons le processus défini page 16 dans la note 110, et désignons par s_1, s_2,... les valeurs de l'indice n correspondant aux succès ($Y_1 = Y_2 = \dots = Y_{s_1 - 1} = 0$, $Y_{s_1} \neq 0$, $Y_{s_1+1} = {}$ $= \dots = 0$ etc...). En posant :

$$Z_n = Y_{s_n}$$

on définit un nouveau processus, qui se déduit du processus initial Y_n en ne conservant que les succès. Il est clair, d'ailleurs, que le processus Z_n apporte moins d'information, au total, que le processus initial, puisque la considération de la longueur des séquences d'échecs successifs est certainement elle aussi instructive. Montrons seulement ici que le processus Z_n est encore du type défini dans la note 110 (avec G remplacé par H et θ par θp, selon les notations de la note 110, p. 16).

Posons :

$$\tau_n = \nu_{s_n}$$

(valeur du paramètre aléatoire ν associé au $n^{\text{ième}}$ succès). A $\tau_1, \dots \tau_n$ fixés, les $Z_1, \dots Z_n$ sont indépendants et Z_k admet la loi

$$\frac{e^{-\tau_k z} H(dz)}{\eta(\tau_k)}$$

comme on l'a déjà indiqué dans la note 110. Le problème consiste donc à déterminer la loi des $\tau_n = \nu_{s_n}$.

A ν_1, ν_2,... fixés, on a $s_1 = 1$ avec la probabilité $\frac{p\eta(\nu_1)}{q + p\eta(\nu_1)}$ et, plus généralement :

$$p(s_1 = r) = \prod_{i=1}^{r-1} \frac{q}{q + p\eta(\nu_i)} \times \frac{p\eta(\nu_r)}{q + p\eta(\nu_r)}$$

D'autre part, les variables

$$\nu_1 \leq \dots \leq \nu_r$$

admettent la densité

$$\theta^r \gamma(\nu_1) \dots \gamma(\nu_r) \; e^{-\theta\phi(\nu_r)}$$

(avec $\gamma(\nu) = q + p\eta(\nu)$). On en déduit que la probabilité d'avoir :

$$s_1 = r \quad \text{et} \quad \tau_1 \in (\tau, \tau + d\tau)$$

est :

$$e^{-\theta\phi(\tau)} p \, q^{r-1} \, \theta^r \, \eta(\tau) \, d\tau \int_0^\tau d\nu_{r-1} \int_0^{\nu_{r-1}} d\nu_{r-2} \dots \int_0^{\nu_2} d\nu_1 =$$

$$q^{r-1} \, \theta^r p\eta(\tau) \, \frac{\tau^{r-1}}{(r-1)!} \, e^{-\theta\phi(\tau)} \, d\tau$$

On en déduit :

a/ A $\tau_1 = \tau$ fixé, le temps d'attente $s_1 - 1$ du premier succès est une variable de Poisson de paramètre :

$$E(s_1 | \tau_1) = q\ \theta\ \tau_1$$

b/ La loi de τ_1 s'obtient en sommant de $r = 1$ à l'infini la densité écrite ci-dessus, ce qui donne :

$$p\ \theta\ \eta(\tau)\ e^{q\theta\tau - \theta\psi(\tau)}$$

Mais on a aussi :

$$\psi(\tau) = q\ \tau + p \int_0^\tau \eta(\lambda)\ d\lambda$$

Donc la loi de τ_1 admet la densité :

$$p\ \theta\ \eta(\tau_1)\ e^{-p\theta \int_0^{\tau_1} \eta(\lambda)\ d\lambda}$$

Autrement dit, τ_1 a même loi que le ν_1 associé au processus construit en remplaçant θ par $p\theta$ et γ par η (G par H).

Or à $\tau_1 = \nu_r$ fixé, la suite du processus en aval a même loi que le processus $(Y_2, Y_3, \dots)$ conditionné par $\nu_1 = \tau_1$. On en déduit sans peine :

a/ A $\tau_1, \tau_2, \dots \tau_n$ fixés, les temps d'attente $s_1 - 1, \dots s_n - s_{n-1} - 1$ des n premiers succès consécutifs sont des <u>variables de Poisson indépendantes</u> admettant les paramètres :

$$E(s_k - s_{k-1} - 1 | \tau_1, \dots \tau_k) = q\ \theta(\tau_k - \tau_{k-1})$$

b/ $\tau_1, \dots \tau_n$ admettent la même loi :

$$(p\theta)^n \, \eta(\tau_1) \, \dots \, \eta(\tau_n) \, e^{-p\theta \int_0^{\tau_n} \eta(\lambda) \, d\lambda}$$

que les $\nu_1, \dots \nu_n$ associés au processus $(p\theta, H)$.

Il en résulte bien que le processus $Z_n = Y_{s_n}$ constitué des seuls succès est identique en loi au processus construit à partir de $p\theta$ et H. Ces remarques permettent une simulation facile : A $\tau_1, \dots \tau_n$ fixés, les longueurs des séquences d'échec séparant les différents succès $s_1, \dots s_n$ sont des variables poissoniennes indépendantes entre elles et des Z_n. Du point de vue de l'inférence statistique, par contre, la connaissance de ces longueurs $s_n - s_{n-1}$ apporte une information supplémentaire, non contenue dans le processus Z_n, car, aux Z_n fixés, les s_n et les τ_n ne sont pas indépendants.

II - Meilleur estimateur linéaire.

Considérons le processus défini par :

$$Y_n = Z_n + X_1 + X_2 + \dots + X_n \qquad (n = 1, 2, \dots.)$$

où les X_n et les Z_n sont deux suites de V.A. indépendantes admettant des espérances nulles et les variances

$$D^2(X_n) = a^2 \quad , \quad D^2(Z_n) = b^2$$

Connaissant $Y_1, \dots Y_n$, on se propose de former le meilleur estimateur linéaire

$$Y^*_{n+p} = \sum_{k=1}^{n} \lambda_k \, Y_k$$

de Y_{n+p} $(p > 0)$. La matrice des covariances est ici :

(a) $$\sigma_{kk'} = a^2 \operatorname{Inf}(k,k') + b^2 \delta_{k,k'}$$

et l'on doit donc résoudre le système :

(a') $$\sum_{k'=1}^{n} \lambda_{k'} \sigma_{kk'} = \sigma_{k,n+p} = a^2 k \qquad (k = 1,2,\dots,n)$$

La solution ne dépend donc pas de p : le meilleur estimateur linéaire est le même pour tous les Y_{n+p}, et coïncide d'ailleurs avec le meilleur estimateur linéaire de la somme :

$$X_1 + \dots + X_n$$

Pour résoudre plus facilement le système (a'), nous compléterons le système d'indice par la valeur $k = 0$, en posant :

$$Y_o = Z_o$$

(Z_o indépendant de Z_1, $Z_2 \dots$ et des X_n). La covariance est encore donnée par (a). Le système (a') est remplacé par :

(b) $$\sum_{k'=0}^{n} \lambda_{k'} \sigma_{kk'} = a^2 k \qquad (k = 0, 1, 2 \dots n)$$

Si la résolution de (b) conduit à $\lambda_o = 0$, (et l'on voit assez clairement qu'il doit en être ainsi), $(\lambda_1, \dots \lambda_n)$ constituera la solution de (a').

1 - Calcul de l'expression :

$$\sum_{i=0}^{n} s^i \operatorname{Inf}(i,j) = \sum_{i=0}^{j-1} i\, s^i + j \sum_{i=j}^{n} \alpha^i$$

On part de

$$\sum_{k=0}^{n} s^k = \frac{1 - s^{n+1}}{1 - s} \quad , \quad \sum_{k=0}^{n} k\, s^k = \frac{s}{(1-s)^2} \left[1 - (n+1)s^n + n\, s^{n+1}\right]$$

D'où :

$$\sum_{i=0}^{j-1} i\, s^i = \frac{s}{(1-s)^2} \left[1 - j\, s^{j-1} + (j-1)\, s^j\right]$$

$$j \sum_{i=j}^{n} s^i = j\, \frac{s^j - s^{n+1}}{1-s}$$

et :

$$\text{(c)} \qquad \sum_{i=0}^{n} s^i \operatorname{Inf}(i,j) = s\, \frac{1 - s^j}{(1-s)^2} - j\, \frac{s^{n+1}}{1 - s}$$

2 - Calcul de l'expression :

$$\sum_{i=0}^{n} s^i \sigma_{ij} = a^2 \sum s^i \operatorname{Inf}(i,j) + b^2\, s^j$$

Compte tenu de (c), on trouve :

$$\text{(d)} \qquad \sum s^i \sigma_{ij} = a^2 \frac{s}{(1-s)^2} - a^2 \frac{s^{n+1}}{1-s}\, j + s^j \left[b^2 - a^2 \frac{s}{(1-s)^2}\right]$$

Désignons par α et β les deux racines de l'équation

(e) $$\frac{x}{(1-x)^2} = \frac{b^2}{a^2}$$

qui vérifient

$$\alpha\beta = 1 \quad , \quad \alpha + \beta = 2 + \frac{a^2}{b^2}$$

Pour s = α ou β, (d) donne :

(f) $$\begin{cases} \sum \alpha^i \sigma_{ij} = b^2 - a^2 \dfrac{\alpha^{n+1}}{1-\alpha} j \\ \sum \beta^i \sigma_{ij} = b^2 - a^2 \dfrac{\beta^{n+1}}{1-\beta} j \end{cases}$$

et par différence :

(f') $$\begin{aligned} \sum(\alpha^i - \beta^i) \sigma_{ij} &= a^2 \left[\frac{\beta^{n+1}}{1-\beta} - \frac{\alpha^{n+1}}{1-\alpha}\right] j = \\ &= b^2 [\beta^n(1-\beta) - \alpha^n(1-\alpha)] j \end{aligned}$$

3 - Calcul des poids λ_k

De (f') résulte aussitôt que la solution du système (b) est :

(g) $$\lambda_i = \frac{a^2(\alpha^i - \beta^i)}{b^2[\beta^n(1-\beta) - \alpha^n(1-\alpha)]} \qquad (i = 0, 1, \dots n)$$

Comme $\lambda_o = 0$, les $\lambda_1, \lambda_2, \dots \lambda_n$ constituent la solution du système initial (a'), et résolvent le problème posé.

4 - Variance du krigeage

Si Y (= Y_{n+p}, ou $X_1 + \ldots + X_n$) est la variable à estimer, et σ_Y^2 sa variance, la variance du krigeage est alors :

$$\sigma_K^2 = \sigma_Y^2 - \Sigma \quad \lambda_j \, \sigma_{j,Z}$$

Mais ici $\sigma_{j,Z} = a^2 \, j$, et :

$$\sigma_K^2 = \sigma_Y^2 - a^2 \sum_{j=0}^{n} j \, \lambda_j$$

On trouve sans difficulté :

$$\sigma_K^2 = \sigma_Y^2 - a^2 \left[n + \frac{\alpha^n - \beta^n}{\alpha^n(1-\alpha) - \beta^n(1-\beta)} \right]$$