

REVUE DE STATISTIQUE APPLIQUÉE

G. MAAREK

A. SEGOND

L'estimation séquentielle dans les modèles linéaires

Revue de statistique appliquée, tome 17, n° 2 (1969), p. 21-45

http://www.numdam.org/item?id=RSA_1969__17_2_21_0

© Société française de statistique, 1969, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

L'ESTIMATION SÉQUENTIELLE DANS LES MODÈLES LINÉAIRES

G. MAAREK et A. SEGOND

REMERCIEMENTS

Sans la collaboration de Mademoiselle Ulmo, Professeur à l'Institut de Statistique de l'Université de Paris, la parution de cet article n'aurait sans doute pas été possible. En effet, grâce à la rigueur de son esprit et la profondeur de ses connaissances, elle a su formuler des critiques constructives, qui nous ont permis de cerner de plus près de nombreuses difficultés mathématiques qui étaient restées dans l'ombre. Et nous avons rédigé ce travail encouragés par l'intérêt qu'elle y a trouvé.

Quelle trouve ici tous nos remerciements.

G. MAAREK et A. SEGOND

Les problèmes d'estimation des coefficients d'influence des facteurs conditionnant un phénomène se présentent fréquemment sous forme séquentielle. En effet, le cas de l'expérimentation mis à part, l'information susceptible d'être traitée par les méthodes statistiques appropriées n'est le plus souvent pas entièrement disponible à un instant donné mais parvient progressivement à l'usager qui est censé l'exploiter pour décider, pour agir. L'intérêt des méthodes statistiques permettant de faire des estimations de plus en plus précises au fur et à mesure qu'arrive l'information, apparaît tout naturellement car l'action le plus souvent n'attend pas.

Les problèmes d'estimation séquentielle abondent dans de très nombreux domaines de l'activité économique. On peut citer :

- La recherche minière : Exploitation des informations (teneur en minerai) provenant de sondages successifs pour décider d'un abandon, d'une intensification de la prospection ou de la mise en exploitation d'un gisement présumé.

- L'industrie pétrolière : Détermination des règles d'exploration systématique d'une région (localisation des forages, importance des campagnes géophysiques) et éventuellement (voir exemple plus loin) des règles d'extension d'un gisement qui vient d'être découvert.

- L'industrie chimique : Estimation dynamique de l'influence de divers paramètres physiques ou physico-chimiques (température, pression, débit) sur le rendement d'une réaction chimique afin d'assurer le contrôle automatique des conditions de fonctionnement du réacteur correspondant.

- L'Industrie de produits de grande consommation : Ajustement à court terme des programmes de production et à moyen terme des capacités de production (personnel, équipement) aux besoins du marché qu'il faut périodiquement réestimer en tenant compte d'influences diverses (publicité, concurrence, revenu des ménages, etc.).

- Etc...

Pourtant, la plupart des méthodes statistiques mises à la disposition des usagers (Analyse de variance et covariance en particulier) se prêtent mal à des remises à jour fréquentes des estimations. En effet, ces méthodes nécessitent pour chaque réestimation la prise en compte de toute l'information élémentaire passée de sorte qu'il ne s'agit pas à proprement parler d'une amélioration de l'estimation précédente mais véritablement d'une réestimation. On imagine aisément le manque de souplesse de telles procédures surtout si l'information est abondante et si le nombre de facteurs dont on cherche à estimer l'influence est élevé (cas de la météo par exemple).

Par ailleurs, un second inconvénient des méthodes classiques d'estimation statistique des coefficients d'influence repose sur leur caractère essentiellement objectif. Or, dans de nombreuses situations, en l'absence de toute information chiffrée, l'usager possède tout de même une certaine connaissance subjective à priori plus ou moins floue du phénomène grâce à son "expérience" ou à la prise en compte de considérations externes impossibles à quantifier.

Le statisticien dira alors qu'il existe une distribution de probabilité a priori des coefficients d'influence cherchés dont il convient de tenir compte, par des méthodes de type bayésien lors de l'estimation effectuée de ces coefficients à l'aide des informations chiffrées. Ceci n'est malheureusement pas possible avec les méthodes classiques de la régression.

Enfin, l'analyse de la régression classique ne permet pas de répondre convenablement au problème du choix des conditions de mesure (niveau des facteurs) qui procurent le maximum d'information relative à une grandeur liée au phénomène étudié. Il est donc apparu nécessaire de mettre au point une méthode d'estimation séquentielle pour les modèles linéaires, susceptible de répondre au triple objectif :

- Souplesse des estimations successives,
- Prise en compte de connaissances subjectives a priori,
- Recherche des conditions d'expériences optimales.

1 - LE MODELE D'HYPOTHESE LINEAIRE

Conformément à la méthodologie de la Recherche Opérationnelle, l'étude commence par le choix d'un modèle mathématique destiné à représenter le réseau des liaisons. Ce modèle sera évidemment de nature probabiliste eu

égard au caractère aléatoire de la réponse : chaque valeur prise par la variable réponse est considérée comme la réalisation d'une variable aléatoire notée y dont on définit l'espérance mathématique $E(y)$ et la variance $V(y) = \sigma^2$.

Le modèle s'appuie sur le corps d'hypothèses suivant :

Hypothèse 1 : La réponse y suit, tous facteurs fixés, une loi de probabilité de Laplace-Gauss.
Sa variance σ^2 est constante sur le domaine exploré.

Hypothèse 2 : Les réponses successives sont indépendantes en probabilité.

Hypothèse 3 : L'influence des facteurs s'exprime sous une forme linéaire :

$$\mathcal{E}(y_u) = \mu + C_1 x_{1u} + C_2 x_{2u} \dots + C_q x_{qu} \quad (1)$$

- \mathcal{E} désigne l'espérance mathématique.
- μ est l'ordonnée à l'origine.
- $x_{1u}, x_{2u}, \dots, x_{qu}$ sont les valeurs prises par les variables, x_1, x_2, \dots, x_q lors de la $u^{\text{ième}}$ mesure.
- C_1, C_2, \dots, C_q sont les coefficients de régression entre y et x_1, x_2, \dots, x_q
- y_u est la valeur de la réponse associée à la $u^{\text{ième}}$ mesure ($u = 1$ à n) dans les conditions $x_{1u}, x_{2u}, \dots, x_{qu}$
- $\varepsilon_u = y_u - (\mathcal{E} y_u)$ est la réalisation d'une variable aléatoire suivant une loi de Laplace-Gauss centrée et de variance σ^2 .

1.1. - Représentation matricielle

Il est commode d'introduire des notations matricielles :

- Le vecteur colonne "réponse" est noté Y si bien que son transposé (vecteur ligne) s'écrit :

$$Y' = (y_1, \dots, y_u, \dots, y_n)$$

pour n observations ou unités statistiques.

- Le vecteur colonne "paramètres théoriques" noté C , est le vecteur dont les composantes sont les valeurs théoriques des paramètres linéairement indépendants du modèle ; il est donc de dimension $d = q + 1$. Son transposé est par suite :

$$C' = (\mu, C_1 \dots C_q)$$

- Le tableau des conditions d'observation est une matrice M à d colonnes et n lignes, formée par la juxtaposition de n vecteurs ligne notés m' dont les composantes sont définies par l'équation suivante :

$$Y = M \cdot C + E \quad (2)$$

où E est le vecteur "écart" de composantes ε_u . Par suite son transposé $E' = (\varepsilon_1, \dots, \varepsilon_u, \dots, \varepsilon_n)$ et, la matrice de covariance de E est $\sigma^2 I$ (I de dimension n) tandis que son espérance est nulle. L'égalité (2) est équivalente à n égalités :

$$y_u = m'_u \cdot C + \varepsilon_u \quad (u = 1 \text{ à } n) \quad (3)$$

avec

$$m'_u = 1, x_{1u}, \dots, x_{qu}$$

On peut noter que le premier élément de m'_u vaut toujours 1, puisque c'est le coefficient de μ . La première colonne de la matrice M des conditions d'observation sera constituée par des 1.

1.2.-Remarques

- Remarque 1

La transcription progressive de l'information consiste à la n + 1^{ième} mesure à accroître :

- La matrice M d'une ligne supplémentaire m'_{n+1} correspondant aux valeurs prises par les facteurs au cours de cette mesure

- le vecteur Y d'une composante y_{n+1} correspondant au n + 1^{ième} résultat.

- Remarque 2

Nous avons supposé que les facteurs explicatifs étaient des variables mesurables, ceci dans le seul but de simplifier la présentation de la méthode. Si certains facteurs sont qualitatifs, on crée autant de variables binaires artificielles qu'il y a de modalités dans les facteurs. En affectant à ces variables les valeurs 1 ou 0 suivant que la modalité correspondante se présente ou ne se présente pas, on se ramène aisément au formalisme des variables mesurables.

2 - ESTIMATION DES PARAMETRES THEORIQUES ET TEST D'INFLUENCES RAPPEL DE LA METHODE DES MOINDRES CARRES.

2.1. - Estimation des paramètres théoriques

La méthode des moindres carrés consiste à trouver le vecteur \hat{C} qui minimise la somme Q des carrés des écarts des résultats effectifs y_i à leurs estimation \hat{y}_i , ou encore en langage vectoriel, qui minimise la norme euclidienne du vecteur $Y - \hat{Y}$ où $\hat{Y} = \hat{M}C$

$$\text{Min}_C Q(C) = \text{Min}_C \sum_i (\hat{y}_i - y_i)^2 = \text{Min}_C ||\hat{Y} - Y||^2 = \text{Min}_C ||MC - Y||^2$$

Le scalaire Q(C) peut encore s'écrire :

$$Q(C) = (MC - Y)' (MC - Y)$$

(MC - Y' étant le vecteur transposé de (MC - Y)

En annulant la dérivée de $Q(C)$ par rapport à C , on obtient :

$$2 M' (M\hat{C} - Y) = 0$$

soit

$$M'M\hat{C} = M'Y$$

en multipliant à gauche par $(M'M)^{-1}$, on en déduit l'estimation \hat{C} du vecteur des paramètres :

$$\boxed{\hat{C} = (M'M)^{-1} M'Y} \quad (4)$$

La présence de la matrice $(M'M)^{-1}$ dans cette formule montre bien qu'il faut traiter l'information simultanément et que ce traitement se traduit surtout par l'inversion d'une matrice $M'M$ (d lignes, d colonnes) d'autant plus importante que le nombre de variables explicatives retenues ($d-1$) est élevé.

Montrons que

$$\hat{Y} = \hat{\mathcal{E}}(Y) = M\hat{C}$$

est une estimation non biaisée de $\mathcal{E}(Y) = MC$

$$(4) \text{ donne } \hat{Y} = M(M'M)^{-1} M'Y$$

$$(2) \text{ s'écrit } MC = Y - E$$

Multiplions à gauche par $(M'M)^{-1} M'$ les deux membres de (2), d'où :

$$(M'M)^{-1} M'M C = (M'M)^{-1} M'Y - (M'M)^{-1} M'E$$

soit

$$C = \hat{C} - (M'M)^{-1} M'E \text{ ou } \hat{C} = C + (M'M)^{-1} M'E$$

On en déduit

$$\mathcal{E}(\hat{C}) = C + \mathcal{E}[(M'M)^{-1} M'E] = C$$

puisque

$$\mathcal{E}[(M'M)^{-1} M'E] = (M'M)^{-1} M' \mathcal{E}(E) = 0$$

On en conclut que le vecteur \hat{C} est une estimation non biaisée de C , ce qui entraîne la même propriété pour \hat{Y} relativement à $\mathcal{E}(Y)$. De plus, C obéit à une loi de Gauss à d dimensions, dont la matrice R des variances et covariances est :

$$\boxed{R = \sigma^2 (M'M)^{-1}} \quad (5)$$

En effet :

$$\begin{aligned} R &= \text{Esp } (\hat{C} - C) (\hat{C} - C)' = \text{Esp } [(M'M) \quad M'E \quad E'M \quad (M'M)^{-1}] \\ &= \sigma^2 I. (M'M)^{-1} M'M (M'M)^{-1} \end{aligned}$$

d'où la formule (5).

On notera que cette matrice R , qui caractérise la précision de l'estimation obtenue, ne dépend pas du vecteur Y , c'est-à-dire des résultats des mesures, mais seulement de M , c'est-à-dire de l'ensemble des conditions d'observation des mesures, σ^2 étant fixé.

L'expression analytique de R ne laisse rien présager de sa convergence, vers 0, lorsque le nombre d'expériences croît.

Intuitivement, on s'attendrait pourtant à ce que l'incertitude sur l'estimation tende vers zéro, lorsque le nombre d'expériences croît sur l'infini. En fait, ce n'est pas toujours le cas : il faut pour cela que la matrice M des conditions d'expérience satisfasse certaines conditions assez générales, étudiées en particulier par Monsieur Malinvaud.

Définissons maintenant successivement :

- La variabilité résiduelle après ajustement, qui en mesure la précision :

$$V_R = \sum_{u=1}^n (y_u - \hat{y}_u)^2$$

V est égal au carré de la norme $\|M\hat{C} - Y\|$ du vecteur $\hat{Y} - Y$, soit :

$$V_R = (M\hat{C} - Y)' (M\hat{C} - Y)$$

Compte tenu de la formule (4), on en déduit, tous calculs effectués :

$$\boxed{V_R = Y'Y - Y'M\hat{C}} \quad (6)$$

Le quotient $\frac{V_R}{n - d}$ estime σ^2 sans biais à la condition toutefois que le modèle soit une bonne représentation de la réalité.

- La variabilité totale des réponses :

$$V_T = \sum_{u=1}^n (y_u - \bar{y})^2$$

\bar{y} étant la moyenne des réponses successives.

- Le taux de variabilité expliquée par l'ajustement :

$$1 - \frac{V_R}{V_T}$$

L'équation (6) conduit à l'égalité :

$$\sum (y_u - \bar{y})^2 = \sum (y_u - \hat{y}_u)^2 + \sum (\hat{y}_u - \bar{y})^2 \quad (7)$$

soit

$$V_T = V_R + V_F$$

Elle exprime que la variabilité totale est égale à la variabilité résiduelle, indicateur de la précision de l'ajustement, augmentée de la quantité $V_F = \sum (\hat{y}_u - \bar{y})^2$ qui représente la variabilité due aux facteurs introduits.

L'égalité (6) entraîne également

$$\sum_u y_u^2 = \sum_u (y_u - \hat{y}_u)^2 + \sum_u \hat{y}_u^2 \quad (8)$$

Cette relation est, comme nous allons le voir, particulièrement appropriée à l'étude des tests de signification des influences des divers facteurs.

2.2. - Test des influences

Nous conviendrons de ce qui suit :

- Un facteur mesurable est sans influence réelle si le coefficient de régression associé est égal à 0.

- Un facteur qualitatif n'a pas d'influence réelle si tous les coefficients de régression correspondant à ses différentes modalités sont égaux.

- L'influence d'un facteur quelconque est significative si un test statistique conduit à rejeter l'hypothèse d'absence d'influence réelle.

Le principe du test de l'influence d'un facteur est le suivant :

Dans un premier stade, on effectue un ajustement de tous les paramètres ce qui fournit un vecteur \hat{C} et une variabilité résiduelle V_R à $(n-d)$ degrés de liberté, d étant le nombre de paramètres indépendants égal au rang de la matrice $M'M$.

(8) donne :

$$\sum_u y_u^2 = V_R + \sum_u \hat{y}_u^2 \quad (9)$$

Ensuite, on annule a priori les paramètres correspondant aux facteurs testés, supposons les facteurs x_1, x_2, x_j , et on effectue un ajustement analogue. On obtient ainsi un vecteur \hat{C}_h et une variabilité résiduelle V_{Rh} à $(n-d_h)$ degrés de liberté : d_h étant le nombre des paramètres indépendants, obtenu en soustrayant à d le nombre J des paramètres annulés.

On obtient cette fois :

$$\sum_u y_u^2 = V_{Rh} + \sum_u y_{uh}^2$$

Il est tout à fait logique de vouloir mesurer l'influence des facteurs testés par la variation positive de variabilité résiduelle $V_{Rh} - V_R$; en particulier, une variation nulle indiquera que les facteurs $x_1 \dots x_j$ sont sans influence sur la réponse.

Cette différence est en fait mesurée en valeur relative par le quotient :

$$F_h = \frac{\frac{V_{Rh} - V_R}{(n-d_h) - (n-d)}}{\frac{V_R}{n-d}} \quad (10)$$

On montre en effet que (compte tenu des hypothèses faites) si l'hypothèse de non influence réelle de $x_1 \dots, x_j$ est vraie, la quantité F_h est une variable aléatoire de Fisher-Snedecor à $n_1 = (n-d_h) - (n-d) = d - d_h$, et $n_2 = n-d$ degrés de liberté.

Le test statistique en découle : si F_h est supérieur à la valeur F correspondant au seuil de signification choisi, on est en droit de conclure à l'influence des facteurs testés.

2.3. - Remarques

Remarque 1

La procédure utilisée consistant à annuler a priori les paramètres relatifs aux facteurs à tester, conduit à des estimations des autres paramètres, identiques à celles directement issues d'une analyse où ces facteurs n'auraient pas été retenus : il revient au même de prendre comme matrice des conditions d'observation la matrice M_h se déduisant de M par suppression des colonnes correspondant aux coefficients annulés. (d_h est aussi égal au rang de la matrice $M_h' M_h$).

Remarque 2

Il est important, sur le plan théorique, de noter que les influences décelées sont des influences conditionnelles : on isole l'influence d'un groupe de facteurs quand celui-ci est introduit dans le modèle après que tous les autres l'aient été.

3 - METHODES D'ESTIMATION SEQUENTIELLE

3.1. - Formules itératives

On a vu que la matrice M se constituait par juxtaposition de vecteurs de conditions d'expérience.

On notera :

- m_n le vecteur caractérisant les conditions de la $n^{\text{ième}}$ expérience.
- La matrice telle qu'elle est constituée après accumulation de n expériences. Elle a d lignes et n colonnes.

- Y_n le vecteur colonne $\begin{pmatrix} y \\ \vdots \\ y_n \end{pmatrix}$ des résultats de n expériences

- \hat{C}_n l'estimation du vecteur C au terme des n premières expériences.

Nous avons vu que l'estimation après n mesures \hat{C}_n était définie par l'équation :

$$\hat{C}_n = (M_n' M_n)^{-1} M_n' Y_n$$

Comme on l'a remarqué, la structure des matrices M_n permet d'écrire :

$$M_n' = |M_{n-1}' \mid m_n \mid$$

$$P_n = |M_n' M_n|^{-1}$$

on a :

$$P_n^{-1} = |M_n' M_n| = |M_{n-1}' M_{n-1}| + m_n m_n'$$

$$P_n^{-1} = P_{n-1}^{-1} + m_n m_n'$$

On peut en déduire la formule itérative reliant P_n et P_{n-1} (cf. annexe).

$$P_n = P_{n-1} - (m_n' P_{n-1} m_n + 1)^{-1} P_{n-1} m_n m_n' P_{n-1}$$

En transposant cette relation dans l'expression de C_n

$$\hat{C}_n = P_n (M_{n-1}' Y_{n-1} + m_n y_n)$$

on obtient

$$C_n = C_{n-1} + (m_n' P_{n-1} m_n + 1)^{-1} P_{n-1} m_n (y_n - m_n' C_{n-1})$$

En résumé :

Si à un instant donné on connaît l'estimation \hat{C}_{n-1} du vecteur paramètre C et la matrice $P_{n-1} = [M_{n-1}' M_{n-1}]^{-1}$ l'adjonction d'une nouvelle mesure caractérisée par un jeu de facteurs représenté par le vecteur ligne m_n' d'une part et un résultat y_n d'autre part, entraîne une réestimation du vecteur paramètre \hat{C} et de la matrice P conformément aux formules ci-après :

$$\hat{C}_n = \hat{C}_{n-1} + (m_n' P_{n-1} m_n + 1)^{-1} P_{n-1} m_n (y_n - m_n' \hat{C}_{n-1})$$

$$P_n = P_{n-1} - m_n' P_{n-1} m_n + 1)^{-1} P_{n-1} m_n m_n' P_{n-1}$$

Les valeurs à introduire pour la première itération sont alors $\hat{C}_0 = (M_d' M_d)^{-1} M_d' Y_d$ obtenues par la méthode classique à partir de premières mesures.

La facilité d'emploi de ces formules vient de ce que l'expression

$$(m_n' P_{n-1} m_n + 1)$$

est un scalaire, et qu'il n'y a aucune inversion de matrice à effectuer pour en calculer l'inverse.

La variation résiduelle peut également se calculer de façon itérative.

En effet :

$$V_{R_n} = Y_n' Y_n - Y_n' M_n \hat{C}_n \quad (12)$$

La variation de V_{R_n} après la $n^{\text{ième}}$ mesure donnée par la formule :

$$\Delta V_{R_n} = y_n^2 - y_n m_n' \hat{C}_{n-1} - H_n' (\hat{C}_n - \hat{C}_{n-1}) \quad (12 \text{ bis})$$

où H_n' est un vecteur ligne à d composantes donné par la formule :

$$H_n' = Y_n' M_n$$

dont les composantes sont :

$$h_j(n) = m_j' Y = \sum_{u=1}^{u=n} m_{ju} y_u \quad j = 1, 2, \dots, d$$

Les composantes $h_j(n)$ peuvent être facilement calculées à chaque itération par addition du scalaire $m_{jn} y_n$.

On remarquera que m_{jn} désigne contrairement à l'habitude l'élément de la j^{e} colonne et de la u^{e} ligne de M .

La formule (12 bis) se démontre facilement à l'aide de (12) en écrivant que ΔV_R est la différence de deux valeurs consécutives de V_R .

3.2. - Conséquences

Les conséquences des deux résultats ci-dessus sont assez importantes :

- Il est possible d'atteindre l'estimation des moindres carrés du vecteur paramètre C de façon itérative par prise en compte progressive de l'information au fur et à mesure qu'elle parvient. L'avantage opérationnel d'une telle procédure est évident.

En effet, dans les formules (11) (12) et (12 bis), les opérations compliquées se limitent à :

- des multiplications matrice par vecteur

$$(P_{n-1} m_n \text{ ou } m_n P_{n-1})$$

- des multiplications vecteur par vecteur

$$m_n' \hat{C}_{n-1} \text{ ou } H_n' (\hat{C}_n - \hat{C}_{n-1})$$

La seule inversion porte sur un scalaire

$$m_n' P_{n-1} m_n + 1$$

ce qui évidemment ne pose aucun problème du point de vue mémoires.

- Il est possible (voir § ci-après) de faire les tests de signification.

3.3. - Tests de signification

La méthode de test exposée précédemment et matérialisée par le test de fisher-Snedecor sur la variable est toujours valable

$$F_h = \frac{\frac{V_{Rh} - V_R}{(n-d_h) - (n-d)}}{\frac{V_R}{n-d}}$$

à condition que l'on soit en mesure de calculer les variances résiduelles V_R et V_{Rh} sans avoir à remanipuler toute l'information antérieurement traitée.

En ce qui concerne V_R il n'y a aucun problème grâce à la formule (12 bis) qui donne sa variation à chaque mesure.

Par contre, le calcul de V_{Rh} est moins simple : il s'agit de la variation résiduelle qu'on aurait obtenue si, tout au long des itérations, on n'avait pas tenu compte des d_h facteurs dont on veut tester la signification. Afin d'obtenir ce résultat tout en conservant l'approche itérative on peut introduire, dans le cas d'un test d'un seul facteur à la fois, ($d = 1$) une mesure fictive à résultat nul et où tous les facteurs sont nuls à l'exception du facteur à tester que l'on prendra égal à un nombre N très grand :

$$y = 0$$

$$m = (0, 0, 0, N, \dots, 0, 0, 0, 0)$$

Cela conduit à une réestimation $\hat{C}^{(h)}$ du vecteur paramètre \hat{C} , qui, quand N tend vers l'infini, tend vers celle que l'on aurait obtenue si on avait négligé le facteur en question depuis le début. Ceci est démontré en annexe.

Grâce à cette nouvelle mesure factice il est possible de calculer d'après la formule (12 bis) :

$$\Delta V_{Rh} = y_{n+1}^2 - y_{n+1} m'_{n+1} \hat{C}_{n+1} - H_{n+1} (\hat{C}^{(h)} - \hat{C}_n)$$

Comme $Y_{n+1} = 0$ et $H_{n+1} = H_n$ puisque $m_{j_{n+1}} y_{n+1} = 0$

Il vient :

$$\Delta V_{R_h} = H_n' (\hat{C}_n - C^{(h)})$$

ΔV_{R_h} est une approximation de l'expression :

$$V_{R_h} - V_R$$

qui apparaît au numérateur de l'expression F_h .

Cette formule est encore vraie dans le cas de l'adjonction successive de plusieurs mesures à résultat nul ; $\hat{C}^{(h)}$ étant la réestimation finale puisque c'est H_n et non $H_n + 1$ qui figure au numérateur de l'expression F_h .

Dans ces conditions, le test devient possible et il n'a pas été nécessaire pour cela de retraiter toute l'information.

4 - APPROCHE BAYESIENNE

La méthode d'estimation des moindres carrés - qu'elle soit itérative ou non - est basée sur l'idée qu'il existe une valeur vraie du vecteur, soit C , et que la mesure observée est sujette à aléa, et, donc, est elle-même une variable aléatoire - Il s'ensuit que l'estimation \hat{C}_n calculée d'après n observations de cette variable aléatoire est également aléatoire.

L'approche bayésienne donne - on le verra plus loin - des résultats pratiques semblables à ceux de l'estimation par les moindres carrés - Néanmoins les idées qui sont à sa base sont bien différentes.

Pour saisir ces idées, on peut se mettre à la place de l'expérimentateur au moment où il va exploiter les résultats d'un ensemble de mesures ; en vue de déterminer un certain nombre de caractéristiques, c'est-à-dire un vecteur C .

Il a déjà une certaine idée a priori des valeurs que peut prendre le vecteur C . Autrement dit, on peut considérer C comme une variable aléatoire, caractérisée par une certaine distribution de probabilité.

En exploitant une expérience, l'expérimentateur peut voir dans quelle mesure le résultat obtenu confirme ou infirme l'idée qu'il avait de la distribution de probabilité de C et lui permet de "corriger" cette distribution. En d'autres termes, on passe d'une distribution de probabilité a priori de C avant expérience à une distribution de probabilité a posteriori de C après expérience.

Les relations qu'on va écrire ont pour but de déterminer celle-ci à partir de celle-là et du résultat de l'expérience.

Ce problème mathématique est en général difficile, et nous ne parlons que d'un cas particulier où :

- la distribution a priori du vecteur C est normale .

- On connaît la précision de la mesure effectuée, c'est-à-dire en employant les notations du paragraphe précédent, on connaît la distribution de probabilité des ε - que l'on suppose, de plus, normale $(0, \sigma^2)$.

Nous allons montrer que, dans ces conditions, la distribution a posteriori de C est également normale. On pourra donc exploiter entièrement une série d'expériences de cette manière. Pour cela :

- On part d'une distribution "a priori" D_0 , à partir de laquelle on calcule, par exploitation d'un premier résultat d'expérience, une distribution "a posteriori" D_1 .

- Celle-ci est prise, à son tour, comme distribution "a priori" lors de l'exploitation de la seconde expérience et ainsi de suite.

Ce procédé prend en compte toute l'information apportée par chaque résultat, de même que l'information initiale qu'on peut avoir avant le démarrage des expériences.

Appelons :

$p(C, y)$ la densité de probabilité de C et de la mesure y

$p(C)$ la densité de probabilité a priori de C avant la mesure (celle d'une loi normale de moyenne \bar{C} et de matrice de covariance (R)).

ε l'écart entre la mesure y et son estimation a priori $m'C$.

$p(\varepsilon)$ la densité de probabilité de l'erreur de mesure. On suppose que c'est celle d'une loi normale de moyenne nulle et de variance σ^2 , connue.

On a, du fait de l'indépendance stochastique entre C et E,

$$P(C) \cdot p(\varepsilon) = p(C) p(y - m'C) = J p(C, y)$$

où J est le jacobien :

$$J = \frac{D(C, y)}{D(C, \varepsilon)} = \frac{D(y)}{D(\varepsilon)}$$

Or, d'après la relation linéaire $y = m'C + \varepsilon$ liant y à ε ce jacobien est égal à l'unité.

On a donc :

$$p(C, y) = p(C) p(y - m'C)$$

D'autre part, la distribution de probabilité de y est normale puisque y est une combinaison linéaire de variables indépendantes en probabilité et normalement distribuées.

$$y = m'C + \varepsilon$$

On a :

$$E(y) = m'E(C) + E(\varepsilon) = m'\bar{C}$$

et

$$\begin{aligned} \text{Var}(y) &= \mathbf{E} \{ (y - \mathbf{E}(y)) \cdot (y - \mathbf{E}(y))' \} \\ \text{Var}(y) &= \mathbf{E} (y - m'\bar{C}) \cdot (y - m'\bar{C})' \\ \text{Var}(y) &= \mathbf{E} \{ (m'(C - \bar{C}) + \varepsilon) (m'(C - \bar{C}) + \varepsilon)' \} \\ \text{Var}(y) &= m'Rm + \sigma^2 \text{ puisque } \mathbf{E}(C - \bar{C}) (C - \bar{C})' = R \\ &\quad \text{et } \mathbf{E}(C - \bar{C}) \varepsilon = 0 \end{aligned}$$

Donc

$$p(y) = K_1 \exp \frac{-\frac{1}{2} (y - m'\bar{C})^2}{\sigma^2 + m'Rm}$$

La distribution a posteriori de C après la mesure y est donnée par la loi de Bayes :

$$P(C|y) = \frac{p(C, y)}{p(y)}$$

Explicitons les lois de probabilités :

$$\begin{aligned} p(C) &= K_2 \exp \left[-\frac{1}{2} (C - \bar{C})' R^{-1} (C - \bar{C}) \right] \\ p(C, y) &= p(C) p(y - m'C) \\ p(C, y) &= K_2 \exp \left[-\frac{1}{2} (C - \bar{C})' R^{-1} (C - \bar{C}) \right] K_3 \exp \left[-\frac{1}{2} \frac{(y - m'C)' (y - m'C)}{\sigma^2} \right] \\ &= K_2 K_3 \exp \left\{ -\frac{1}{2} \left[(C - \bar{C})' R^{-1} (C - \bar{C}) + \frac{(y - m'C)' (y - m'C)}{\sigma^2} \right] \right\} \\ p(C|y) &= K_4 \exp \left[-\frac{1}{2} \left[(C - \bar{C})' R^{-1} (C - \bar{C}) + \frac{(y - m'C)' (y - m'\bar{C})}{\sigma^2} - \frac{(y - m'\bar{C})' (y - m'\bar{C})}{\sigma^2 + m'Rm} \right] \right] \quad (13) \end{aligned}$$

L'exponentielle portant sur une forme quadratique de C, on en déduit que la loi de probabilité a posteriori de C après une mesure y est également normale.

Pour déterminer la moyenne $\bar{C}^{(1)}$ et la matrice de covariance de $R^{(1)}$ de cette loi, il faut identifier l'expression (13) à :

$$k \exp \left(-\frac{1}{2} (C - \bar{C}^{(1)})' R^{(1)-1} (C - \bar{C}^{(1)}) \right)$$

Il vient :

$$R^{(1)-1} = R^{-1} + \frac{m'm}{\sigma^2}$$

expression analogue à celle de P_n^{-1} établie en 3-1-

$$\begin{aligned}
 R^{(1)} &= R - \frac{R m m' R}{m' R m + \sigma^2} \\
 \bar{C}^{(1)} &= \bar{C} + \frac{R m (y - m' C)}{m' R m + \sigma^2}
 \end{aligned}
 \tag{14}$$

Cette loi de probabilité a posteriori sera utilisée comme loi a priori pour calculer la loi de probabilité a posteriori de C après observation d'une nouvelle valeur de y et puisqu'elle a la même forme que la loi a priori initiale, les formules (14) sont encore applicables et peuvent être utilisées comme formules de récurrence.

C'est l'identité de forme de la loi a posteriori et de la loi a priori, qui n'est réalisée pour une loi a priori normale que lorsque σ^2 est supposé connu, qui permet l'utilisation pratique de l'estimation séquentielle par la méthode de Bayes.

On remarquera l'analogie entre les formes de récurrence établies dans l'approche classique et dans l'approche bayésienne. Si dans les équations (14) on pose :

$$R = \sigma^2 P,$$

le terme σ^2 s'élimine, et les équations obtenues deviennent formellement identiques aux équations (11) ce qui, pratiquement, permet d'utiliser les mêmes programmes numériques dans l'une et l'autre méthode.

Mais il importe de souligner les différences d'optique. Dans l'approche classique \hat{C} est aléatoire et suit une loi normale de moyenne C, tandis que dans l'approche bayésienne, c'est C qui est aléatoire et a une distribution a posteriori normale de moyenne $C^{(1)}$.

En particulier, dans l'approche bayésienne, les tests d'hypothèses sur C, tels qu'on les pratique, dans l'estimation statistique classique, n'ont pas de sens - Par contre, on peut examiner si une valeur particulière C_0 , se trouve à l'intérieur d'un domaine de probabilité pour C.

On peut dire toutefois que dans les deux cas, $(C - \bar{C}^{(1)})$ et $(\hat{C}_n - C)$, sont des variables aléatoires normales centrées, pour lesquelles les équations d'itérations sont identiques.

Il serait tentant de dire que ces deux suites de vecteurs convergent l'une vers l'autre.

Néanmoins, les propriétés de convergence de l'approche bayésienne n'ont pas été étudiées - du moins à notre connaissance. Il est vraisemblable que cela soulèverait des difficultés comparables à celles rencontrées lors d'études semblables sur la convergence de l'estimateur classique.

5 - EXEMPLE D'APPLICATION DANS L'INDUSTRIE PETROLIERE (*)

Un gisement de pétrole vient d'être découvert et le problème se pose de mettre en place une procédure "d'extension" du gisement permettant d'en

(*) Cette étude a été faite à la demande de la Compagnie Française des Pétroles.

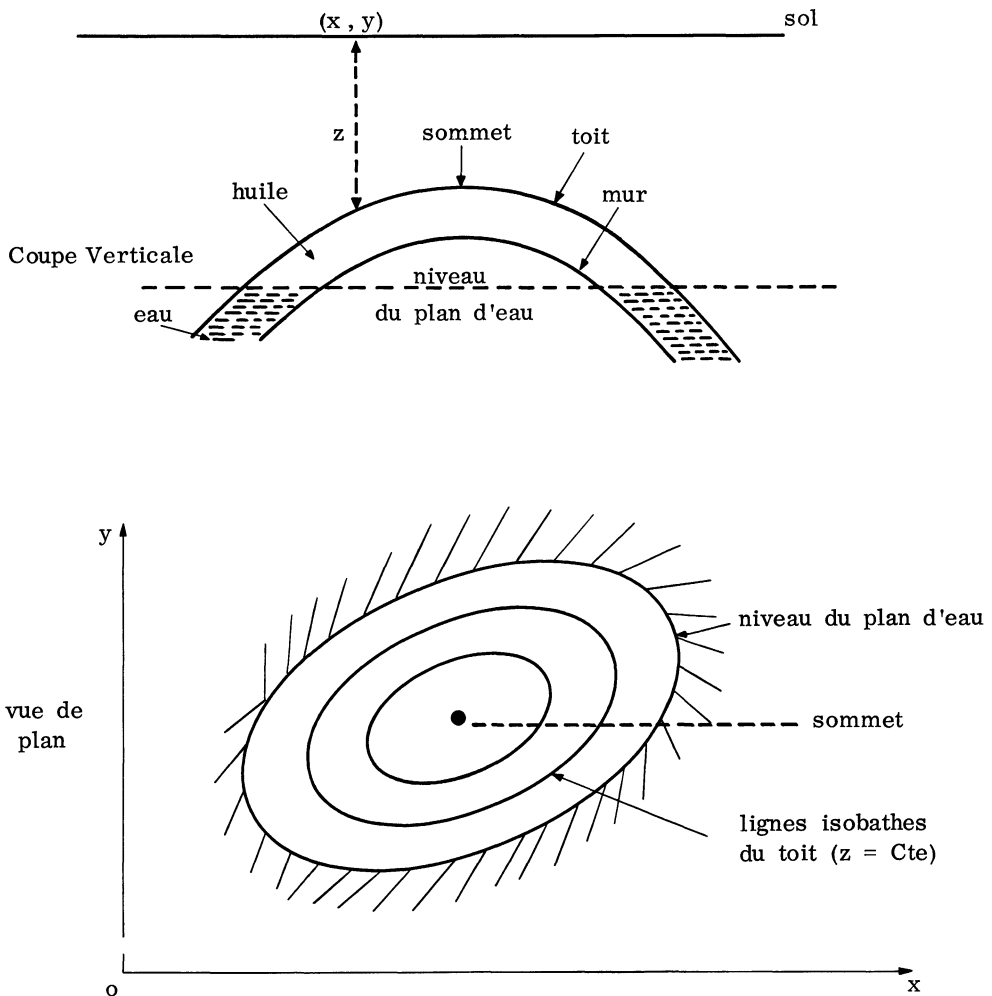
apprécier les réserves et d'ajuster en conséquence les investissements en infra-structure d'exploitation (stations de pompage, pipes-lines, etc...).

La connaissance de la structure géologique du gisement repose sur deux types d'information :

- la géophysique (sismique reflexion)
- Les forages.

Les informations géophysiques ont l'avantage d'être peu coûteuses et étendues à l'ensemble du gisement, mais elles sont imprécises. Par contre les forages, sont certes très coûteux et ne donnent qu'une information très locale mais cette information est bien plus précise.

L'importance des réserves est liée au volume du réservoir qui dépend de la forme de celui-ci, du niveau du plan d'eau qui limite la hauteur d'imprégnation en huile et de l'épaisseur de la couche imprégnée.



La position géographique des points d'informations-mesure géophysique ou forage - est repérée par un système d'axes cartésiens (Oxy).

Des considérations géologiques permettent d'assimiler en première approximation la forme du toit du réservoir à un paraboloïde à axe vertical d'équation

$$z = C_1 x^2 + C_2 xy + C_3 y^2 + C_4 x + C_5 y + C_6$$

où z est la profondeur du toit du réservoir en chaque point x y' et d'admettre que le mur du réservoir est décalé verticalement par rapport au toit de résultat nul et où z apparaît comme un produit du vecteur ligne

$$m = (x^2, xy, y^2, x, y, 1)$$

par le vecteur colonne

$$C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix}$$

Ce vecteur C matérialise la géométrie du réservoir et par conséquent le volume $V = V(C)$ de celui-ci, une fois connues sont épaisseur et le niveau du plan d'eau.

Pour pouvoir estimer ce volume, il est donc nécessaire d'estimer les composantes du vecteur C en fonction des informations disponibles.

Le niveau du plan d'eau et l'épaisseur du réservoir entreront implicitement dans l'expression analytique de V parce qu'ils ne sont pas à estimer. Le niveau du plan d'eau est en effet connu à la suite d'un forage qui a eu "les pieds dans l'eau". Quant à l'épaisseur du réservoir, elle résulte de considérations géologiques externes.

Les problèmes qui se posent sont les suivants :

1/ Disposant à un stade donné de l'extension, d'informations géophysiques et de forages, estimer les réserves du gisement et l'incertitude sur cette estimation.

2/ On désire poursuivre la reconnaissance du gisement. Où placer le forage suivant pour minimiser l'incertitude sur l'estimation des réserves ?

Problème 1 - Estimation des réserves

L'équation représentative du paraboloïde limitant le réservoir est linéaire par rapport aux coefficients C_i

On a donc un modèle linéaire où les facteurs d'influence sont représentés par la position géographique d'un point sur la carte et où le résultat est la profondeur du gisement en ce point. Les méthodes d'estimation séquentielle exposées plus haut sont donc applicables.

Les coefficients C_i sont estimés au départ par une régression portant sur l'ensemble des informations géophysiques.

Cette première estimation est ensuite séquentiellement réajustée pour tenir compte des résultats des forages successifs déjà réalisés.

Ainsi l'ensemble des informations disponibles à un stade donné de la reconnaissance est matérialisé par

- Une estimation \bar{C} de C
- Une matrice de covariance R .

La variance σ^2 des résultats de forage est évidemment nettement inférieure à la variance résiduelle résultant de l'ajustement des moindres carrés portant sur les mesures géophysiques. Cela n'est pas gênant dans l'optique bayésienne que nous adoptons ici et qui consiste à considérer C comme un vecteur aléatoire dont l'espérance mathématique est \bar{C} et dont la dispersion est matérialisée par la matrice de covariance R .

Les réajustements successifs de la loi de probabilité de C (moyenne \bar{C} et matrice de covariance R) nécessitent alors la connaissance a priori de la variance des résultats des forages.

Les responsables ont pu nous donner, compte tenu de leur expérience, une bonne estimation de cette variance.

En admettant, dans le cadre de l'optique bayésienne, que la dispersion du vecteur coefficient C autour de son espérance mathématique est suffisamment faible pour que $C - \bar{C}$ reste petit avec une forte probabilité, V peut s'écrire :

$$V(C) = V(\bar{C}) + \sum_i (C_i - \bar{C}_i) \frac{\partial V(\bar{C})}{\partial C_i}$$

au second ordre près.

Le vecteur \bar{C} étant connu à un stade donné de la reconnaissance, l'incertitude sur les réserves ne porte que sur les différences $C_i - \bar{C}_i$ qui sont en moyenne nulles puisque \bar{C} est l'espérance mathématique de C .

La formule approchée ci-dessus permet donc de calculer moyenne et variance de $V(C)$, soit

$$E [V(C)] = V(\bar{C})$$

$$\text{Var} [V(C)] = E \left[\sum_{ij} \frac{\partial V(\bar{C})}{\partial C_i} \frac{\partial V(\bar{C})}{\partial C_j} (C_i - \bar{C}_i) (C_j - \bar{C}_j) \right] = \alpha' R \alpha$$

où α est le gradient de $V(\bar{C})$ dans l'espace des C_i et R la matrice de covariance du vecteur aléatoire C .

Problème 2 - Localisation du puits suivant

La poursuite de la reconnaissance étant décidée, où placer le puits suivant pour avoir le maximum d'information ?

En d'autres termes, quelles doivent être les coordonnées géographiques (x,y) du futur puits pour que la réduction de la variance du volume des réserves soit la plus grande possible ?

Les formules récurrentes sur les matrices de covariance établies plus haut permettent d'écrire la relation qui lie les matrices de covariance R et $\xi R^{(1)}$ avant et après forage soit :

$$R^{(1)} = R - \frac{Rm(xy) m'(xy) R}{m'(xy) R m(xy) + \sigma^2}$$

où σ^2 est la variance de l'incertitude sur le résultat du forage et $m(x,y)$ le vecteur de composantes $(x^2, xy, y^2, x, y, 1)$ qui matérialise la position géographique (x,y) du nouveau puits.

La variation de la variance sur le volume V est alors

$$\Delta \text{Var}(V) = \frac{\alpha' R m(xy) m'(xy) R \alpha}{m'(xy) R m(xy) + \sigma^2}$$

Il suffit de maximiser la valeur absolue de $\Delta \text{Var}(V)$ par rapport à x et y pour localiser la position optimale du puits.

Il apparaît que la localisation optimale calculée résulte d'une double attraction :

- Celle des zones où les sondages sont peu nombreux,
- Celle des zones situées le long du grand axe du gisement.

Le calcul ci-dessus permet de réaliser le meilleur compromis entre ces deux attractions.

Si le premier type d'attraction est évident, le second résulte de considérations géométriques qui montrent que l'incertitude sur le volume est plus sensible à l'incertitude du niveau z le long du grand axe que le long du petit axe.

Etant donné la complexité de la formule ci-dessus, le calcul des coordonnées optima x et y du futur forage n'a pu être réalisé par des procédés analytiques classiques (annulation des dérivées partielles par rapport à x et y de l'expression de $\Delta \text{Var}(v)$).

Par ailleurs, pour des raisons diverses d'ordre géologique (porosité des roches en particulier) ou juridique (limitation géographique du permis de recherche) les responsables souhaitent se réserver la possibilité de s'écarter du point géographique optimal et connaître la perte d'information qui en résulterait.

Aussi, avons-nous bâti un programme machine pour calculer ΔVar en n'importe quel point (x,y) et tracer le réseau des lignes isovariance. L'examen de ce réseau permet ainsi aux responsables d'intégrer plus facilement des critères externes dans le choix de la localisation du forage.

6 - CONCLUSION

Les méthodes d'estimation séquentielle paraissent devoir connaître un grand développement dans les années à venir car la gestion automatisée des systèmes à comportement aléatoire impose un couplage permanent entre l'estimation et le contrôle.

Un programme, appelé ESOPE, actuellement opérationnel sur Control Data 3600, a été mis au point sur la base de ces méthodes. Il peut traiter des problèmes d'estimation séquentielle comportant jusqu'à 180 facteurs explicatifs et un nombre de mesures pratiquement illimité.

ANNEXE

1 - DEMONSTRATION DE LA FORMULE (14)

$$R^{(1)} = R (m' R m + \sigma^2)^{-1} R m m' R$$

à partir de

$$R^{(1)-1} = R + \frac{m m'}{\sigma^2}$$

où m est un vecteur colonne .

On pose :

$$A = R^{(1)} \quad \text{et} \quad A = R - B$$

et on cherche B telle que :

$$(1) \quad (R - B) \left(R^{-1} + \frac{m m'}{\sigma^2} \right) = I$$

La recherche est basée sur la remarque que $m' R m$ est un scalaire .

(1) est équivalente à :

$$(1') \quad B \left[R^{-1} + \frac{m m'}{\sigma^2} \right] = R \frac{m m'}{\sigma^2}$$

d'où on déduit

$$B \left[I + \frac{m m' R}{\sigma^2} \right] = \frac{R m m' R}{\sigma^2}$$

puis

$$B \left[m + \frac{m (m' R m)}{\sigma^2} \right] = \frac{R m m' R m}{\sigma^2}$$

soit :

$$B m \left[1 + \frac{m' R m}{\sigma^2} \right] = R m \frac{m' R m}{\sigma^2}$$

$$(2) \left[1 + \frac{m' R m}{\sigma^2} \right]$$

étant un scalaire, on peut écrire

$$(3) B - \frac{R m m' R}{\sigma^2 + m' R m} = 0$$

Les solutions de (3) sont

$$B = \frac{R m m' R}{\sigma^2 + m' R m} + B_0$$

B_0 étant telle que $B_0 m = 0$

Il faut alors déterminer B_0 de telle sorte que B soit solution de (1) ou de (1') qui lui est équivalente.

On sait que cette solution est unique puisque $R^{(1)}$ n'a qu'une seule inverse et on voit en reportant dans (1') que l'on doit avoir :

$$\frac{R m m' R}{\sigma^2 + m' R m} \left(R^{-1} + \frac{m m'}{\sigma^2} \right) + B_0 \left(R^{-1} + \frac{m m'}{\sigma^2} \right) = \frac{R m m'}{\sigma^2}$$

soit

$$\frac{R m m'}{\sigma^2 + m' R m} + \overbrace{\frac{R m m' R m m'}{\sigma^2 (\sigma^2 + m' R m)}}^{\text{scalaire}} + B_0 \left(R^{-1} + \frac{m m'}{\sigma^2} \right) = \frac{R m m'}{\sigma^2}$$

ou

$$R m m' \left[\frac{1}{\sigma^2 + m' R m} + \frac{m' R m}{\sigma^2 (\sigma^2 + m' R m)} - \frac{1}{\sigma^2} \right] + B_0 \left[R^{-1} + \frac{m m'}{\sigma^2} \right] = 0$$

soit

$$R m m' \left[\frac{\sigma^2 + m' R m - (\sigma^2 + m' R m)}{\sigma^2 (\sigma^2 + m' R m)} \right] + B_0 \left[R^{-1} + \frac{m m'}{\sigma^2} \right] = 0$$

donc

$$B_0 \left[R + \frac{m m'}{\sigma^2} \right] = 0$$

ce qui est satisfait en particulier si $B_0 = 0$

Comme la solution est unique, c'est celle que l'on obtient avec $B_0 = 0$ (c'est du reste la seule puisque $R^{-1} + \frac{m m'}{\sigma^2} = A^{-1}$ est régulière).

soit :

$$B = \frac{R m m' R}{\sigma^2 + m' R m}$$

et

$$A = R^{(1)} = R - R m m' R + \frac{1}{\sigma^2 + m' R m}$$

ce qu'il fallait démontrer.

Remarque :

C'est parce que m étant un vecteur, $m m'$ est de rang 1 donc non inversible qu'on ne peut multiplier les deux membres de (2) à droite par $m' (m m')^{-1}$ pour trouver B .

2 - ADJONCTION D'UNE MESURE ARTIFICIELLE - CONSEQUENCES

Considérons la mesure artificielle suivante :

$$\text{résultat} \quad y = 0$$

$$\text{vecteur facteur} \quad m' = (0, 0, 0, N, 0 \dots 0)$$

Les facteurs sont tous nuls sauf un qui est très grand (N)

Calculons d'après les formules d'itération la nouvelle estimation $\hat{C}_N^{(h)}$ de C

On a :

$$\hat{C}_N^{(h)} = \hat{C} + P m (m' P m + 1)^{-1} (y - m' \hat{C})$$

$m' P m$ est une forme quadratique qui donne : $p_{kk} N^2$

si k est le facteur dont la valeur est égale à N , $P m$ est un vecteur : $N \cdot \vec{p}_k$ où (\vec{p}_k) est un vecteur dont les composantes sont les éléments de la $k^{\text{ième}}$ ligne de la matrice P .

Enfin, le produit scalaire $m' \hat{C}$ donne $N \hat{C}_k$ ou \hat{C}_k est la $k^{\text{ième}}$ composante du vecteur paramètre estimé avant la mesure artificielle.

Il vient finalement :

$$\hat{C}_N^{(h)} = \hat{C} - \frac{N C_k \cdot (\vec{p}_k)}{p_{kk} N^2 + 1}$$

Quand N tend vers l'infini on obtient pour limite :

$$\hat{C}^{(h)} = \hat{C} - \frac{C_k \cdot (\vec{p}_k)}{p_{kk}}$$

La $k^{\text{ième}}$ composante est manifestement nulle.

De même :

$$P_N^{(h)} = P - Pm(m' P m + 1)^{-1} m' P$$

Il vient :

$$P_N^{(h)} = P - \frac{N \cdot \vec{p}_k, p'_k N}{p_{kk} N^2 + 1}$$

soit pour limite quand N tend vers l'infini :

$$P^{(h)} = P - \frac{\vec{p}_k p'_k}{p_{kk}}$$

La $k^{\text{ième}}$ ligne et la $k^{\text{ième}}$ colonne sont également nulles.

Nous montrons, ci-dessous, que la nouvelle estimation $\hat{C}^{(h)}$ du vecteur paramètre obtenu après adjonction de la mesure artificielle a des composantes $C_j^{(h)}$ relatives aux facteurs i autres que la $k^{\text{ième}}$, identiques à celles que l'on aurait estimées directement si l'on avait négligé tout au long des mesures l'information relative au $k^{\text{ième}}$ facteur.

Nous savons que :

$$\hat{C} = (M'M)^{-1} M'Y$$

ou encore :

$$(M'M) \cdot \hat{C} = M'Y$$

soit $m_{p,i}$ les éléments de la matrice M

et a_{ij} les éléments de la matrice $M'M$ (on a évidemment : $a_{ij} = \sum_r m_{ri} m_{rj}$).

On a donc :

$$\sum_j a_{ij} \hat{C}_j = \sum_r m_{ri} y_r \quad i = 1, 2, \dots, k, \dots, d$$

Si on avait négligé le $k^{\text{ième}}$ facteur tout au long des mesures, la matrice M correspondante se déduit de M par suppression pure et simple de la $k^{\text{ième}}$ colonne. Soit m_{rj}^* un élément de M^* , a_{ij}^* un élément de $M'M$ et ξ_j^* le vecteur d'estimation correspondant.

On a pour une de ses composantes j

$$\sum_j \alpha_{ij}^* \xi_j^* = \sum_r m_{ri} y_r \quad (i, j = 1 \dots d - 1)$$

On notera que si le vecteur C d dimensions ξ^* n'en possède que $(d - 1)$. Il en est de même pour $M^* M^{*'} \text{ vis à vis de } MM'$.

A partir de ces vecteurs et matrices à $(d - 1)$ dimensions, on peut en bâtir à d dimensions par insertion d'éléments ou de vecteurs nuls en $k^{\text{ième}}$ position, soit :

$$\begin{aligned} \hat{\xi} & \text{ d'élément } \hat{\xi}_j & (\xi_k = 0) \\ M & \text{ d'élément } \mu_{r,j} & (\mu_{r,k} = 0) \\ MM' & \text{ d'élément } \alpha_{ij} & (\alpha_{i,k} = a_{j,k} = 0) \end{aligned}$$

On a alors évidemment :

$$\begin{aligned} \mu_{r,j} &= m_{r,j} & \text{ pour } j \neq k \\ \alpha_{i,j} &= a_{ij} & \text{ pour } i, j \neq k \end{aligned}$$

et l'équation (A₂.2.1) s'écrit :

$$\sum_{j=k} a_{ij} \hat{\xi}_j = \sum_r m_{ri} y_r \quad (r, j = 1, 2 \dots k-1, K+1 \dots d)$$

Montrons que $\xi_j = G_j^{(R)}$ pour $j \neq k$

On doit avoir alors :

$$\sum_j \alpha_{ij} \left(\hat{C}_j - C_k \frac{p_{kj}}{p_{kk}} \right) = \sum_r \mu_{ri} y_r$$

soit pour $i \neq k$

$$\sum_{j=k} a_{ij} \left(\hat{C}_j - \hat{C}_k \frac{p_{kj}}{p_{kk}} \right) = \sum_r m_{r,i} y_r$$

où

$$\sum_j a_{ij} C_j = \sum_r m_{r,i} y_r \quad \text{pour tout } i$$

donc il suffit de vérifier que

$$\sum_{j=k} a_{ij} \frac{C_j}{p_{kk}} p_{k,j} + a_{i,k} C_k = 0 \quad \text{pour } i \neq k$$

soit

$$\sum_j a_{ij} p_{kj} \frac{C_k}{p_{kk}} = 0$$

ou

$$\sum_j a_{ij} p_{k,j}$$

est un élément (i,k) de la matrice $M'M P = I$ (matrice unité)

donc est nul pour ($i \neq k$), et la proposition est démontrée.

De même, on peut vérifier que tout terme de la matrice $Q^{(h)} = p^{(h)}$ ($M'M$) est nul, sauf ceux situés sur la diagonale, qui sont égaux à un, sauf celui d'indice k qui est nul.

Il s'ensuit que la matrice $P^{(k)}$, privée de la $k^{\text{ième}}$ ligne et de la $k^{\text{ième}}$ colonne est l'inverse de la matrice M^*M^* , et est bien la matrice de covariance du vecteur ξ^* . La matrice $P^{(h)}$, donc la ligne et la colonne d'indice k sont nulles, est la matrice de covariance du vecteur ξ , dont la $k^{\text{ième}}$ composante est une variable certaine de valeur nulle.

La vérification de l'égalité ci-dessus peut se faire comme suit : soit $q_{ij}^{(h)}$ un élément de la matrice $Q^{(h)}$.

Alors :

$$q_{i,j}^{(h)} = \sum_1 \alpha_{i,1} p_{1,j}^{(h)}. \text{ Si } i = k, \text{ cette expression est nulle } (\alpha_{i1} = 0)$$

Si $i \neq k$

$$q_{i,j}^{(h)} = \sum_1 a_{i,1} p_{1,j}^{(h)} - a_{i,k} p_{k,j}^{(h)} \text{ puisque } \alpha_{i1} = a_{i1} \text{ si } i \neq k$$

$$\alpha_{ik} = 0$$

$$= \sum_1 a_{i,1} p_{1,j}^{(h)} - a_{i,k} \left[p_{k,j} - \frac{p_{kj} p_{kk}}{p_{kk}} \right]$$

Le terme entre crochets est nul, d'où

$$q_{i,j}^{(h)} = \sum_1 a_{i,1} p_{1,j} - \frac{p_{i,k} p_{k,j}}{p_{kk}}$$

$$= \sum_1 a_{i,1} p_{1,j} - \frac{p_{k,j}}{p_{kk}} \sum_1 a_{i,1} p_{1,k}$$

Si $i = k, i \neq k$

les expressions du type $\sum_1 a_{i,1} p_{1,p}$ sont les éléments (i, j) du produit $M^*M \cdot P = I$

donc si

$$\begin{array}{ll} i \neq j & q_{i,j}^{(h)} = 0 \\ i = j & i \neq k \quad q_{i,j}^{(h)} = i \end{array}$$

BIBLIOGRAPHIE

- LEE R.C.K. - Optimal Estimation Identification and Control M.I.T. Press
- MALINVAUD E. - Estimation et prévision dans les modèles économiques régressifs - Revue de l'Institut International de Statistique Vol. 29 n° 2 - La Haye
- GRILICHES Z - A note on serial correlation bias in estimates of distributed lags. - Econometrica - Janvier 1961, p. 65 - 73.