

REVUE DE STATISTIQUE APPLIQUÉE

J.-P. MARCIANO

Plans d'expériences bayésiens en économétrie

Revue de statistique appliquée, tome 19, n° 4 (1971), p. 45-71

http://www.numdam.org/item?id=RSA_1971__19_4_45_0

© Société française de statistique, 1971, 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/>

PLANS D'EXPÉRIENCES BAYESIENS EN ÉCONOMÉTRIE (1)

J.-P. MARCIANO

Maître-Assistant de Mathématiques Statistiques
Faculté des Sciences Économiques d'Aix-en-Provence
Centre d'Économétrie

INTRODUCTION

Il existe en Econométrie de nombreux problèmes qui entraînent l'étude de I groupes de données avec, dans chacun, J observations ; c'est par exemple le cas pour l'étude de rendements en fonction de l'ouvrier et de la machine, pour l'étude de salaires en fonction de l'entreprise et du métier, ou en fonction de l'année et de la région.

La théorie classique des plans d'expériences, blocs complets au hasard, étudie des modèles de tels phénomènes sous la forme

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

où les y_{ij} sont les valeurs observées, μ l'effet général, α_i celui du traitement i, β_j celui du bloc j et e_{ij} "l'erreur" attachée à la valeur observée y_{ij} ; souvent, l'hétérogénéité des blocs offre moins d'intérêt. On se limite alors à l'étude du modèle $y_{ij} = \mu + \alpha_i + e_{ij}$ ou en incluant l'effet général dans celui de chaque traitement : $y_{ij} = \theta_i + e_{ij}$, modèle d'abord adopté ici.

D'autre part, les modèles traités ici sont à effet aléatoire ; les θ_i sont alors considérés comme des "observations" indépendamment distribuées d'une distribution dont l'étude se substitue à celle des valeurs individuelles des θ_i . Généralement, on suppose que ce sont des valeurs observées d'une variable ayant une distribution $N(\theta, \sigma_1^2)$, de même que les e_{ij} ont une distribution $N(0, \sigma_2^2)$; le problème est alors d'estimer θ et σ_2^2 .

De plus, c'est dans l'optique Bayésienne que cette étude est faite ; en effet, fréquemment en Economie il existe des informations a priori sur les paramètres ; sinon, leurs distributions a priori sont supposées non informatives.

Dans une première partie, est recherchée une estimation des espérances révisées et des variances des θ_i , dans le cadre d'une étude de l'effet des traitements. Un tel modèle avait déjà été partiellement étudié par Box et Tiao [1] ; mais ici de nouvelles différences fondamentales avec le modèle à effet fixe sont mises en relief, notamment dans un exemple original tandis qu'est abordé le problème de l'extension des résultats à un modèle hiérarchique $y_{ijk} = \theta_i + b_{ij} + e_{ijk}$ voire à un modèle plus développé.

(1) Cet article est extrait d'une thèse de Doctorat es Sciences Mathématiques devant être soutenue en 1972 à l'Université Paul Sabatier de Toulouse. Des procédures FORTRAN de traitement informatique des modèles étudiés peuvent être demandées au Centre d'Économétrie de la Faculté de Sciences Économiques d'Aix-en-Provence, où travaille l'auteur.

Dans la deuxième partie, l'intérêt se porte non plus sur les θ_1 mais sur les variances composantes, surtout σ_2^2 . Tiao collaborant avec Tan [2] a surtout recherché des propriétés asymptotiques de σ_2^2 . Ici l'étude est poussée plus loin, jusqu'à une estimation de la variance et de l'espérance révisée de σ_2^2 , qu'il est alors intéressant de comparer au mode de la distribution ; le modèle est maintenant similaire à celui de la première partie, mais avec l'effet général dégagé : $y_{1j} = \mu + \theta_1 + e_{1j}$, avec la distribution de $\theta_1 \sim N(0, \sigma_2^2)$. Cette partie s'achève sur la poursuite du traitement de l'exemple et sur l'extension des résultats sur σ_2^2 à des modèles hiérarchiques développés.

Au-delà de l'examen de quelques propriétés particulières des plans Bayésiens apparaît la définition d'une méthode exploitable basée sur un modèle Bayésien à effet aléatoire, simple ou hiérarchisé. Contrairement au cas des modèles à effet fixe, nous obtenons dans le modèle $y_{1j} = \theta_1 + e_{1j}$ avec $\theta_1 \sim N(\theta, \sigma_2^2)$, non seulement une valeur espérée révisée pour chaque θ_1 mais encore une estimation de la variance de chaque θ_1 et même des covariances entre les θ_1 , en plus de l'espérance et de la variance de σ_2^2 .

Première Partie

ETUDE DE L'EFFET DES TRAITEMENTS

1.1 - LE MODELE

Soient des groupes de I données correspondant à différents traitements et J observations par traitement. L'étude de l'effet du traitement i pour la j^{ème} observation peut conduire à un modèle

$$y_{1j} = \theta_1 + e_{1j} \quad \text{avec} \quad \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{array}$$

Les e_{1j} sont les "erreurs" sur chaque observation (i, j) ; on les suppose indépendamment et Normalement distribuées

$$e_{1j} \sim N(0, \sigma_1^2) \quad (1)$$

θ_1 représente l'effet associé au i^{ème} groupe. Les effets espérés sont des "observations" indépendamment distribuées d'une distribution à étudier, supposée ici Normale ; là réside la différence entre ce modèle, à effet aléatoire, et le modèle à effet fixe. Ainsi, en (8), sera traitée une étude, dans une taille de commune fixée, du nombre de commerces d'activité i par commune k ; il s'agit alors d'estimer l'effet espéré de chaque activité i, mais en supposant qu'il existe un processus p

(1) Le symbole \sim signifiera distribué comme une variable du type suivant.

$$p \sim N(\theta, \sigma_2^2) \tag{2}$$

générateur des effets espérés ; σ_2^2 est alors la variance traitement à traitement, σ_1^2 et σ_2^2 seront appelés composants de la variance ; on suppose de plus ici qu'il y a le même nombre d'observations par groupe.

Il est possible de construire un modèle similaire

$$y_{ij} = \mu + a_i + e_{ij}$$

avec cette fois ici μ paramètre de localisation commun et $a_i \sim N(0, \sigma_2^2)$. On est ramené au modèle précédent en posant μ égal à θ , espérance du processus générateur.

On a le tableau de départ suivant

Observations Traitements	1 J	
1	y_{11}	y_{1J}
⋮	⋮	⋮
I	y_{I1}	y_{IJ}

soit, en faisant apparaître pour le modèle, une matrice X de variables explicatives artificielles

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1J} \\ \vdots \\ y_{21} \\ \vdots \\ y_{2J} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{IJ} \end{bmatrix} = \begin{bmatrix} 10 \dots 0 \\ \vdots \\ 10 \dots 0 \\ \vdots \\ 01 \dots 0 \\ \vdots \\ 01 \dots 0 \\ \vdots \\ 0 \dots 01 \\ \vdots \\ 0 \dots 01 \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_2 \\ \vdots \\ \theta_I \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1J} \\ \vdots \\ e_{21} \\ \vdots \\ e_{2J} \\ \vdots \\ e_{I1} \\ \vdots \\ e_{IJ} \end{bmatrix}$$

ce que l'on pourra noter

$$\underline{y} = X \cdot T + E$$

On adoptera de plus certaines notations de l'analyse classique avec

$$T' = [\theta_1, \dots, \theta_I]$$

$$S_1 = \Sigma \Sigma (y_{1j} - \bar{y}_1)^2 = \underline{y}' \underline{y} - \underline{y}' X P \quad v_1 = I(J - 1) \quad m_1 = \frac{S_1}{v_1}$$

$$S_2 = J \Sigma (\bar{y}_1 - \bar{y})^2 = \underline{y}' X P - Y' N \underline{Y} \quad v_2 = I - 1 \quad m_2 = \frac{S_2}{v_2}$$

$$\text{avec } N = \begin{bmatrix} 1/IJ & 1/IJ \\ \vdots & \vdots \\ 1/IJ & 1/IJ \end{bmatrix} \quad \text{et } P' = [\bar{y}_1, \dots, \bar{y}_I], \quad \bar{y}_1 = \frac{\sum_j y_{1j}}{J}$$

On remarque toutefois qu'en analyse de variance classique, l'estimateur non biaisé de σ_2^2 est

$$\hat{\sigma}^2 = \left[\frac{S_2}{I - 1} \frac{S_1}{I(J - 1)} \right] / J = (m_2 - m_1) / J$$

Parfois dans certains plans, on obtient alors le résultat suivant : l'estimation du carré moyen interbloc < estimation du carré moyen de l'erreur, ce qui entraîne dans l'analyse de variance σ_2^2 négatif, résultat qui n'a pas de sens en pratique. Cette difficulté ne se présentera pas dans les modèles Bayesiens à effet aléatoire. Avant l'étude de σ_2^2 , sera entreprise ici celle de l'espérance et de la variance de chaque effet θ_1 .

1.2 - DEFINITION D'UNE DISTRIBUTION A PRIORI POUR $\mu, \sigma_1^2, \sigma_2^2$

Soit $\underline{\theta} = T' = [\theta_1, \theta_2, \dots, \theta_I]$ le vecteur des effets des traitements. Le tableau \underline{y} des observations défini plus haut est connu et $\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu$ sont inconnus.

Comme on a un processus générateur des θ_1 qui est $N(\mu, \sigma_2^2)$, et la distribution des e_{1j} qui est $N(0, \sigma_1^2)$, toujours indépendante de la première, le principe des probabilités composées donne ici

$$p(\mu, \sigma_1^2, \sigma_2^2, \underline{\theta}, \underline{y}) = p(\mu, \sigma_1^2, \sigma_2^2) p(\underline{\theta} / \sigma_2^2, \mu) p(\underline{y} / \underline{\theta}, \sigma_1^2) \quad (3)$$

avec $p(\mu, \sigma_1^2, \sigma_2^2)$ comme distribution a priori. Il paraît nécessaire de revenir à la définition d'une distribution a priori des paramètres, distribution qui caractérise l'analyse Bayésienne.

Si l'on a une densité Normale avec précision h et espérance μ inconnues, le tirage d'un échantillon x_1, \dots, x_n , donne connaissance d'une moyenne m' alors qu'au départ on ne sait rien sur μ . Un autre échantillon au hasard donnerait une autre valeur de m' mais pour μ fixé, m' est Normalement distribué

$$m' \sim e^{-\frac{h n}{2} (\mu - m')^2}$$

Ceci, qui est en analyse classique la vraisemblance de l'échantillon, sera appelé en analyse Bayésienne vraisemblance de μ , si le premier échantillon est connu.

Si un deuxième échantillon est tiré de moyenne m , on aura de même

$$m \sim e^{-\frac{h \cdot n}{2} (\mu - m)^2} \text{ pour } \mu \text{ fixé}$$

ce que l'on appelle dans notre cas, où μ est inconnu, la vraisemblance de μ . Si deux échantillons de moyennes m' puis m ont été prélevés, la vraisemblance de μ prend la forme

$$L(\mu) \sim \underbrace{e^{-\frac{h \cdot n'}{2} (\mu - m')^2}}_{1^\circ \text{ échantillon}} \underbrace{e^{-\frac{h \cdot n}{2} (\mu - m)^2}}_{2^\circ \text{ échantillon}} = e^{-\frac{h(n+n')}{2} \left\{ \frac{\mu - \frac{n'm + n'm'}{n+n'}}{\frac{n+n'}{n+n'}} \right\}^2} \quad (4)$$

Cette théorie, classique si ce n'est pour la dénomination vraisemblance de μ , est exposée ici pour une distribution Normale ; elle peut être appliquée à une distribution exponentielle, χ^2 ou autre.

L'analyse Bayésienne introduit de plus sur μ une information que le praticien peut traduire au point de vue mathématique. Cette information peut être considérée comme une connaissance équivalente à celle qu'aurait fourni un échantillon, ce qui conduit à donner à la connaissance a priori de μ la forme d'un résumé exhaustif objectif

$$e^{-\frac{h \cdot n}{2} (\mu - m')^2}$$

Les informations a priori sur le paramètre inconnu μ prennent alors la forme d'une distribution de probabilités. De plus, la connaissance d'un échantillon $x_1 \dots x_n$ de moyenne m permet alors d'exprimer la vraisemblance de μ par rapport à cet échantillon. La probabilité révisée de μ s'écrit alors comme au (4)

$$p''(\mu) \sim e^{-\frac{h \cdot n}{2} (\mu - m')^2} \quad e^{-\frac{h \cdot n}{2} (\mu - m)^2}$$

à priori de μ vraisemblance de μ
(équivalente à un (2° échantillon)
1° échantillon)

1.3 - LA DISTRIBUTION A PRIORI NON INFORMATIVE

Ci-dessus a été expliquée la construction d'une a priori Normale et son rôle pour un paramètre μ . Or face à 2 distributions $N(\mu, \sigma_1^2)$ et $N(0, \sigma_2^2)$ avec $\mu, \sigma_1^2, \sigma_2^2$ inconnus, il faut construire une a priori $p(\mu, \sigma_1^2, \sigma_2^2)$ avec $\sigma_1^2 > 0$ et $\sigma_2^2 > 0$.

Soit alors une loi Multinormale centrée de dimension 2 (plus généralement n), avec 3 paramètres, les variances des 2 variables et la corrélation ρ . Si l'on tire un échantillon de n vecteurs $X_1 \dots X_n$ avec $X_1 = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix}$ on peut former la matrice V symétrique

$$V = \begin{pmatrix} \sum (u - \bar{u})^2 & \sum (u - \bar{u}) (v - \bar{v}) \\ \sum_{ij} (u - \bar{u}) (v - \bar{v}) & \sum (v - \bar{v})^2 \end{pmatrix}$$

L'échantillon des $X_i \begin{pmatrix} u_i \\ v_i \end{pmatrix}$ est tiré au hasard donc les éléments de la matrice V sont, avant le tirage, aléatoires. La probabilité de l'échantillon $X_1 \dots X_n$ s'exprime simplement en fonction de celle des éléments de V qui suivent eux-même une distribution de Wishart ; on trouvera à ce sujet une démonstration des plus récentes de ce résultat dans l'ouvrage de DeGroot "Optimal Statistical Decisions" (p. 57). Comme dans le cas d'une loi Normale simple, si les paramètres sont inconnus, on considèrera que des informations a priori à leur sujet équivalent à la connaissance d'un échantillon et se présenteront donc sous la forme d'une distribution de probabilité de Wishart appelée distribution à priori.

Si la loi Multinormale n'est pas centrée, on a 5 paramètres, supposés maintenant inconnus, groupés dans un vecteur des moyennes M et d'autre part dans V , ou bien dans R matrice des précisions. L'a priori jointe de M et R , connaissance jugée équivalente à celle d'un échantillon tiré dans ces conditions, prend la forme d'une distribution Normal-Wishart, produit d'une distribution Multinormale par une distribution de Wishart. La distribution a posteriori jointe de M et R est le produit de l'a priori par la vraisemblance obtenue par l'échantillon ; on obtient à nouveau une distribution Normal-Wishart (DeGroot, ouvrage cité plus haut, p. 178, théorème 1).

Si l'on a très peu de connaissance a priori des paramètres ou même si l'on veut minimiser les risques des erreurs, on peut principalement faire tendre les variances vers l'infini dans le produit de l'a priori par la vraisemblance ; il n'est évidemment pas question de faire tendre la variance vers l'infini dans la distribution a priori seule, qui est Normal-Wishart et se comporterait comme une distribution Normale en ce cas. Mais en faisant tendre la variance vers l'infini dans le produit de l'a priori par la vraisemblance, le résultat obtenu pour la distribution a posteriori est encore une densité. On remarque qu'on aurait obtenu le même résultat en effectuant le produit de la vraisemblance par la quantité $\frac{1}{\sigma_1^2 \sigma_{12}^2}$ avec $\sigma_{12}^2 = \sigma_1^2 + j\sigma_2^2$ (DeGroot, ouvrage cité plus haut, 10.3, p. 197). On a donc

$$p''(M \text{ et } R) = \text{à priori} \cdot \text{vraisemblance}$$

Si la connaissance a priori est faible ou si on veut minimiser les risques d'erreur

$$p''(M \text{ et } R) = (\text{a priori} \cdot \text{vraisemblance})$$

Variance $\longrightarrow \infty$

$$\text{avec } (\text{a priori} \cdot \text{vraisemblance}) \longrightarrow \left(\frac{1}{\sigma_1^2 \sigma_{12}^2} \cdot \text{vraisemblance} \right)$$

le facteur multiplicatif est appelé distribution a priori non informative. Il est évident que ce facteur multiplicatif n'est pas intégrable sur D si

$$D = \{ \mu \in]-\infty, +\infty[, \sigma_1^2 > 0, \sigma_{12}^2 > \sigma_1^2 \}$$

ce n'est donc pas une densité et en ce sens l'appellation de distribution a priori pour ce facteur (1), n'est peut-être pas heureuse. Néanmoins, nous retiendrons qu'en cas de diffusion de la connaissance a priori, le produit (a priori . Vraisemblance) est encore une densité qui prend la forme $\left(\frac{1}{\sigma_1^2 \sigma_{12}^2} \cdot \text{Vraisemblance} \right)$. Cette densité sera appelée "révisée jointe des paramètres".

1.4 - L'ANALYSE BAYESIENNE

D'après la relation (3) une fois donné l'échantillon \underline{y} , on a la distribution révisée jointe de $(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu)$. Comme il existe un processus générateur de θ , avec ensuite $\underline{\theta}$ générateur de \underline{y}

$$p(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu / \underline{y}) = \text{à priori} \cdot \text{vraisemblance}$$

Si pour minimiser les risques d'erreurs, on suppose l'a priori non informative.

$$p(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu / \underline{y}) \sim \frac{1}{\sigma_1^2 \sigma_{12}^2} p(\underline{\theta} / \sigma_2^2, \mu) \cdot p(\underline{y} / \underline{\theta}, \sigma_1^2) \quad (5)$$

avec $p(\underline{y}) = p(y_{11} \text{ et } y_{12} \dots \text{ et } y_{1j} \text{ et } \dots \text{ et } y_{11} \text{ et } \dots y_{1j})$

$$p(\underline{y} / \underline{\theta}, \sigma_1^2) = (\sigma_1^2)^{-\frac{1}{2}j} \exp \left[-\frac{1}{2\sigma_1^2} \left(S_1 + J \sum_1 (\theta_1 - \bar{y}_1)^2 \right) \right]$$

et $p(\underline{\theta}) = p(\theta_1, \text{ et } \theta_2 \text{ et } \dots \theta_1)$

$$p(\underline{\theta}) = (\sigma_2^2)^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma_2^2} \sum_1 (\theta_1 - \mu)^2 \right]$$

Si après avoir remplacé la vraisemblance par sa valeur dans (5), on intègre par rapport à μ , il vient

$$\int_{\mathbb{R}} p(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu / \underline{y}) \cdot d\mu = p(\underline{\theta}, \sigma_1^2, \sigma_2^2 / \underline{y}) \\ \sim (\sigma_1^2)^{-\frac{1}{2}(j+1)} \sigma_2^2^{-\frac{1}{2}(j+1)} \sigma_{12}^{-2} \exp \left[-\frac{1}{2\sigma_1^2} \left(S_1 + J \sum (\theta_1 - \bar{y}_1) + \frac{\sigma_1^2}{\sigma_2^2} \sum (\theta_1 - \bar{\theta})^2 \right) \right] \quad (5')$$

On pose $Z = \frac{\sigma_1^2}{\sigma_{12}^2} = \frac{\sigma_1^2}{\sigma_1^2 + J\sigma_2^2}$ d'où $\frac{\sigma_1^2}{\sigma_2^2} = \frac{J Z}{1 - Z}$ et

(1) Utilisée couramment par la littérature Anglo-Saxonne à ce sujet.

$$\begin{aligned}
p(\underline{\theta}, \sigma_1^2, Z/\underline{y}) &\sim (\sigma_1^2)^{-\frac{1(J+1)+1}{2}} \sigma_1^{1-1} (\sigma_2^2)^{-\frac{1-1}{2}} \sigma_{12}^2 \exp [\dots] \\
&\sim (\sigma_1^2)^{-\frac{1(J+1)-1}{2}} \left(\frac{JZ}{1-Z}\right)^{\frac{1-1}{2}} Z \exp \left[-\frac{1}{2\sigma_1^2} \left(S_1 + J \left(\sum_1 (\theta_1 - \bar{y}_1) \right)^2 + \frac{Z}{1-Z} \sum (\theta_1 - \bar{\theta})^2 \right) \right]
\end{aligned}
\tag{5''}$$

Avec l'expression (5') sous cette forme, il sera plus facile de chercher la distribution conditionnelle de $\underline{\theta}$ pour une valeur de Z donnée puis en intégrant, grâce à la distribution marginale de Z, obtenir la distribution marginale de $\underline{\theta}$. Le cheminement suivi est donc

$$p(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu/\underline{y}) = \text{révisée jointe} = \text{à priori} \cdot \text{vraisemblance}$$

$$p(\underline{\theta}, \sigma_1^2, \sigma_2^2/\underline{y}) = \int_D p(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu/\underline{y}) d\mu$$

$$\text{changement de variable en } Z = \frac{\sigma_1^2}{\sigma_2^2}$$

$$\begin{aligned}
p(\underline{\theta}/\underline{y}, \forall z) &= \int_D p(\underline{\theta}, z/\underline{y}) dz = \int p(Z/\underline{y}) \cdot p(\underline{\theta}/z/\underline{y}) dz \tag{6} \\
&\quad \uparrow \qquad \qquad \qquad \uparrow \qquad \qquad \qquad \uparrow \\
&\quad \text{marginale} \qquad \qquad \text{marginale} \qquad \text{conditionnelle} \\
&\quad \text{de } \underline{\theta} \qquad \qquad \text{de } Z \qquad \qquad \text{de } \underline{\theta}
\end{aligned}$$

On cherche à définir pour chaque traitement θ_1 une espérance et une variance révisées. Pour obtenir la densité marginale de $\underline{\theta}$, il faut étudier les densités, conditionnelle de $\underline{\theta}$ et marginale de Z.

1.5 - LA DISTRIBUTION CONDITIONNELLE DE $\underline{\theta}$ SUR Z

La connaissance de Z équivaut à celle de la précision du processus. Il faut sortir un instant du problème considéré et étudier quelques propriétés d'un processus Multinormal quelconque, à précision connue. Ces propriétés seront utiles par la suite.

1.5.1 - Quelques propriétés d'un processus Multinormal

Soit un processus multinormal de dimension j, défini comme générant I variables de densités identiques X_1, X_2, \dots, X_I avec pour chacune même densité

$$F_N(X/M, H) = (2\pi)^{-\frac{I}{2}} |H|^{1/2} \exp \left(-\frac{1}{2} (X - M)' H (X - M) \right)$$

et $E(X/M, H) = M$

Var $(X/M, H) = H^{-1}$, H est la précision

On définit aussi le scalaire précision moyenne $h = |H|^{\frac{1}{J}}$ et la matrice des précisions relatives

$$\eta = \frac{H}{h} = \begin{vmatrix} \frac{h_{11}}{h} & \dots & \frac{h_{1j}}{h} \\ \vdots & & \vdots \\ \frac{h_{j1}}{h} & & \frac{h_{jj}}{h} \end{vmatrix} \quad \text{on a } |\eta| = 1$$

X est le vecteur des X_1, \dots, X_I avec $X_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1j} \end{pmatrix}$

Avec ces notations

$$F_N(X/M, h \eta) = (2\pi)^{-J/2} h^{J/2} \exp \left\{ -\frac{h}{2} (X - M)' \eta (X - M) \right\}$$

$$E(X/M, h \eta) = M$$

$$\text{Var}(X/M, h \eta) = (h \eta)^{-1}.$$

Supposons que seule la précision relative soit connue. Posons

$$\underline{m} = \frac{1}{I} \sum_{k=1}^I X^k \quad \text{si le processus génère } X_1, \dots, X_I$$

$$N = I \eta$$

$$v = J(I - 1)$$

$$V = \frac{1}{v} \sum_{k=1}^I (X^k - \underline{m})' \eta (X^k - \underline{m})$$

La vraisemblance de l'échantillon devient

$$L(\underline{m}, N, v, v) \sim h^{\frac{J+v}{2}} \exp \left\{ \frac{h}{2} v - \frac{1}{2} h(\underline{m} - M)' N (\underline{m} - M) \right\}$$

Les paramètres \underline{m} , N , O , v sont alors quatre statistiques connues par l'échantillon

Comme nous souhaitons utiliser l'information contenue dans les statistiques incomplètes (\underline{m}, N) ou (v, v) , on peut aussi décomposer la vraisemblance

$$L(\underline{m}, N) \sim h \quad e^{-h/2(\underline{m}-M)' N (\underline{m}-N)}$$

$$L(V, v) \sim h^{v/2} e^{-\frac{hvv}{2}}$$

$$\text{Soit } p \text{ le rang de } N \quad \begin{cases} = 0 & \text{si } I = 0 \\ = J & \text{si } I > 0 \end{cases}$$

Alors l'expression de la densité devient

$$e^{-\frac{h}{2}(\underline{m}-M)'N(\underline{m}-M)} e^{-\frac{h\nu v}{2}} \frac{p+\nu}{h^{\frac{p+\nu}{2}}}$$

et avec les conventions $\nu = 0$ si v est inconnu

$$N = 0 \text{ si } \underline{m} \text{ est inconnu}$$

cette nouvelle densité se réduit aux vraisemblances $L(\underline{m}, N)$ ou $L(v, \nu)$. De plus ici ν est supposé connu.

Etudions alors la distribution conjointe à 2 variables (M, h) , quand la précision relative η est connue, l'espérance M et la précision moyenne h étant aléatoires. La distribution à priori qui leur convient le mieux est, d'après les études de Raiffa et la forme de la vraisemblance, la Gamma Normale (Raiffa et Schlaifer, "Applied Statistical Decision Theory" (Chapitre 12. A), ainsi définie

$$F_N^J(M, h/\underline{m}, v, N, \nu) = F_N(M/\underline{m}, hN) F_{\gamma_2}(h/v, \nu) \quad (6')$$

$$\sim h^{p/2} e^{-\frac{h}{2}(M-\underline{m})'N(M-\underline{m})} \frac{v}{h^{\frac{p}{2}+1}} e^{-\frac{h\nu v}{2}}$$

Si le juge de décision prend $N' = 0$ (variance infinie), comme p est le rang de N , on est alors ramené à une distribution a priori non-informative.

Si la distribution a priori de (M, h) est Gamma Normale de paramètres (m', v', ν', N') où N' de rang p' , si l'échantillon donne une statistique (m, v, ν, N) où N est de rang p , alors Raiffa montre que la distribution a posteriori de (M, h) , produit de l'a priori par la vraisemblance, est Gamma-Normale de paramètres

$$N'' = N + N' \quad p'' \text{ rang de } N''$$

$$\underline{m}'' = N''^{-1} (N' \underline{m}' + N \underline{m})$$

$$\nu'' = \nu' + p' + \nu + p - p''$$

$$\nu'' = \frac{J' v' + (\underline{m}')^t N \underline{m}' + \nu v + \underline{m}^t N \underline{m} - (\underline{m}'')^t N \underline{m}''}{\nu''} \quad (1)$$

Raisonnant dès lors uniquement sur des distributions a posteriori, on notera sans risque de confusion les paramètres \underline{m}'' , N'' , O'' , ν'' d'une manière moins lourde \underline{m} , N , O , ν

De la densité jointe ainsi obtenue de (M, h) , on déduit la distribution marginale de h seul qui est Gamma² ainsi que la conditionnelle de h pour M connu.

Comme la conditionnelle de M pour h donné est

$$F_N(M/\underline{m}, hN)$$

 (1) (m') représente la matrice transposée de m' , le symbole, utilisé jusqu'ici pour la transposée d'une matrice ayant dû être remplacée par t pour éviter toute confusion.

on obtient la densité marginale de M par l'intégration

$$\int_0^{\infty} F_N(M/\underline{m}, hN) \cdot F_{\gamma_2}(h/v, v) dh$$

↓
↓
 conditionnelle marginale
 de M en h de h

On reconnaît dans l'intégrale de ce produit la définition d'une densité de Student dans $R^{(j)}$. Cette densité se développe sous la forme

$$F_S^{(j)}(M/\underline{m}, N, v) = \frac{v^{v/2} \left(\frac{v+j}{2} + 1\right)!}{\pi^{j/2} \left(\frac{v}{2} - 1\right)!} [v + (M - \underline{m})' N (M - \underline{m})]^{-\frac{v+j}{2}} |N|^{1/2} \quad (6'')$$

L'expression d'une densité de Student classique est

$$\int_0^{\infty} F_N^{(j)}(M/\underline{m}, hN) F_{\gamma_2}(h/1, v) dh$$

Il faudra donc faire dans (6') le changement de variable $h_1 = h_v$ pour se ramener à la forme classique ci-dessus, avec

$$F_{\gamma_2}(h/v, v) = \frac{1}{\left(\frac{v}{2} - 1\right)!} \frac{v v}{2} e^{-\frac{vvh}{2}} \left(\frac{vvh}{2}\right)^{\frac{v}{2}-1}$$

Pour calculer les moments, il suffit de renverser l'ordre de l'intégration dans la densité de Student

$$\begin{aligned} \mu &= \int_{-\infty}^{+\infty} \left[\int_0^{\infty} M F_N^{(r)}(M/\underline{m}, hN) \right] f_{\gamma_2}(h/v, v) dh dz \\ &= \int_0^{\infty} \underline{m} \cdot f_{\gamma_2}(h/v, v) dh = \underline{m} \end{aligned} \quad (7)$$

On obtient $E(M/\underline{m}, v, N, v) = \underline{m}$ avec $v > 1$

On montrerait de même

$$\text{Var}(M/\underline{m}, v, N, v) = N^{-1} v \frac{v}{v-2} \quad (8)$$

On a ainsi calculé la variance et l'espérance révisée de \underline{M} ; Dunet et Sobel ont étudié dans *Biometrika* (1954, p. 153) des propriétés plus intéressantes de la distribution de Student mais qui ne s'appliquent que dans R^2 .

1.5.2 - Application à la conditionnelle de $\underline{\theta}$ sur Z

Pour revenir au problème de cette première partie, on cherche la distribution marginale révisée de $\underline{\theta}$, avec une distribution a priori de $(\underline{\theta}, \underline{c}_1^2)$,

pour Z fixé, supposée non-informative, ce qui est un cas particulier de la distribution a priori Normal Gamma étudiée au 1.5.1.

Par le produit de l'a priori par la vraisemblance, on a obtenu l'a posteriori jointe de $(\underline{\theta}, \sigma_1^2, \sigma_2^2, \mu)$.

En intégrant par rapport à μ , on a obtenu l'a posteriori du triplet $(\underline{\theta}, \sigma_1^2, \sigma_2^2)$.

Pour faciliter la suite des calculs, on a posé $Z = \frac{\sigma_1^2}{\sigma_{12}^2}$.

La distribution marginale révisée de $\underline{\theta}$ sera donnée par

$$p(\underline{\theta}/y, Z) = \int p(Z) \cdot p(\underline{\theta}/Z) dz$$

\uparrow \uparrow \uparrow
 marginale marginale conditionnelle
 de $\underline{\theta}$ révisée de Z de $\underline{\theta}$ pour Z donné

La conditionnelle de $\underline{\theta}$ pour Z donné est obtenue à partir de l'a posteriori du triplet $(\underline{\theta}, \sigma_1^2, \sigma_2^2)$ devenu par le changement de variable $(\underline{\theta}, \sigma_1^2, Z)$.

Il s'agit d'étudier ici cette conditionnelle de $\underline{\theta}$; on peut utiliser les résultats du paragraphe précédent; dans la conditionnelle de $\underline{\theta}/Z$ est fixé, on est en présence d'un processus Multinormal à précision connue η avec $M = \underline{\theta} \dots \dots$; on a vu plus haut que l'a posteriori correspondante de M pour Z fixé, c'est-à-dire la conditionnelle de $\underline{\theta}$ pour Z fixé, est une t distribution de dimension I à $\nu_1 + \nu_2$ degrés de liberté. Il faut remarquer que si la précision relative, (ici Z) est connue, la précision moyenne R peut être quelconque dans son domaine de définition.

L'expression (6'') de cette densité nous donne, après changement de variable

$$p(\underline{\theta}/y, Z) \sim \left\{ 1 + \frac{J(\underline{\theta} - \hat{\underline{\theta}}(Z))' (I - Z J \underline{1}' \underline{1}) (\underline{\theta} - \hat{\underline{\theta}}(Z))}{\nu s^2(Z)} \right\}^{-\frac{\nu_1 + \nu_2 + I}{2}}$$

avec $\underline{1}$ est un vecteur $(I \times 1)$ de 1

$$s^2(Z) = (S_1 + ZS_2) (1 + Z)/\nu$$

$$\nu = \nu_1 + \nu_2 \quad \hat{\underline{\theta}}(Z) = (\hat{\theta}_1(Z), \dots, \hat{\theta}_I(Z))$$

$$\hat{\theta}_1(Z) = \bar{y}_1 - Z(\bar{y}_1 - \bar{y})$$

Le rôle de M est ici joué par $\underline{\theta}$, celui de la statistique \underline{m} par $\hat{\underline{\theta}}(Z)$; on peut utiliser les résultats (7) et (8) pour l'espérance et la variance révisée de θ_1 : on obtient

$$E(\theta_1/Z, y) = \hat{\theta}_1(Z) \quad \text{pour tout } i \tag{7e}$$

$$= \bar{y}_1 - Z(\bar{y}_1 - \bar{y})$$

$$\text{Var}(\theta_1/Z, y) = \frac{\nu}{J(\nu - 2)} \left[1 + \frac{Z}{I(1 - Z)} \right] s^2(Z) \quad (8')$$

et même une covariance commune

$$\text{Cov}(\theta_1, \theta_j / Z, y) = \frac{\nu}{J(\nu - 2)} \left[\frac{Z}{I(1 - Z)} \right] s^2(Z)$$

1.6 - LES DISTRIBUTIONS MARGINALES DE Z ET $\underline{\theta}$

On a maintenant la distribution conditionnelle de θ . L'expression (6) nous rappelle comment obtenir la marginale révisée de $\underline{\theta}$

$$\begin{array}{ccccc} p(\underline{\theta}/y, \forall z) & = & \int_0^1 p(z/y) & p(\underline{\theta}/z, y) dz \\ \uparrow & & \uparrow & \uparrow \\ \text{marginale révisée} & & \text{marginale} & \text{conditionnelle de } \underline{\theta} \\ \text{cherchée ou } \underline{\theta} & & \text{de Z à trouver} & \text{trouvée ci-dessus} \end{array}$$

C'est l'expression (5'') de 1.3 de $p(\theta, \sigma_1^2, z/y)$ qui est utilisée pour le calcul de la distribution marginale de z ; on obtient en intégrant par rapport à θ et en faisant intervenir la contrainte $C (\sigma_1^2 < \sigma_{12}^2)$ dans p

$$p_1(\sigma_1^2, \sigma_{12}^2/y) = \frac{p(\sigma_1^2, \sigma_{12}^2/y)}{p(C/y)} = \frac{(\nu_1 m_1) p\left(\chi_{\nu_1}^{-2} = \frac{\sigma_1^2}{\nu_1 m_1}\right) (\nu_2 m_2)^{-1} p\left(\chi_{\nu_2}^{-2} = \frac{\sigma_{12}^2}{\nu_2 m_2}\right)}{p\left(\chi_{\nu_2}^2 > \frac{\nu_1 m_1}{\nu_2 m_2}\right)}$$

avec χ^{-2} chi deux inversé

soit après transformations et intégration

$$p(z/y) = \frac{\frac{m_2}{m_1} p\left(F_{\nu_2, \nu_1} = z \frac{m_2}{m_1}\right)}{p\left(F_{\nu_2, \nu_1} < \frac{m_2}{m_1}\right)} \quad \text{avec } 0 < z < 1$$

on remarque que la contrainte $\sigma_1^2 < \sigma_{12}^2$ devient $0 < z < 1$.

$p(F_{\nu_2, \nu_1})$ est la densité d'une variable F ; le moment d'ordre r de z est alors

$$\mu_r' = E(z^r) = \frac{B\left(\frac{\nu_2}{2} + r, \frac{\nu_1}{2} - r\right) \text{Ix}\left(\frac{\nu_2}{2} + r, \frac{\nu_1}{2} - r\right)}{B\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)}, \quad \text{Ix}\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right) \frac{S_1}{S_2}$$

ce qui permet de calculer espérance et variance de z ; $B(p, q)$ étant la fonction Beta et $\text{Ix}(p, q)$ la mesure Beta incomplète tabulée par Pearson, avec

$$x = \frac{S_2}{S_1 + S_2}$$

Nous obtenons la distribution marginale de $\underline{\theta}$ par intégration

$$p(\underline{\theta}/y) = \int_0^1 p(\underline{\theta}/z, y) \cdot p(z/y) dz$$

On obtient notamment les résultats fondamentaux :

$$E(\theta_1/y) = \hat{\theta}_1 = \bar{y}_1 - \mu_1' (\bar{y}_1 - \bar{y})$$

$$\begin{aligned} \text{Var}(\theta_1/y) = E \frac{\text{Var} \theta_1/z}{Z} + E \frac{(\hat{\theta}_1(z) - \hat{\theta}_1)^2}{Z} &= \frac{I(S_1 + \mu_1' S) - (I - 1) (\mu_1' S_1 + \mu_2' S_2)}{IJ(\nu_1 + \nu_2 - 2)} \\ &+ (\bar{y}_1 - \bar{y})^2 \mu_2 \text{ avec } \mu_2 = \mu_2' - \mu_1'^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(\theta_i \theta_j/y) &= E \frac{\text{Cov}(\theta_i \theta_j/z)}{Z} + E \frac{(\hat{\theta}_i(z) - \hat{\theta}_i) (\hat{\theta}_j(z) - \hat{\theta}_j)}{Z} \\ &= \frac{I(S_1 + \mu_1' S_2) - (I - 1) (\mu_1' S_1 + \mu_2' S_2)}{ij(\nu_1 + \nu_2 - 2)} + (\bar{y}_i - \bar{y}) (\bar{y}_j - \bar{y}) \mu_2 \end{aligned}$$

1.7 - EXEMPLE NUMERIQUE

Il semble que ces résultats n'aient jamais été examinés en détail en vue d'un traitement numérique systématique de modèles dans l'optique Bayésienne à effet aléatoire.

Plusieurs remarques très importantes s'imposent alors ; en effet fixe, on a l'espérance révisée de θ_1 sous la forme $E(\theta_1/y) = \bar{y}_1$; en effet aléatoire l'espérance révisée devient $\bar{y}_1' - \mu_1' (\bar{y}_1 - \bar{y})$; or μ_1' est l'espérance de z , variable comprise entre 0 et 1. Ainsi, en effet aléatoire, dans l'espérance révisée de θ_1 le terme $\mu_1' (\bar{y}_1 - \bar{y})$ joue le rôle de terme correcteur qui rapproche les espérances de \bar{y} .

Il faut également noter qu'en effet fixe, il y a une seule variance intra $\text{Var}(\theta_1/y) = \frac{S_1}{J(\nu_1 - 2)}$; en effet aléatoire, une variance intra est définie pour chaque θ_i ; par suite, si en effet fixe, la covariance entre deux θ_i est nulle, il n'en est pas du tout de même en effet aléatoire et le tableau des covariances est une nouvelle source de renseignements.

La conduite des calculs sera étudiée sur un exemple dont les données nous ont été aimablement proposées par le Laboratoire C.N.R.S. d'Economie et de Sociologie du Travail à Aix-en-Provence [5].

Auparavant une remarque s'impose en ce qui concerne le calcul des moments de z

$$\mu_r' = \frac{B\left(\frac{\nu_2}{2} + r, \frac{\nu_1}{2} - r\right)}{B\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \frac{I_x\left(\frac{\nu_2}{2} + r, \frac{\nu_1}{2} - r\right)}{I_x\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \frac{S_1}{S_2} \text{ avec } x = \frac{S_2}{S_1 + S_2}$$

Pour éviter d'avoir à étudier un rapport de valeurs numériques très petites des Fonctions Beta, il est préférable de décomposer ces dernières en Fonctions Gamma et de simplifier les rapports de la forme $(x + 1) / (x)$. Il vient alors

$$\mu'_1 = \frac{\nu_2}{\nu_1 - 2} \frac{I_x\left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2} - 1\right)}{I_x\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \frac{S_1}{S_2} \quad \nu \neq 2$$

$$\mu'_2 = \frac{\nu_2(\nu_2 + 2)}{(\nu_1 - 4)(\nu_1 - 2)} \frac{I_x\left(\frac{\nu_2}{2} + \frac{\nu_2}{2} - 2\right)}{I_x\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \frac{S_1}{S_2} \quad \nu_1 \neq 4$$

Nous utiliserons ces expressions dans l'exemple traité maintenant. Il s'agit d'étudier le taux de croissance annuel des salaires par entreprise et par métier suivant un modèle

$$s_{ij} = \theta_i + e_{ij}$$

où s_{ij} est l'indice de croissance pour l'entreprise i et le métier j , θ_i est l'effet de l'entreprise i . De tels modèles ont dans ce cas toujours été traités en analyse de variance classique uniquement, avec estimation des variances inter et intra-classes ; or, la méthode Bayésienne à effet aléatoire exposée dans cette première partie permettra de définir en outre une variance par entreprise et une covariance entre les entreprises.

Le tableau 1 nous fournit les données.

Tableau 1
Taux de croissance des salaires sur une
année par entreprise et par métier

Entreprise	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Métier																
	4.06	4.02	4.74	3.30	5.39	1.35	2.06	2.44	2.36	2.21	4.14	4.42	4.59	2.51	2.30	3.13
	5.77	2.38	4.55	2.62	5.21	1.74	3.22	2.52	4.48	4.32	5.38	3.75	4.62	3.27	5.23	2.10
	4.92	3.44	5.24	2.93	5.01	2.28	4.01	2.28	4.72	4.38	4.66	3.57	3.77	2.92	2.89	1.94
	4.79	3.48	5.33	3.44	5.85	1.81	3.86	3.17	3.92	3.40	4.59	3.76	4.40	3.10	3.10	2.32
Moyenne	4.88	3.33	4.97	3.07	5.37	1.79	3.29	2.60	3.87	3.58	4.69	3.87	4.33	2.95	3.33	2.37

Nous avons ici 4 métiers et 16 entreprises, c'est-à-dire, avec les notations classiques

$$S_1 = \sum \sum (s_{ij} - \bar{s}_i) = 21,64 \quad \nu_1 = I(J - 1) = 48$$

$$S_2 = J \sum (\bar{s}_i - \bar{s})^2 = 60,24 \quad \nu_2 = I - 1 = 15$$

En effet fixe, l'espérance révisée de chaque θ_1 est \bar{s}_1 donné dans le tableau 2.

En effet aléatoire $E(\theta_1/\underline{s}) = \bar{s}_1 - \mu_1' (\bar{s}_1 - \bar{s})$

avec
$$\mu_1' = \frac{\nu_2}{\nu_1^{-2}} \frac{I_x\left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2} - 1\right)}{I_x\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right)} \frac{S_1}{S_2} \quad \text{et} \quad x = \frac{S_2}{S_1 + S_2}$$

$\mu_1' = 0.103$; il est alors possible de dresser le tableau 2 des espérances révisées après le calcul de \bar{s} ($\bar{s} = 3.65$), les mesures Beta incomplètes ayant été calculées, pour les valeurs indiquées, par une table de Pearson.

Tableau 2
Espérance révisée des taux,
par entreprise

Entreprise	Effet fixe $E(\theta_1/\underline{s}) = \bar{s}_1$	Effet aléatoire $E(\theta_1/\underline{s}) = \bar{s}_1 - 0,103(\bar{s}_1 - \bar{s})$
1	4.88	4.76
2	3.33	3.36
3	4.97	4.84
4	3.07	3.13
5	5.37	5.20
6	1.79	1.98
7	3.29	3.33
8	2.60	2.71
9	3.87	3.85
10	3.58	3.59
11	4.69	4.59
12	3.87	3.85
13	4.33	4.26
14	2.95	3.02
15	3.33	3.36
16	2.37 $\bar{s} = 3.65$	2.50

On remarquera dans ce tableau 2 l'effet correcteur déjà signalé qui rapproche les espérances révisées en effet aléatoire, de \bar{s} .

Pour l'étude de la variance de chaque θ_1 , on obtient d'abord

$$\mu_2' \simeq \frac{15 \times 17}{44 \times 46} \times 1 \times \frac{S_1}{S_2} \simeq 0,045$$

et
$$\mu_2 = \mu_2' - \mu_1'^2 \simeq 0,033$$

d'où la variance
$$\text{Var}(\theta_1/\underline{s}) = a + \mu_2(\bar{s}_1 - \bar{s})^2$$

avec
$$a = \frac{I(S_1 + \mu_1' S_2) - (I - 1)(\mu_1' S_1 + \mu_2' S_2)}{IJ(\nu_1 + \nu_2 - 2)} = 0,097$$

alors qu'en effet fixe, il y a une seule variance intra

$$\text{Var}(\theta_1/\underline{s}) = \frac{S_1}{J(v_1 - 2)} = 0,117$$

d'où le tableau 3.

Tableau 3
Variances révisées intra-entreprises

Entreprise	Effet fixe $\text{Var}(\theta_1/\underline{s}) = \frac{S_1}{J(v_1 - 2)}$	Effet aléatoire $\text{Var}(\theta_1/s) = (\bar{s}_1 - \bar{s})^2 \cdot 0,033 + 0,097$
1	0.117	0.149
2	0.117	0.101
3	0.117	0.154
4	0.117	0.110
5	0.117	0.195
6	0.117	0.207
7	0.117	0.101
8	0.117	0.135
9	0.117	0.099
10	0.117	0.098
11	0.117	0.132
12	0.117	0.099
13	0.117	0.110
14	0.117	0.112
15	0.117	0.101
16	0.117	0.149

Enfin, il existe ici une matrice de covariances Λ ,

$$\text{Cov}(\theta_1 \theta_j / \underline{s}) = \frac{I(S_1 + \mu_1' S_2) - (I - 1)(\mu_1' S_1 + \mu_2' S_2)}{IJ(v_1 + v_2 - 2)} + \mu_2 (\bar{s}_1 - \bar{s})(\bar{s}_j - \bar{s})$$

ou $\text{Cov}(\theta_1 \theta_j / \underline{s}) = 0,097 + 0,033(\bar{s}_1 - \bar{s})(\bar{s}_j - \bar{s})$

il est alors possible de définir une matrice de coefficients de corrélations entre les différentes entreprises par la formule classique

$$\rho_{1j} = \frac{\text{Cov}(\theta_1 \theta_j / \underline{s})}{[\text{Var}(\theta_1 / \underline{s}) \text{Var}(\theta_j / \underline{s})]^{1/2}}$$

Le tableau 4 donne les valeurs de ces matrices pour cet exemple. Il fait apparaître des corrélations très fortes entre certaines entreprises et au contraire les caractères spécifiques des entreprises 5 d'une part, 6 d'autre part.

Tableau 4

MATRICE DES COVARIANCES

0.149	0.084	0.150	0.073	0.166	0.021	0.082	0.054	0.105	0.094	0.139	0.105	0.124	0.068	0.084	0.045
0.084	0.101	0.083	0.103	0.078	0.116	0.100	0.108	0.094	0.097	0.086	0.094	0.089	0.104	0.100	0.110
0.150	0.083	0.154	0.071	0.171	0.015	0.081	0.051	0.106	0.093	0.142	0.106	0.126	0.066	0.083	0.041
0.073	0.103	0.071	0.110	0.064	0.132	0.103	0.117	0.092	0.098	0.077	0.092	0.083	0.110	0.103	0.121
0.166	0.078	0.171	0.064	0.195	-0.008	0.076	0.037	0.109	0.093	0.156	0.109	0.135	0.057	0.078	0.024
0.021	0.116	0.015	0.132	-0.008	0.207	0.119	0.161	0.083	0.101	0.033	0.083	0.055	0.139	0.116	0.175
0.082	0.100	0.081	0.103	0.076	0.119	0.101	0.109	0.094	0.097	0.084	0.094	0.088	0.105	0.100	0.112
0.054	0.108	0.051	0.117	0.037	0.161	0.109	0.135	0.089	0.099	0.060	0.089	0.073	0.121	0.108	0.141
0.105	0.094	0.106	0.092	0.109	0.083	0.094	0.089	0.099	0.096	0.104	0.098	0.101	0.091	0.094	0.087
0.094	0.097	0.093	0.098	0.093	0.101	0.097	0.099	0.096	0.098	0.094	0.096	0.095	0.098	0.097	0.099
0.139	0.086	0.142	0.077	0.156	0.033	0.084	0.060	0.104	0.094	0.132	0.104	0.120	0.072	0.086	0.053
0.105	0.094	0.106	0.092	0.109	0.083	0.094	0.089	0.098	0.096	0.104	0.099	0.101	0.091	0.094	0.087
0.124	0.089	0.126	0.083	0.135	0.055	0.088	0.073	0.101	0.095	0.120	0.101	0.110	0.081	0.089	0.068
0.068	0.104	0.066	0.110	0.057	0.139	0.105	0.121	0.091	0.098	0.072	0.091	0.081	0.112	0.104	0.126
0.084	0.100	0.083	0.103	0.078	0.116	0.100	0.108	0.094	0.097	0.086	0.094	0.089	0.104	0.101	0.110
0.045	0.110	0.041	0.121	0.024	0.175	0.112	0.141	0.087	0.099	0.053	0.087	0.068	0.126	0.110	0.149

MATRICE DES CORRELATIONS

1.00	0.68	0.99	0.57	0.97	0.12	0.67	0.38	0.87	0.77	0.99	0.87	0.97	0.53	0.68	0.30
0.68	1.00	0.66	0.97	0.56	0.80	0.99	0.92	0.94	0.98	0.74	0.94	0.85	0.98	0.99	0.90
0.99	0.66	1.00	0.55	0.99	0.08	0.65	0.35	0.86	0.76	0.99	0.96	0.97	0.50	0.66	0.27
0.57	0.97	0.55	1.00	0.43	0.87	0.98	0.96	0.88	0.94	0.63	0.88	0.76	0.99	0.97	0.94
0.97	0.56	0.99	0.43	1.00	-0.04	0.54	0.23	0.78	0.67	0.97	0.78	0.92	0.38	0.56	0.14
0.12	0.80	0.08	0.87	-0.04	1.00	0.82	0.96	0.58	0.71	0.20	0.58	0.36	0.91	0.80	0.99
0.67	0.99	0.65	0.98	0.54	0.82	1.00	0.93	0.94	0.98	0.73	0.94	0.84	0.99	0.99	0.91
0.38	0.92	0.35	0.96	0.23	0.96	0.93	1.00	0.77	0.86	0.45	0.77	0.60	0.98	0.92	0.99
0.87	0.94	0.86	0.88	0.78	0.58	0.94	0.77	1.00	0.97	0.91	0.99	0.97	0.87	0.94	0.72
0.77	0.98	0.76	0.94	0.67	0.71	0.98	0.86	0.97	1.00	0.83	0.97	0.91	0.94	0.98	0.82
0.99	0.74	0.99	0.63	0.97	0.20	0.73	0.45	0.91	0.83	1.00	0.91	0.99	0.60	0.74	0.37
0.87	0.94	0.86	0.88	0.78	0.58	0.94	0.77	0.99	0.97	0.91	1.00	0.97	0.87	0.94	0.72
0.97	0.85	0.97	0.76	0.92	0.36	0.84	0.60	0.97	0.91	0.99	0.97	1.00	0.73	0.85	0.53
0.53	0.98	0.50	0.99	0.38	0.91	0.99	0.98	0.87	0.94	0.60	0.87	0.73	1.00	0.98	0.97
0.68	0.99	0.66	0.97	0.56	0.81	0.99	0.92	0.94	0.98	0.74	0.94	0.85	0.98	1.00	0.90
0.30	0.90	0.27	0.94	0.14	0.99	0.91	0.99	0.72	0.82	0.37	0.72	0.53	0.97	0.90	1.00

1.8 - EXTENSION DU MODELE

Un travail similaire à celui effectué dans l'exemple ci-dessus aurait consisté en l'examen des indices de croissance annuels de salaires dans 4 métropoles différentes, chaque fois pour 5 qualifications professionnelles différentes, et pour chaque qualification, dans 16 entreprises représentant 16 branches bien définies. Alors que dans le modèle précédent, l'utilisateur posait

comme hypothèse une nette prépondérance de l'effet de la branche, ou de la métropole, il s'agit d'examiner maintenant un modèle permettant de dégager 2 effets prépondérants, par exemple, ici l'effet de la métropole, d'une part, celui de la qualification dans la métropole, d'autre part.

Un tel traitement nécessite l'étude d'un nouveau modèle

$$y_{ijk} = \theta_i + b_{ij} + e_{ijk} \quad \text{avec} \quad \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, \dots, J \\ k = 1, \dots, k \end{array}$$

les y_{ijk} sont les valeurs observées,

les θ_i et b_{ij} sont deux sortes d'effets aléatoires

e_{ijk} est l'"erreur" sur l'observation (i, j, k).

On suppose les θ_i , b_{ij} , e_{ijk} tous indépendants, ce qui est parfois assez restrictif en Econométrie. On fait les hypothèses suivantes

$$\begin{aligned} V_i \theta_i &\sim N(\theta, \sigma_3^2) \\ b_{ij} &\sim N(0, \sigma_2^2) \\ e_{ijk} &\sim N(0, \sigma_1^2) \end{aligned}$$

σ_1^2 , σ_2^2 , σ_3^2 sont encore ici appelés composants en variance. On suppose qu'il y a le même nombre d'observations par groupe et, particularité du modèle à effet aléatoire, les θ_i sont des "observations" indépendamment distribuées d'une même variable aléatoire $N(\theta, \sigma_3^2)$.

Soit le modèle ainsi étudié

$$y_{ijk} = \theta_i + b_{ij} + e_{ijk}$$

il est possible de changer l'indexation des valeurs observées

$$y_{111} ; y_{112} ; \dots ; y_{11,k} ; y_{121} ; y_{122} ; \dots ; y_{1,2,k} \dots$$

qui deviennent

$$y_{11} ; y_{12} \quad y_{1,k} ; y_{1,k} , y_{1,k+2} \quad y_{1,2k} ; \dots$$

soit en général y_{ijk} devient y_{i1} avec $l = k(j - 1) + k$. Comme d'autre part, on a supposé les effet indépendants, on obtient

$$b_{ij} + e_{ijk} = c_{il} \quad \text{avec} \quad c_{il} \sim N(0, \sigma_1^2 + \sigma_2^2)$$

On est ramené au modèle à effet aléatoire déjà étudié précédemment

$$\begin{aligned} y_{i1} &= \theta_i + c_{i1} \quad \text{avec} \quad \theta_i \sim N(\theta, \sigma_3^2) \\ & \quad \quad \quad c_{i1} \sim N(0, \sigma_1^2 + \sigma_2^2) \end{aligned}$$

et l'on pourra de même dégager les espérances et variances révisées des θ_i

$$E(\theta_i/y) = y_i - \mu_1'(\bar{y}_i - \bar{y})$$

avec cette fois ci

$$\bar{y}_i = \sum_j \sum_k y_{ijk}$$

$$\bar{y} = \sum \sum \sum y_{ijk}$$

$$S_2 = JK \sum (\bar{y}_i - \bar{y})^2 \quad \text{et} \quad \nu_2 = I - 1$$

$$S_1 = \sum \sum (y_{i1} - \bar{y}_i)^2 \quad \nu_1 = I(K - 1)$$

et comme dans le premier modèle $\mu'_1 = \frac{\nu_2}{\nu_1 - 2} \frac{I_x \left(\frac{\nu_2}{2} + 1, \frac{\nu_1}{2} - 1 \right)}{I_x \left(\frac{\nu_2}{2}, \frac{\nu_1}{2} \right)} \frac{S_1}{S_2}$

on calculera de même variances et covariances sans autre difficulté, ce qui permet d'envisager l'étude d'effets aléatoires sur des modèles de plus en plus hiérarchisés

$$y_{ijk1\dots} = \theta_1 + b_{ij} + c_{ijk} + e_{ijk1\dots}$$

Deuxième Partie

ETUDE DE LA VARIANCE INTER-TRAITEMENTS

2.1 - LE MODELE ET L'ANALYSE BAYESIENNE

Dans la première partie, l'étude de l'effet des traitements, espérance révisée et variance intra, a été faite sur un modèle de la forme

$$y_{ij} = \theta_1 + e_{ij} \quad \text{avec} \quad \theta_1 \sim N(\theta, \sigma_2^2) \quad \text{et} \quad e_{ij} \sim N(0, \sigma_1^2).$$

Ici, c'est l'étude de la variance inter-traitements qui sera poursuivie, mais sur un modèle similaire

$$y_{ij} = \mu + a_i + e_{ij}$$

avec μ paramètre de localisation commun, a_i effet du traitement i avec $a_i \sim N(0, \sigma_2^2)$ et $e_{ij} \sim N(0, \sigma_1^2)$. On supposera, d'une part $\theta = \mu$ ce qui rend les deux modèles similaires et d'autre part les a_i et e_{ij} normalement et indépendamment distribués.

Si la distribution a priori de $(\mu, \sigma_1^2, \sigma_2^2)$ est non informative, on obtient

$$p(\mu, \sigma_2^2, \sigma_1^2) \sim \frac{1}{\sigma_1^2 \sigma_{12}^2} \quad \text{avec} \quad \sigma_{12}^2 = \sigma_1^2 + J\sigma_2^2$$

La fonction de vraisemblance jointe se note en analyse Bayesienne :

$$L(\mu, \sigma_2^2, \sigma_1^2 / \underline{y}) \sim (\sigma_1^2)^{-\frac{I(J-1)}{2}} (\sigma_{12}^2)^{-\frac{I}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_{12}^2} + \frac{IJ(\bar{y} - \mu)^2}{\sigma_{12}^2} \right] \right\}$$

avec les mêmes notations d'analyse classique que dans la première partie, notamment pour S_1 et S_2 .

On peut alors calculer la distribution a posteriori

$$p''(\mu, \sigma_2^2, \sigma_1^2 / \underline{y}) = p'(\mu, \sigma_2^2, \sigma_1^2) \times L(\mu, \sigma_2^2, \sigma_1^2 / \underline{y})$$

soit

$$p(\mu, \sigma_2^2, \sigma_1^2 / \underline{y}) \sim (\sigma_1^2)^{-\frac{I(J-1)+1}{2}} (\sigma_{12}^2)^{-\frac{I+J}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_{12}^2} + \frac{IJ(\bar{y} - \mu)^2}{\sigma_{12}^2} \right] \right\}$$

Si l'on intègre cette expression par rapport à μ , il vient

$$p(\sigma_1^2 \sigma_2^2 / \underline{y}) \sim C (\sigma_1^2)^{-\frac{I(J-1)+1}{2}} (\sigma_{12}^2)^{-\frac{I+1}{2}} \exp \left[-\frac{1}{2} \left(\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_2^2} \right) \right] \quad (7'')$$

où C est la constante normalisante appropriée

$$C = \frac{J(S_1)^{\frac{I(J-1)}{2}} S_2^{\frac{I-1}{2}} \frac{I-1}{2} \frac{IJ-1}{2}}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right) I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)}$$

avec $\Phi = \frac{S_2}{S_1}$ et I_x mesure Beta incomplète pour $x = \frac{\Phi}{1 + \Phi}$

En raison de la définition de $\sigma_{12}^2 = \sigma_1^2 + J\sigma_2^2$, il existe une contrainte $\sigma_{12}^2 > \sigma_1^2$ dont il faut tenir compte.

2.2 - LA DISTRIBUTION A POSTERIORI DE LA VARIANCE INTER-TRAITEMENTS

Opérons la transformation $r_1 = \frac{S_1}{\sigma_1^2}$ et $u = \frac{2J\sigma_2^2}{S_2}$

où σ_2^2 et u sont des variables aléatoires ; l'expression (7'') devient

$$p(r, u / \underline{y}) = \frac{2^{-\frac{IJ-1}{2}} r^{\frac{I(J-1)}{2}} \frac{\Phi r}{\left(1 + \frac{1}{2} \Phi ru\right)^{\frac{I+1}{2}}}}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right) I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \exp \left\{ -\frac{1}{2} \left[r + \frac{\Phi r}{1 + \frac{1}{2} \Phi ru} \right] \right\} \quad (8'')$$

on remarque en effet que $\frac{S_2}{\sigma_{12}^2} = \frac{\Phi r}{1 + \frac{1}{2} \Phi ru}$.

Si l'on effectue alors un nouveau changement de variable

$$\Psi = \frac{1}{2} \frac{\Phi r}{1 + \frac{1}{2} \Phi ru} \quad \beta = \frac{I-1}{2} \quad 0 < \Psi < \infty$$

avec $\varphi = \frac{r}{2} \quad \alpha = \frac{I(J-1)}{2} \quad 0 < \varphi < \infty$

il vient

$$(r, u/y) = \frac{1}{\Gamma(\alpha)} \varphi^{\alpha-1} e^{-\varphi} \frac{1}{\Gamma(\beta)} \Psi^{\beta-1} e^{-\Psi} \frac{1}{I_x(\alpha, \beta)} \quad (9)$$

forme plus fréquente d'une distribution χ^2 .

La distribution a posteriori de u est obtenue par intégration de l'expression (8) par rapport à r

$$p(u/y) = \int_0^\infty p(u, r/y) dr = \frac{\int_0^\infty k(u, z) z^{\frac{I(J-1)}{2}} e^{-z} dz}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right) I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)}$$

avec $z = \frac{r}{2} = \frac{S_1}{2\sigma_1^2}$ et $k = \left(u + \frac{1}{\Phi z}\right)^{-\frac{I+1}{2}} \exp\left\{-\frac{1}{u + \frac{1}{\Phi z}}\right\}$

Si la distribution a priori initialement supposée non-informative est appropriée, la distribution $p(u/y)$ rassemble toute notre connaissance a posteriori de u donc de σ_2^2 ; comme la distribution est définie sur $(0, \infty)$, il n'y a aucun problème de "variance négative".

Par intégration directe pour $\frac{I-1}{2} > r$, le moment d'ordre p de u est

$$\mu_p = \frac{1}{I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \sum_{i=0}^p \binom{p}{i} (-1)^i \Phi^{-i} I_x\left(\frac{I-1}{2} - p + i, \frac{I(J-1)}{2} - i\right) \frac{\Gamma\left(\frac{I-1}{2} - p + i\right) \Gamma\left(\frac{I(J-1)}{2} - i\right)}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right)}$$

Tiao, collaborant avec Tan [2], a surtout recherché des propriétés asymptotiques de la distribution de σ_2^2 négligeant le calcul de l'espérance et

de la variance révisées de σ_2^2 . Or si ce calcul peut paraître fastidieux, il se simplifie jusqu'à offrir une comparaison intéressante avec le mode de la distribution de (r, u) .

En effet

$$\mu_1' = \frac{1}{H_\Phi\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \left[I_x\left(\frac{I-1}{2} - 1, \frac{I(J-1)}{2}\right) \frac{\Gamma\left(\frac{I-1}{2} - 1\right) \Gamma\left(\frac{I(J-1)}{2}\right)}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right)} \right. \\ \left. - \frac{1}{\Phi} I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2} - 1\right) \frac{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2} - 1\right)}{\Gamma\left(\frac{I-1}{2}\right) \Gamma\left(\frac{I(J-1)}{2}\right)} \right]$$

soit après simplification, en remarquant que $\Gamma(x+1)/\Gamma(x) = x$

on obtient

$$\mu_1' = \frac{2}{H_\Phi\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \left[\frac{1}{I-3} I_x\left(\frac{I-3}{2}, \frac{I(J-1)}{2}\right) \right. \\ \left. - \frac{1}{\Phi} \frac{1}{I(J-1)-2} I_x\left(\frac{I-1}{2}, \frac{I(J-1)-2}{2}\right) \right]$$

et comme $\mu_1' = E(u/y) = E\left(\frac{2J\sigma_2^2}{S_2} / y\right)$

$$E(\sigma_2^2/y) = \frac{1}{J I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \left[\frac{S_2}{I-3} I_x\left(\frac{I-3}{2}, \frac{I(J-1)}{2}\right) \right. \\ \left. - \frac{S_1}{I(J-1)-2} I_x\left(\frac{I-1}{2}, \frac{I(J-1)-2}{2}\right) \right]$$

De même le calcul de μ_2' nous conduit à

$$\mu_2'(\sigma_2^2/y) = \frac{1}{J^2 I_x\left(\frac{I-1}{2}, \frac{I(J-1)}{2}\right)} \left[\frac{S_2^2 I_x\left(\frac{I-5}{2}, \frac{I(J-1)}{2}\right)}{(I-3)(I-5)} \frac{2S_1 S_2 I_x\left(\frac{I-3}{2}, \frac{I(J-1)-2}{2}\right)}{(I-3)(I(J-1)-2)} \right. \\ \left. + \frac{S_1^2 I_x\left(\frac{I-1}{2}, \frac{I(J-1)-4}{2}\right)}{(I(J-1)-4)(I(J-1)-2)} \right]$$

ce qui permet de calculer la variance révisée par la relation

$$\text{Var}(\sigma_2^2/y) = \mu_2'(\sigma_2^2) - [\mu_1'(\sigma_2^2)]^2$$

2.3 - APPLICATION NUMERIQUE

Avant de poursuivre l'exploitation de l'exemple de la première partie, quelques importantes remarques originales s'imposent ici aussi.

Si l'on considère le mode de la distribution a posteriori de (σ_1^2, σ_2^2) , c'est-à-dire de (r, u) , on obtient par différentiation de la distribution jointe (9)

$$\frac{\partial P}{\partial \varphi} = \frac{1}{I_x(\alpha, \beta)} \frac{1}{\Gamma(\alpha)} [(\alpha - 1) \varphi^{\alpha-2} e^{-\varphi} - \varphi^{\alpha-1} e^{-\varphi}] \frac{1}{\Gamma(\beta)} \Psi^{\beta-1} e^{-\beta}$$

Pour un extremum $\frac{\partial P}{\partial \Phi} = 0 \implies (\alpha - 1) \varphi^{\alpha-2} = \varphi^{\alpha-1}$

soit $(\alpha - 1) \varphi^{-1} = 1$

d'où $\varphi_0 = \alpha - 1 = \frac{r_0}{2}$ d'où $r_0 = 2(\alpha - 1) = I(J - 1) - 2$.

De même

$$\frac{\partial P}{\partial \Psi} = \frac{1}{I_x(\alpha, \beta)} \frac{1}{\Gamma(\alpha)} \varphi^{\alpha-1} e^{-\varphi} \frac{1}{\Gamma(\beta)} [(\beta - 1) \Psi^{\beta-2} - \Psi^{\beta-1} e^{-\beta}]$$

Pour un extremum $\frac{\partial P}{\partial \Psi_0} = 0 \implies (\beta - 1) \Psi^{\beta-2} = \Psi^{\beta-1} \quad (\beta - 1) \Psi^{-1} = 1$

d'où $\Psi_0 = \beta - 1 = \frac{1}{2} \frac{\Phi r_0}{1 + \frac{\Phi r_0 u}{2}}$

soit $u_0 = \frac{2}{I - 3} - \frac{2}{\Phi [I(J - 1) - 2]}$

les conditions du deuxième ordre permettent de constater qu'il s'agit bien d'un maximum.

D'autre part, la condition $u \geq 0$ entraîne $\Phi \geq \frac{I - 3}{I(J - 1)}$. Si cette dernière condition n'est pas satisfaite, on peut alors faire $u_0 = 0$, ce qui donne $r_0 = \frac{IJ - 1}{1 + \Phi}$. La condition $u_0 < 0$ équivaudrait en effet à $\sigma_2^2 < 0$; la distribution (8) n'est donc définie que pour $0 < u < \infty$.

Comme $u = \frac{2J\sigma_2^2}{S_2}$, on obtient pour $(\sigma_2^2)_0$

$$(\sigma_2^2)_0 = \frac{1}{J} \left(\frac{S_2}{I - 3} - \frac{S_1}{I(J - 1) - 2} \right)^+$$

avec la convention $a^+ = \max(a, 0)$.

Comparons cette expression avec l'espérance révisée de σ_2^2

$$E(\sigma_2^2/y) = \frac{1}{J} \left(K_1 \frac{S_2}{I-3} - K_2 \frac{S_1}{I(J-1)-2} \right)$$

$$\text{avec } K_1 = \frac{I_x \left(\frac{I-3}{2}, \frac{I(J-1)}{2} \right)}{I_x \left(\frac{I-1}{2}, \frac{I(J-1)}{2} \right)} \quad \text{et} \quad K_2 = \frac{I_x \left(\frac{I-1}{2}, \frac{I(J-1)-2}{2} \right)}{I_x \left(\frac{I-1}{2}, \frac{I(J-1)}{2} \right)}$$

K_1 et K_2 sont deux expressions de formes respectives

$\frac{I_x(\alpha-1, \beta)}{I_x(\alpha, \beta)}$ et $\frac{I_x(\alpha, \beta-1)}{I_x(\alpha, \beta)}$. Plus α et β sont grands, c'est-à-dire plus I et J sont grands, plus K_1 et K_2 sont voisins de 1 ; le mode de la distribution est alors une bonne approximation de l'espérance révisée de σ_2^2 .

Ces résultats sont à rapprocher de ceux de l'analyse classique où l'unique estimateur non biaisé pour σ_2^2 , basé sur (y, S_1, S_2) est

$$\hat{\sigma}_2^2 = \frac{1}{J} \left(\frac{S_2}{I-1} - \frac{S_1}{I(J-1)} \right)^+$$

Il faut une fois de plus signaler qu'un tel estimateur peut prendre une valeur numérique négative avec une probabilité positive ! un autre estimateur très utilisé et qui n'est jamais négatif est celui obtenu par la méthode du Maximum de vraisemblance

$$\hat{\sigma}_2^2 = \frac{1}{J} \left(\frac{S_2}{I} - \frac{S_1}{I(J-1)} \right)^+$$

Si l'on considère l'exemple de la première partie concernant la croissance annuelle de salaires dans différentes entreprises et métiers, on obtient l'estimation suivante de la variance inter-entreprises (tableau 4).

Tableau 4

Estimation de la variance inter-entreprises
dans l'exemple de la 1ère partie

Analyse	Méthode	Expression	Valeur numérique
Classique	Estimateur non biaisé	$\frac{1}{J} \left(\frac{S_2}{I-1} - \frac{S_1}{I(J-1)} \right)$	0,89
	Maximum de vraisemblance	$\frac{1}{J} \left(\frac{S_2}{I} - \frac{S_1}{I(J-1)} \right)$	0,83
Bayésienne	Approximation par le mode de l'à posteriori	$\frac{1}{J} \left(\frac{S_2}{I-3} - \frac{S_1}{I(J-1)-2} \right)^+$	1,04
	Espérance révisée	$\frac{1}{J} \left(\frac{K_1 S_2}{I-3} - \frac{K_2 S_1}{I(J-1)-2} \right)$ défini pour $0 < u < \infty$	1,04 avec $K_1 \approx 1 \text{ à } 10^{-2}$ près $K_2 \approx 1 \text{ à } 10^{-2}$ près

Ayant μ_1' et μ_2' , on pourra calculer la variance révisée de σ_2^2 ; il faut noter qu'en analyse classique, on a comme "erreur" sur l'estimateur non biaisé

$$\frac{1}{J} \left[\frac{2\sigma_2^2}{I-1} + \frac{2\sigma_2^4}{I(J-1)} \right]$$

ce qui nécessite d'ailleurs, comme pour le calcul de l'"erreur" dans la méthode du Maximum de Vraisemblance, le calcul préalable d'une estimation de σ_1^2 . Ici μ_1' et μ_2' suffisent à établir une variance révisée du composant σ_2^2

Enfin, on pourra même dégager une variance inter-traitements dans un modèle à 3 effets, par exemple villes, qualifications professionnelles et types d'entreprises. Soit le modèle similaire à celui de la première partie

$$y_{ijk} = \mu + a_i + b_{ij} + e_{ijk}$$

avec $a_i \sim N(0, \sigma_3^2)$ $b_{ij} \sim N(0, \sigma_2^2)$ $e_{ijk} \sim N(0, \sigma_1^2)$

en posant $l = K(j-1) + k$, il vient le modèle

$$y_{il} = \mu + a_i + e_{il}$$

avec $a_i \sim N(0, \sigma_3^2)$ $e_{il} \sim N(0, \sigma_1^2 + \sigma_2^2)$

déjà étudié dans cette deuxième partie pour l'estimation d'une variance inter-traitements σ_3^2 ; dans l'exemple cité il faut noter qu'on est ainsi amené à privilégier par hypothèse un effet -celui des villes ou celui des entreprises- étudié alors en détail par les espérances révisées, les variances intra et la variance inter-traitements.

En conclusion, les travaux de Tiao et Box [1], Tiao et Tan [2], par l'étude de quelques propriétés, ont ouvert la voie à l'utilisation des plans d'expérience Bayésiens à effet aléatoire ; ici, a été définie toute une méthodologie qui permet, dans un plan à 2 ou plusieurs dimensions, pour les différents traitements, le calcul d'espérances révisées, de variances intra-traitements, alors qu'il y en a une seule dans le modèle à effet fixe. La seconde partie de ce travail vise, de plus, dans un tel modèle, à un calcul d'une variance inter-traitements. L'ensemble de la méthode est un nouvel outil statistique à la disposition de l'Econométrie.

BIBLIOGRAPHIE

- [1] BOX et TIAO - Bayesian estimation of means for the random effect model - J.A.S.A. - p. 174 - 181 ; Mars 1968.
- [2] TIAO et TAN - Bayesian analysis of random effect models - Biometrika, n° 52, 37-53, 1965.
- [3] JEFFREYS (1961) - Theory of Probability - Oxford - Clarendon Press, 1961.

- [4] RAIFFA et SCHLAIFER - Applied statistical decision theory - Harvard, 1961.
- [5] DAUBIGNEY - Croissance des salaires et évolution de la hiérarchie ouvrière - Cahiers de l'I. S. E. A. - Economie et Société, tome III, n° 12, Genève. 1969.
- [6] DeGROTT - "Optimal Statistical Decisions" Mac-Graw Hill, 1970.
- [7] DUNET et SOBEL - Quelques propriétés de la Distribution de Student Biometrika p. 153-169 - 1954.
- [8] MARCIANO - Analyse Bayesienne de plans d'expériences à effet aléatoire - Thèse de Doctorat d'Etat, Toulouse, 1971.