LES CAHIERS DE L'ANALYSE DES DONNÉES

J.-P. BENZÉCRI
C. BOURGARIT
J.-C. MADRE

Problème : ajustement d'un tableau à ses marges d'après la formule de reconstitution

Les cahiers de l'analyse des données, tome 5, n° 2 (1980), p. 163-172
<http://www.numdam.org/item?id=CAD_1980__5_2_163_0>

Les Cahiers de l'Analyse des Données
Vol. V - 1980 - n° 2 - p. 163-172

PROBLÈME : AJUSTEMENT D'UN TABLEAU A SES MARGES D'APRÈS LA FORMULE DE RECONSTITUTION [AJUS. MARGES FAC.]

par J.-P. Benzécri (1),
C. Bourgarit (2)
et J.-C. Madre (3).

0 Il arrive communément dans la pratique que pour une loi de correspondance g_{IJ} on ait d'une part, une estimation approchée f_{IJ} de l'ensemble du tableau ; et d'autre part, une estimation très précise des lois marginales g_I, g_J. On doit alors reconstituer le tableau g_{IJ} sous les deux contraintes d'avoir pour marges g_I et g_J ; et de s'écarter le moins possible de f_{IJ} : c'est ce qu'on appelle *ajuster le tableau* f_{IJ}, *à des marges* nouvelles g_I , g_J. L'objet du présent problème est d'étudier une procédure d'ajustement particulière, fondée sur l'analyse de la correspondance f_{IJ}.

Pour nous la procédure d'ajustement présente encore un autre intérêt. Soit f_{IJ} et g_{IJ} deux lois de correspondance. Afin de comparer g_{IJ} à f_{IJ} il est commun, d'adjoindre les lignes et colonnes de g_{IJ} en éléments supplémentaires à f_{IJ}: il en résulte des facteurs modifiés :

$$F^g_\alpha(i) = F_\alpha(g^i_J) \quad ; \quad G^g_\alpha(j) = G_\alpha(g^j_I).$$

Et une question se pose : les différences interprétables observées entre F^g_α et F_α (ou G^g_α et G_α) se retrouvent-elles lorsqu'on remplace g_{IJ} par un tableau g'_{IJ} construit par ajustement de f_{IJ} aux marges g_I et g_J de g_{IJ} ? Si tel est le cas on dira que la différence entre f_{IJ} et g_{IJ} est principalement un effet de marge. On sait d'ailleurs, cf Yagolnitzer (*in Cahiers* Vol II n° 3 pp 251-264) que l'analyse d'un tableau en quatre blocs fournit d'une part les facteurs de la correspondance moyenne $(f+g)/2$; et d'autre part des facteurs issus de la différence $(f-g)/2$: et parmi ceux-ci les deux premiers, et souvent les seuls interprétables, sont communément des facteurs de marges (liés directement à la différence des lois marginales de f et g sur I et J). Dans la partie 3 du présent problème, on calcule l'expression exacte de F^g_α dans le cas d'un modèle continu simple ; la confrontation avec des données réelles sera faite ailleurs.

f	g
g	f

1 Enoncé du problème *

1.1 On suppose donnée une loi de correspondance f_{IJ} (i.e. un système de nombres positifs ou nuls f_{ij} dont la somme est 1) ; et deux lois g_I et g_J. On note λ_α , $F_\alpha(i)$, $G_\alpha(j)$ les valeurs propres et facteurs

(1) *Professeur de statistique. Université P. et M. Curie*
(*) *Problème donné aux étudiants du D.E.A.. Partiel de Mars 1980.*
(2) *Assistante. Université P. et M. Curie.*
(3) *Attaché de Recherches - CNRS - CREDOC. 140, rue du Chevaleret.*

issus de l'analyse de f_{IJ} ; pour tout profil s_I sur I (i.e. système de nombres positifs s_i de somme 1) on note $G_\alpha(s_I)$ la valeur du facteur G_α, pour s_I considéré comme une colonne supplémentaire adjointe au tableau f_{IJ} ; et on définit de même $F_\alpha(s_J)$ pour un profil sur J considéré comme ligne supplémentaire.

1.1.1 Calculer en fonction des g_i, g_j, λ_α, $F_\alpha(i)$, $G_\alpha(j)$ les valeurs des facteurs $F_\alpha(g_J)$ et $G_\alpha(g_I)$ (pour g_J et g_I considérés comme éléments supplémentaires).

1.1.2 On considère l'expression :

$$g'_{ij} = g_i g_j (1+\Sigma_\alpha\{\lambda_\alpha^{-1/2}(F_\alpha(i) - A_\alpha)(G_\alpha(j) - B_\alpha)\}) \;;$$

(où la somme Σ_α est étendue soit à tous les facteurs non triviaux issus de f_{IJ} ; soit seulement aux p premiers de ces facteurs ; p étant choisi librement).

Déterminer A_α, B_α afin que le tableau g'_{IJ} ait pour lois marginales g_I et g_J.

1.2 Dans la 2° partie, on conserve les hypothèses de la 1° partie et on suppose que :

$$\forall\, j \in J \;:\; g_j = f_j(1 + b\,\varphi_1(j))$$

où b est une constante réelle ; et $\varphi_1 = \lambda_1^{-1/2}\, G_1$ est le facteur normalisé de variance 1.

1.2.1 Exprimer les nombres B_α (cf 1.1.2) en fonction de b et des valeurs propres λ_α.

1.2.2 On considère le profil g'^{i}_J d'une ligne du tableau g'_{IJ} construit au § 1.1.2 (avec les marges exactes g_I, g_J) :

$$g'^{i}_J = \{g'^{i}_j \mid j \in J\} \;;\quad g'^{i}_j = g'_{ij}/g'_i = g'_{ij}/g_i \;;$$

Donner la valeur $F_\alpha(g'^{i}_J)$ du facteur F_α pour le profil g'^{i}_J (adjoint en ligne supplémentaire à f_{IJ} ; cf 1.1.1). On exprimera cette valeur d'abord en fonction des g'^{i}_j, λ_α et $G_\alpha(j)$; puis en fonction des coefficients A_β et B_β (pour β quelconque) et de l'ensemble des valeurs propres et facteurs issus de l'analyse du tableau f_{IJ} ; puis donner cette valeur en remplaçant les B_β par leur valeur (en fonction de b) trouvée en 1.2.1.

N.B. Dans cette question, les relations d'orthogonalité entre facteurs jouent un rôle essentiel.

1.2.3 Les facteurs normalisés de variance 1 étant désignés par $\varphi_\alpha(j)$, on note :

$$T_{\alpha\beta\gamma} = \Sigma\{\varphi_\alpha(j)\;\varphi_\beta(j)\;\varphi_\gamma(j)\;f_j \mid j \in J\} \;;$$

Montrer que les facteurs $F_\alpha(g'^i_J)$, calculés en 1.2.2, s'expriment sous la forme :

$$F_\alpha(g'^i_J) = C_{\alpha 0} + \sum_\beta C_{\alpha\beta} F_\beta(i)$$

avec des coefficients C dont le calcul est très simple en fonction des A_α, B_α, λ_α et des $T_{\alpha\beta\gamma}$. On donnera également les coefficients C en fonction de b, et des A_α, λ_α et $T_{\alpha\beta\gamma}$.

1.2.4 Généraliser la formule trouvée en 1.2.3 au cas où on ne suppose plus que g_j a la forme particulière $f_j(1+b\varphi_1(j))$ mais une forme générale $g_j = f_j(1+\sum_\alpha b_\alpha \varphi_\alpha(j))$.

1.3 L'objet de la 3° partie est d'appliquer les formules trouvées au § 1.2.3, à une correspondance continue symétrique simple sur le carré $(-1, 1) \times (-1,1)$. On note donc :

$$X = Y = (-1, 1) \text{ (intervalle } -1, +1) ;$$

$$f(x,y) = (1/4)(1 + (1/2)\, xy + (x^2 - (1/3))\,(y^2 - (1/3))) .$$

$$g(x) = (1+x)/2 \; ; \; g(y) = (1+y)/2 \; ;$$

$$f_{XY} = f(x,y)\, dx\, dy \; ; \quad g_X = g(x)\, dx \; ; \; g_Y = g(y)\, dy .$$

1.3.1 Déterminer les lois marginales $f_X = f(x)dx$ et $f_Y = f(y)dy$ de la loi de correspondance f_{XY}.

1.3.2 Déterminer une constante k_1 telle que la fonction $\varphi_1(x) = k_1 x$ ait variance 1 (sur X muni de la loi f_X).

1.3.3 Déterminer une constante k_2 telle que la fonction $\varphi_2(x) = k_2(x^2 - (1/3))$ ait variance 1 (sur X muni de la loi f_X).

1.3.4 Quels sont les facteurs 1 et 2 issus de la correspondance f_{XY} (on donnera λ_1, $F_1(x)$, $G_1(y)$ et λ_2, $F_2(x)$, $G_2(y)$).

1.3.5 Construire la correspondance g'_{XY} associée à f_{XY} et aux marges g_X, g_Y suivant le procédé expliqué au § 1.1.2.

1.3.6 Utiliser les résultats de 1.2.3 pour calculer les deux fonctions :

$$F_1(g'^x_Y) = F^g_1(x) \quad ; \quad \text{et} \quad : \quad F_2(g'^x_Y) = F^g_2(x) \quad ;$$

et représenter dans le plan des axes 1×2 issu de l'analyse de f_{XY} les deux courbes d'équation paramétrique :

$$E = \{(F_1(x), F_2(x)) \,|\, x \in X\} \; ; \; E' = \{(F^g_1(x), F^g_2(x)) \,|\, x \in X\} .$$

2 *Solution du problème*

2.1 *Formule d'ajustement d'un tableau à ses marges*

2.1.1 Valeur des facteurs pour les profils des nouvelles marges : La formule de transition donne :

$$F_\alpha(g_J) = \lambda_\alpha^{-1/2}\ \Sigma\{g_j\ G_\alpha(j)\,|\,j\in J\}\ ;$$
$$G_\alpha(g_I) = \lambda_\alpha^{-1/2}\ \Sigma\{g_i\ F_\alpha(i)\,|\,i\in I\}\ .$$

2.1.2 Calcul des constantes dans la formule d'ajustement : Etudier la condition pour que g'_{IJ} ait pour loi marginale $g'_I = g_I$. On a :

$$g'_{ij} = g_i g_j(1+\Sigma_\alpha\{\lambda_\alpha^{-1/2}(F_\alpha(i)-A_\alpha)(G_\alpha(j)-B_\alpha)\})\ ;\ \text{d'où :}$$

$$g'_i = \Sigma_j g'_{ij} = g_i(1+\Sigma_\alpha\{\lambda_\alpha^{-1/2}(F_\alpha(i)-A_\alpha)(\overline{G}_\alpha-B_\alpha)\})\ ;\ \text{où on a}$$

noté :

$$\overline{G}_\alpha = \Sigma\ \{g_j\ G_\alpha(j)\,|\,j\in J\} = \lambda_\alpha^{1/2}F_\alpha(g_J)\ ;$$

il est évident que l'égalité $g_I = g'_I$ est vérifiée si on a $B_\alpha = \overline{G}_\alpha$, pour chacun des p facteurs retenus dans la formule de reconstitution; on peut de plus montrer (sous réserve que les g_i soient tous non-nuls) que cette condition est également nécessaire, i.e. que $g'_I = g_I$ implique : $B_\alpha = \overline{G}_\alpha$. Voici la démonstration. Notons :

$$\Delta_i = (g'_i/g_i)-1 = \Sigma_\alpha\{\lambda_\alpha^{-1/2}(F_\alpha(i)-A_\alpha)(\overline{G}_\alpha-B_\alpha)\}\ ;\ \text{on a :}$$

$$(g'_I = g_I) \Leftrightarrow (\forall\, i\in I : \Delta_i = 0)\ ;$$

de la nullité de tous les Δ_i résulte la nullité de la somme :

$$\begin{aligned}S_\beta &= \Sigma\{f_i\ F_\beta(i)\ \Delta_i\,|\,i\in I\}\\ &= \Sigma_\alpha\{\lambda_\alpha^{-1/2}(\overline{G}_\alpha-B_\alpha)\Sigma\{f_iF_\beta(i)(F_\alpha(i)-A_\alpha)|\ i\in I\}\}\\ &= \Sigma_\alpha\{\lambda_\alpha^{-1/2}(\overline{G}_\alpha-B_\alpha)\ \lambda_\alpha\ \delta_{\alpha\beta}\} = \lambda_\beta^{1/2}\ (\overline{G}_\beta-B_\beta)\ ;\end{aligned}$$

d'où finalement la nullité de $(\overline{G}_\beta - B_\beta)$ pour tout β.

Le raisonnement fait pour la marge sur I valant pour la marge sur J nous conclurons :

$$A_\alpha = \Sigma\{g_i\ F_\alpha(i)\,|\,i\in I\} = \lambda_\alpha^{1/2}G_\alpha(g_I)\ ;$$

$$B_\alpha = \Sigma\{g_j\ G_\alpha(j)\,|\,j\in J\} = \lambda_\alpha^{1/2}F_\alpha(g_J)\ .$$

2.2 *Adjonction des lignes du tableau reconstitué en éléments supplémentaires au tableau de base*

2.2.1 Calcul des B_α dans un cas particulier : $g_j = f_j(1+b\,\lambda_1^{-1/2}G_1(j))$; on a :

$$B_\alpha = \Sigma_j\ \{g_j\ G_\alpha(j)\} = \Sigma_j\{f_j\ G_\alpha(j)\} + b\lambda_1^{-1/2}\Sigma_j\{f_j\ G_1(j)G_\alpha(j)\}\ ;$$

or les facteurs G_α sont de moyenne nulle, de variance λ_α, et deux à deux non corrélés ; d'où en notant comme d'usage, $\delta_{\alpha\beta}= 1$ si $\alpha=\beta$ et 0 sinon) :

$B_\alpha = b\,\lambda_1^{-1/2}\,\lambda_\alpha\,\delta_{\alpha 1}$; i.e. $B_1 = b\,\lambda_1^{1/2}$; et les autres B_α sont nuls.

2.2.2 *<u>Adjonction des lignes supplémentaires dans le même cas particulier</u>* : On a :

$$F_\alpha(g'^{i}_{J}) = \lambda_\alpha^{-1/2}\,\Sigma\{g'^{i}_{j}\,G_\alpha(j) \mid j \in J\} \;;\; \text{où } g'^{i}_{j} = g'_{ij}/g_i \;;$$

en remplaçant g'_{ij} par l'expression donnée pour lui dans l'énoncé, il vient :

$$\lambda_\alpha^{1/2}\,F_\alpha(g'^{i}_{J}) = \Sigma_j\{g_j\,G_\alpha(j)\,(1 + \Sigma_\beta\{\lambda_\beta^{-1/2}\,(F_\beta(i) - A_\beta)\,(G_\beta(j) - B_\beta)\})\} \;;$$

et, en tenant compte des expressions pour les A et les B trouvées au § 2.1.2 :

$$\lambda_\alpha^{1/2}\,F_\alpha(g'^{i}_{J}) = B_\alpha + \Sigma_\beta\{\lambda_\beta^{-1/2}\,(F_\beta(i) - A_\beta)\,(\Sigma_j\{g_j\,G_\alpha(j)G_\beta(j)\} - B_\alpha B_\beta)\} \;;$$

ici apparaît la moyenne du produit de deux facteurs *relativement à la loi* g_J ; le calcul de cette moyenne est facile, en fonction de l'expression donnée ici de la densité de g_J par rapport à f_J ; il vient :

$$\Sigma_j\{g_j\,G_\alpha(j)G_\beta(j)\} = \Sigma_j\{f_j\,G_\alpha(j)\,G_\beta(j)\} + b\,\Sigma_j\{f_j\,\varphi_1(j)G_\alpha(j)G_\beta(j)\}$$

$$= \lambda_\alpha\,\delta_{\alpha\beta} + b\lambda_1^{-1/2}\,\Sigma_j\{f_j\,G_1(j)G_\alpha(j)G_\beta(j)\} \;;\; \text{d'où :}$$

$$F_\alpha(g'^{i}_{J}) = \lambda_\alpha^{-1/2}\,B_\alpha + \Sigma_\beta\{(\lambda_\alpha\,\lambda_\beta)^{-1/2}\,(F_\beta(i) - A_\beta)\,(\lambda_\alpha\delta_{\alpha\beta} + b\,\lambda_1^{-1/2}\Sigma_j\{f_j\,G_1G_\alpha G_\beta\} - B_\alpha B_\beta)\};$$

il vaut la peine de préciser cette expression, en tenant compte des valeurs des B_α trouvées au § 2.2.1 ; et en distinguant le cas $\alpha = 1$; on a :

$$F_1(g'^{i}_{J}) = b + (F_1(i) - A_1)\,(1 - b^2)$$

$$+ \lambda_1^{-1}\,\Sigma_\beta\{\lambda_\beta^{-1/2}\,(F_\beta(i) - A_\beta)\,b\,\Sigma_j\{f_j\,G_1^2\,G_\beta\}\} \;;$$

$$F_\alpha(g'^{i}_{J}) = (F_\alpha(i) - A_\alpha)$$

$$+ \lambda_1^{-1/2}\,\lambda_\alpha^{-1/2}\,\Sigma_\beta\{\lambda_\beta^{-1/2}(F_\beta(i) - A_\beta)\,b\,\Sigma_j\{f_j\,G_1\,G_\alpha\,G_\beta\}\} \,,$$

où dans la dernière formule, α désigne un indice *différent* de 1.

N. B. La formule construite ici s'apparente à la formule de reconstitution des données en fonction des facteurs, mais en diffère en ce que les λ_α , $F_\alpha - A_\alpha$, $G_\alpha - B_\alpha$ ne sont pas en général les véritables v.p. et facteurs que donnerait l'analyse du tableau g'_{IJ} ; on s'en souviendra dans la suite en calculant les facteurs F^g_α , G^g_α , relatifs au tableau g' adjoint en supplément au tableau f .

2.2.3 *<u>Les facteurs calculés sur les lignes supplémentaires sont des combinaisons linéaires des facteurs calculés sur les lignes principales</u>* : On reprend les formules du § 2.2.2, en remplaçant les sommes $\Sigma_j\{f_j\,G_\alpha\,G_\beta\,G_\gamma\}$ par $(\lambda_\alpha\,\lambda_\beta\,\lambda_\gamma)^{1/2}\,T_{\alpha\beta\gamma}$; il vient :

$$F_\alpha(g'^{i}_{J}) = \lambda_\alpha^{-1/2}\ B_\alpha + (F_\alpha(i) - A_\alpha)$$

$$+ \Sigma_\beta\{(F_\beta(i) - A_\beta)(bT_{1\alpha\beta} - (\lambda_\alpha\lambda_\beta)^{-1/2} B_\alpha B_\beta)\}\ ;\ \text{soit:}$$

$$F_\alpha(g'^{i}_{J}) = C_{\alpha 0} + \Sigma_\beta\{C_{\alpha\beta}\ F_\beta(i)\}\ ;\ \text{avec :}$$

$$C_{\alpha 0} = \lambda_\alpha^{-1/2}\ B_\alpha - A_\alpha - \Sigma_\beta\{A_\beta(b\ T_{1\alpha\beta} - (\lambda_\alpha\lambda_\beta)^{-1/2} B_\alpha B_\beta)\}\ ;$$

$$C_{\alpha\beta} = \delta_{\alpha\beta} + b\ T_{1\alpha\beta} - (\lambda_\alpha\lambda_\beta)^{-1/2} B_\alpha B_\beta\ ;$$

ici encore, les valeurs particulières des B nous suggèrent de distinguer le cas $\alpha = 1$; on a :

$$C_{10} = b - A_1 + A_1\ b^2 - b\ \Sigma_\beta\{A_\beta T_{11\beta}\}\ ;$$

$$C_{\alpha 0} = -A_\alpha - b\ \Sigma_\beta\{A_\beta\ T_{1\alpha\beta}\}\ ;\ \text{où } \alpha \neq 1\ ;$$

$$C_{11} = 1 + b\ T_{111} - b^2\ ;$$

$$C_{1\beta} = b\ T_{11\beta}\ ;\ \text{où } \beta \neq 1\ ;$$

$$C_{\alpha\beta} = \delta_{\alpha\beta} + b\ T_{1\alpha\beta}\ ;\ \text{où } \alpha \neq 1\ ;$$

2.2.4 *Calcul du cas général où* : $g_j = f_j\ (1 + \Sigma_\alpha\{b_\alpha\ \varphi_\alpha(j)\})$. Revenons à l'expression des facteurs $F(g'^{i}_{J})$ trouvée au début du § 2.2.2, avant toute hypothèse sur la forme de g_J :

$$\lambda_\alpha^{1/2} F_\alpha(g'^{i}_{J}) = B_\alpha + \Sigma_\beta\{\lambda_\beta^{-1/2}(F_\beta(i) - A_\beta)(\Sigma_j\{g_j\ G_\alpha(j)\ G_\beta(j)\} - B_\alpha B_\beta)\}\ ;$$

il nous faut calculer en fonction des b_α (et des $T_{\alpha\beta\gamma}$), d'une part les B_α, d'autre part les moyennes $\Sigma_j\{g_j\ G_\alpha(j)\ G_\beta(j)\}$. On a :

$$B_\alpha = \Sigma\{g_j\ G_\alpha(j) \mid j \in J\}$$

$$= \Sigma\{f_j\ (1 + \Sigma_\beta\{b_\beta\ \varphi_\beta(j)\})\ G_\alpha(j) \mid j \in J\} = \lambda_\alpha^{1/2} b_\alpha$$

(où on a tenu compte de l'orthogonalité des facteurs).

$$M_{\alpha\beta} = \Sigma_j\{g_j\ G_\alpha(j)\ G_\beta(j)\}$$

$$= \Sigma_j\{f_j\ (1 + \Sigma_\gamma\{b_\gamma\varphi_\gamma(j)\})\ G_\alpha(j)\ G_\beta(j)\}$$

$$= \lambda_\alpha\ \delta_{\alpha\beta} + \lambda_\alpha^{1/2}\ \lambda_\beta^{1/2}\ \ \Sigma_\gamma\{b_\gamma\ T_{\alpha\beta\gamma}\}.$$

En résumant la valeur du facteur F_α pour une ligne supplémentaire, il vient :

$$F_\alpha(g'^{i}_{J}) = b_\alpha + \Sigma_\beta\{(F_\beta(i) - A_\beta)(\delta_{\alpha\beta} - b_\alpha b_\beta + \Sigma_\gamma\{b_\gamma\ T_{\alpha\beta\gamma}\})\}.$$

$$F_\alpha(g'^{i}_{J}) = C_{\alpha 0} + \Sigma_\beta\ C_{\alpha\beta}\ F_\beta(i)\ ;\ \text{où on a noté :}$$

$$C_{\alpha 0} = b_\alpha - A_\alpha + \Sigma_\beta\{A_\beta\ (b_\alpha b_\beta - \Sigma_\gamma\{b_\gamma\ T_{\alpha\beta\gamma}\})\}\ ;$$

$$C_{\alpha\beta} = \delta_{\alpha\beta} - b_\alpha\ b_\beta + \Sigma_\gamma\{b_\gamma\ T_{\alpha\beta\gamma}\}.$$

2.3 Etude d'un modèle continu : Notre but est d'apprécier qualitativement l'écart entre le nuage $N(I)$ des profils f^i_J issus du tableau principal f_{IJ}, et le nuage $N'(I)$ des profils g'^i_J issus du tableau ajusté g'_{IJ}, et adjoints en éléments supplémentaires à l'analyse principale. Pour une telle comparaison le mieux est de considérer un modèle continu, aussi simple que possible, mais présentant toutefois des effets intéressants. Nous choisissons d'étudier une correspondance symétrique, rentrant dans le cas particulier du début du § 2.2 : ainsi on aura $A_1 = B_1$ et les autres A_α, B_α nuls. Mais pour que l'écart entre $N(I)$ et $N'(I)$ soit plus frappant, on désire qu'il n'y ait pas un simple changement d'échelle sur les facteurs, mais un mélange de ceux-ci: autrement dit que les $C_{\alpha\beta}$ extradiagonaux ($\alpha \neq \beta$) ne soient pas tous nuls ! Plus précisément, cf § 2.2.3, on a :

$$C_{12} = C_{21} = b\, T_{112} \quad ;$$

il faut donc que T_{112} soit non-nul. Or T_{112} qui est la moyenne de $(\varphi_1)^2\,\varphi_2$, sera non-nul si φ_2 est corrélé à $(\varphi_1)^2$: on réalisera au mieux cela, si φ_2 est égal (à une constante additive près) à $(\varphi_1)^2$; c'est-à-dire dans le cas de l'effet Guttman. Compte-tenu de ce qui précède, le modèle qu'on va étudier nous paraît être aussi simple que possible !

2.3.1 Les lois marginales de la correspondance de base f_{XY} : Avant tout, vérifions que la densité proposée $f(x,y)$ est bien positive sur le carré XY : en effet, le minimum du produit xy sur le carré est -1. Et le minimum du produit $(x^2-(1/3))\,(y^2-(1/3))$ est atteint quand l'un des termes est maximum positif (il vaut alors 2/3) et l'autre minimum négatif (il vaut alors -1/3) : le minimum de ce produit est donc -2/9. Finalement, le min. de $f(x,y)$ est supérieur à $(1/4)(1-(1/2)-(2/9))$; donc positif.

Le calcul de la densité marginale $f(x)$ est rendu très simple par le fait que la fonction $(y^2-(1/3))$ a été choisie pour avoir moyenne nulle sur l'intervalle $Y = (-1, +1)$. On a :

$$f(x) = \int_Y f(x,y)\,dy = \frac{1}{4}\int_Y (1 + (xy/2) + (x^2 - (1/3))(y^2 - (1/3)))\,dy = 1/2 \;;$$

$$f_X = f(x)\,dx = (1/2)dx \;;\quad f_Y = f(y)dy = (1/2)dy.$$

2.3.2 Normalisation de la fonction linéaire :

$\varphi_1(x) = k_1\,x$; φ_1 a moyenne nulle sur $X = (-1, 1)$;

$\mathrm{var}(\varphi_1) = \int_X (\varphi_1(x))^2\, f(x)\,dx = k_1^2/3$; on posera donc :

$k_1 = 3^{1/2}$; $\varphi_1(x) = 3^{1/2}\,x$.

2.3.3 Normalisation de la fonction quadratique

$\varphi_2(x) = k_2(x^2 - (1/3))$; φ_2 a moyenne nulle sur X ;

$\mathrm{var}(\varphi_2) = (k_2)^2 \int_X (x^2-(1/3))^2\,(dx/2) = (k_2)^2(4/45)$; on posera :

$$k_2 = (3/2)5^{1/2} \ ; \ \varphi_2(x) = (3/2)5^{1/2}(x^2-(1/3)) \ .$$

2.3.4 *Analyse du tableau* f_{XY} : Il suffit de mettre la loi donnée sous la forme suivante :

$$f(x,y) = f(x)\,f(y)\,(1 + \Sigma_\alpha\{\lambda_\alpha^{1/2}\ \varphi_\alpha(x)\ \varphi_\alpha(y)\}),$$

où les fonctions φ_α sont des fonctions de moyenne nulle et variance 1 deux à deux non corrélées. Ici le modèle est conçu pour qu'il n'y ait que deux facteurs, qui sont les fonctions φ_1 et φ_2 qu'on vient de normaliser. Il faut seulement vérifier que φ_1 et φ_2 sont orthogonales entre elles (ce qui est évident puisque φ_1 est une fonction *impaire* : $\varphi_1(-x) = -\varphi_1(x)$; tandis que φ_2 est *paire* : $\varphi_2(x) = \varphi_2(-x)$; d'où $\int_{-1}^{1} \varphi_1\ \varphi_2\ dx = 0$) ; et calculer les λ_1, λ_2 pour rejoindre la densité $f(x,y)$ donnée. Il suffit de poser

$$\lambda_1^{1/2} = (1/2)(1/k_1^2) = 1/6 \ ; \ \lambda_1 = (1/36) = 0{,}028 \ ;$$

$$\lambda_2^{1/2} = (1/k_2^2) = (4/45) \ ; \ \lambda_2 = (4/45)^2 = 0{,}008 \ ;$$

$$f(x,y) = (1/4)\,(1 + (1/6)\,(3^{1/2}x)\,(3^{1/2}y) + (4/45)\,((3/2)5^{1/2}\,(x^2-(1/3))\ (3/2)5^{1/2}(y^2-(1/3))))$$

On a pour les facteurs F, G :

$$F_1(x) = \lambda_1^{1/2}\ \varphi_1 = 3^{-1/2}\,2^{-1}\,x \ ; \ G_1(y) = 3^{-1/2}\,2^{-1}\,y$$

$$F_2(x) = \lambda_2^{1/2}\ \varphi_2 = (2/3)5^{-1/2}(x^2-(1/3)) \ ; \ G_2(y) = (2/3)5^{-1/2}(y^2-(1/3)) \ .$$

On notera que l'on a bien $\lambda_2 < \lambda_1$; comme le suggère notre numérotage.

2.3.5 *Ajustement de la correspondance aux nouvelles marges* : On a posé:

$$g(x) = (1+x)/2 = f(x)\,(1 + 3^{-1/2}\,\varphi_1(x)) \ ;$$

$$g(y) = (1+y)/2 = f(y)\,(1 + 3^{-1/2}\,\varphi_1(y)) \ ;$$

on rentre donc dans le cas du § 1.2.1 avec $b = 3^{-1/2}$. D'où :

$$B_1 = A_1 = b\,\lambda_1^{1/2} = (1/6)\,3^{-1/2} \ ; \ B_2 = A_2 = 0$$

et pour la densité du tableau ajusté :

$$g'(x,y) = (1/4)\,(1+x)\,(1+y)\,(1 + (1/2)\,(x-(1/3))(y-(1/3)) + (x^2-(1/3))\,(y^2-(1/3))) \ ;$$

il est facile de vérifier comme au § 2.3.1 qu'il s'agit bien d'une densité positive sur le carré XY .

2.3.6 *Adjonction du tableau ajusté en supplément au tableau de base* : Pour appliquer les résultats du § 2.2.3, il nous suffit de calculer les $T_{\alpha\beta\gamma}$: plus exactement ceux utiles ici ; c'est-à-dire T_{111}, T_{112} et T_{122}.

On a :

$$T_{111} = \int_X f(x)\,(\varphi_1(x))^3\,dx = ((k_1)^3/2)\int_X x^3\,dx = 0$$

$$T_{122} = \int_X f(x)\,(\varphi_2(x))^2\,\varphi_1(x)\,dx$$

$$= (k_1(k_2)^2/2)\int_X (x^2-(1/3))^2 x\,dx = 0\ ;$$

en effet, ici encore il s'agit de l'intégrale sur (-1, +1) d'une fonction *impaire*; reste à calculer T_{112}

$$T_{112} = \int_X f(x)\,(\varphi_1(x))^2\,\varphi_2(x)\,dx$$

$$= \int_X k_1^2\ k_2\ x^2(x^2-(1/3))\,(dx/2)$$

$$= \int_X k_1^2\ k_2(x^2-(1/3))\ (x^2-(1/3))\ (dx/2)$$

$$= (k_1^2/k_2) = 2(5^{-1/2})\ ;$$

dans ce calcul on a tenu compte de ce que φ_2 a moyenne nulle sur X pour remplacer x^2 par $(x^2-(1/3))$; et utiliser ensuite le fait que φ_2 a variance 1. Ceci posé on a (cf § 2.2.3)

$$C_{10} = b - A_1 + A_1\,b^2 - b\,\Sigma_\beta\{A_\beta\ T_{11\beta}\}$$

$$= b - A_1 + A_1\,b^2 = 3^{-1/2}\,(1-(1/6)+(1/18))$$

$$= 3^{-1/2}\,(8/9) = 0{,}51$$

$$C_{20} = -A_2 - b\,\Sigma_\beta\{A_\beta\ T_{12\beta}\} = -\,b\,A_1\,T_{121}$$

$$= -5^{-1/2}/9 = -0{,}05\ ;$$

$$C_{11} = 1 + b\,T_{111} - b^2 = 1 - b^2 = (2/3)\ ;$$

$$C_{22} = 1 + b\,T_{122} = 1\ ;$$

$$C_{12} = C_{21} = b\,T_{112} = 2(15^{-1/2}) = 0{,}516.$$

D'où pour les facteurs des éléments supplémentaires :

$$F^g_1(x) = C_{10} + C_{11}\,F_1(x) + C_{12}\,F_2(x)$$
$$= 0{,}46 + 0{,}192\,x + 0{,}154\,x^2\ ;$$

$$F^g_2(x) = C_{20} + C_{21}\,F_1(x) + C_{22}\,F_2(x)$$
$$= 0{,}15\ (-1 + x + 2x^2)\,.$$

Reste à représenter les deux arcs E et E'.

$$E = \{(3^{-1/2}.\ 2^{-1}x\ ;(2/3)5^{-1/2}\ (x^2-(1/3))\ |\ x \in (-1,\ 1)\}\ .$$

Il s'agit d'un arc de parabole, symétrique par rapport au deuxième axe ; c'est l'effet Guttman sous sa forme la plus pure.

$$E' = \{F^g_1(x) \ ; \ F^g_2(x) \mid x \in (-1, 1)\} \ ;$$

puisque F^g_1 et F^g_2 sont tous deux des trinômes du deuxième degré en le paramètre x, il s'agit ici encore d'un arc de parabole : mais cet arc n'est pas aussi régulièrement disposé que E : il y a de E à E' une importante déformation où le coefficient C_{12} lié à T_{112} (donc on l'a dit a l'effet Guttman) joue le rôle principal.

Le tableau ci-dessous qui donne les variations de $F^g_1(x)$ et $F^g_2(x)$ avec x permet de tracer la courbe E' qui est donnée, ainsi que E sur la figure ci-après :

x	$-\infty$		-1		-0.62		-0.25		0		0.5		1		$+\infty$
$F'^g_1(x)$		$-$		$-$	0	$+$		$+$		$+$		$+$		$+$	
$F'^g_2(x)$		$-$		$-$		$-$	0	$+$		$+$		$+$		$+$	
$F_1{}^g(x)$			0,424	↘	0,40	↗	0,42	↗	0,46	↗	0,59	↗	0,81		
$F_2{}^g(x)$			0	↘	-0,13	↘	-0,17	↗	-0,15	↗	0	↗	0,30