REVUE DE STATISTIQUE APPLIQUÉE

K. M. M. DORKENOO
J.-R. MATHIEU

Étude d'un modèle factoriel d'analyse de la variance comme modèle linéaire généralisé

Revue de statistique appliquée, tome 41, nº 2 (1993), p. 43-57
<http://www.numdam.org/item?id=RSA_1993__41_2_43_0>

Rev. Statistique Appliquée, 1993, **XLI** (2), 43-57

ETUDE D'UN MODÈLE FACTORIEL D'ANALYSE DE LA VARIANCE COMME MODÈLE LINÉAIRE GÉNÉRALISÉ

K.M.M. Dorkenoo, J.-R. Mathieu

Laboratoire de Statistique et Probabilités
U.R.A. C.N.R.S. D0745, Université Paul Sabatier
31062 Toulouse Cedex, France

RÉSUMÉ

On considère le modèle gaussien (M_d) d'analyse de la variance avec interaction modélisée par d termes multiplicatifs. On montre que ce modèle peut être considéré comme un modèle linéaire généralisé et on étudie les propriétés asymptotiques des estimateurs du maximum de vraisemblance d'une part lorsque le nombre de répétitions tend vers l'infini et d'autre part lorsque la variance des résidus tend vers zéro. Ces résultats sont ensuite appliqués à un problème de choix de modèle dans (M_d).

Mots-clés : *Analyse de la variance, Interaction multiplicative, Modèle linéaire généralisé, Distribution asymptotique, Choix de modèles.*

SUMMARY

The ANOVA model with d multiplicative terms for the interaction effect (M_d) is considered. It is shown that this model can be considered as a generalized linear model; then the asymptotic properties of the maximum likelihood estimators are given firstly when the number of replications goes to infinity and secondly when the variance function of the observations tends to zero. The results are applied to solve a problem of model choice in (M_d).

Key-words : *ANOVA, Multiplicative Interaction, Generalized Linear Model, Asymptotic Distribution, Model Choice.*

1. Introduction

L'objet de ce travail est l'étude du modèle d'analyse de la variance à deux facteurs avec interactions modélisées par d termes multiplicatifs. Si on note y_{ijk} l'observation de la variable aléatoire Y_{ijk} relative à la $k^{ème}$ répétition faite pour le niveau i du premier facteur et le niveau j du second, le modèle étudié s'écrit :

$$(M_d) \begin{cases} Y_{ijk} = \mu + a_i + b_j + \sum_{s=1}^{d} \lambda_s \alpha_{i,s} \beta_{j,s} + \varepsilon_{ijk} \\ i = 1, \dots, p;\ j = 1, \dots, q \leqslant p;\ k = 1, \dots, r; \\ \varepsilon_{ijk} \sim N(0, \sigma^2) \\ \text{avec des contraintes d'identification sur les paramètres.} \end{cases}$$

Ce modèle d'analyse de la variance a connu différentes approches quant à l'étude des interactions. Turkey [27] propose la modélisation des interactions comme produit des effets différentiels a_i par b_j puis Williams [28] le modèle (M_1). Depuis, ce dernier a fait l'objet de plusieurs travaux dont ceux de Mandel [20, 21, 22], Gollob [13], Johnson [16, 17, 15], Corsten [8], Denis [9], Shuurmann [26], Gabriel [6, 12], Krishnaiah [18], Robert [25], Boik [4, 5] sur l'estimation et sur les tests d'hypothèses.

L'intérêt d'un tel modèle est, entre autres, l'interprétation aisée des interactions et la représentation des niveaux de facteurs par des techniques d'analyse de données.

Malgré une bibliographie enrichie par l'analyse multivariée, principalement l'Analyse en Composantes Principales, ce n'est que récemment et parallèlement à notre travail que des résultats ont été publiés sur la distribution asymptotique des paramètres multiplicatifs de (M_1) par Chadœuf et Denis [7] puis Goodman et Haberman [14].

Pour déterminer la distribution asymptotique des paramètres multiplicatifs de (M_1), nous avons appliqué les résultats de Aitchison et Silvey [1] sur l'estimation sous contraintes dans le modèle linéaire généralisé (MLG) (cf. Nelder et Wedderburn [24] puis Mathieu [23]) modèle dont nous montrons que $M(d)$ est un cas particulier.

La distribution asymptotique des estimateurs lorsque le nombre r des répétitions tend vers l'infini permet d'approcher la distribution des estimateurs lorsque r est grand. Mais nous nous intéressons aussi à approcher la distribution des estimateurs dans le cas où $r = 1$. Une théorie applicable dans ce cas consiste à étudier les propriétés asymptotiques des estimateurs maximum de vraisemblance (MV) lorsque le paramètre d'échelle du MLG tend vers zéro, ce qui correspond à considerer le cas où σ^2 tend vers zéro dans le modèle (M_d).

Enfin, compte tenu du fait que le modèle étudié peut comporter plusieurs termes multiplicatifs pour modéliser l'interaction, nous nous sommes intéressés au problème du choix optimal (M_d) en vue de la prédiction (cf. Allen [2, 3] et Mallows [19]). Des exemples sont traités pour illustrer les solutions apportées à ce problème en comparaison avec les tests d'hypothèses proposés dans la littérature.

2. Modèle linéaire généralisé

2.1. *Définition du modèle*

Dans le modèle linéaire généralisé (MLG) défini par Nelder et Wedderburn [24], la moyenne de la $i^{ème}$ observation est une fonction numérique monotone d'un prédicteur linéaire qui dépend des variables explicatives et d'un paramètre inconnu θ de R^p. On définit alors la fonction de lien comme l'inverse de cette fonction numérique. Le MLG ainsi défini regroupe le cas d'observations issues de $v.a$ gaussienne, binomiale, de Poisson, gamma, gaussienne inverse, etc...

Dans une extension du MLG (cf. Mathieu [23]), on suppose que l'espérance mathématique des observations est fonction de plusieurs prédicteurs linéaires. La formulation la plus générale du MLG pour les observations réelles est la suivante :

Soit $y = (y_1, \ldots, y_n)$ un échantillon tel que

– y_i soit la réalisation d'une $v.a$ réelle Y_i et ces n $v.a$ soient indépendantes;

– il existe sur $\mathbf{R}$ une mesure par rapport à laquelle Y_i possède une densité $f_i(., \theta, \xi)$ non nulle sur une partie U de $\mathbf{R}$, indépendante des paramètres inconnus θ et ξ.

– $f_i(., \theta, \xi)$ est définie par

$$\forall u \in U \subset \mathbf{R}, \quad f_i(u, \theta, \xi) = \exp\{\xi^{-1}[\gamma_i u - \mu(\gamma_i) - \eta_i(u)] + \psi(u, \xi)\}$$

avec $\gamma_i = \phi(t_i\theta)$ où :

– θ est inconnu et appartient à $\mathbf{R}^p$;

– ξ est inconnu et appartient à $\mathbf{R}_+^* = \mathbf{R}_+ - \{0\}$;

– t_i est une application linéaire connue de $\mathbf{R}^p$ dans $\mathbf{R}^s$;

– ϕ est une application numérique connue définie sur $\mathbf{R}^s$;

– μ, ψ et η_i sont des fonctions numériques connues.

Cette formulation du MLG peut se résumer par le schéma suivant :

$$\theta \in \mathbf{R}^p \xrightarrow{t_i} t_i\theta \in \mathbf{R}^s \xrightarrow{\phi} \gamma_i \in \mathbf{R} \xrightarrow{\mu} \mu(\gamma_i) \in \mathbf{R}.$$

Le logarithme de la vraisemblance de y s'écrit :

$$L_n(y, \theta, \xi) = \xi^{-1} \sum_{i=1}^{n} [y_i\gamma_i - \mu(\gamma_i) - \eta_i(y_i)] + \sum_{i=1}^{n} \psi(y_i, \xi).$$

Sous des conditions de régularités sur μ et ϕ, on montre (cf. [24]) la

Proposition 2.1

$$\begin{aligned}&\forall(\theta,\xi)\in\mathbf{R}^p\times\mathbf{R}_+^*\\&\qquad\mathbf{E}_{\theta,\xi}(Y_i)=\mu'(\gamma_i)=(\mu'\circ\phi)(t_i\theta)\\&\qquad\mathbf{Var}_{\theta,\xi}(Y_i)=\xi(\mu''\circ\phi)(t_i\theta)\end{aligned}\qquad\square$$

Remarque 1. *Le paramètre ξ n'affectant pas la moyenne des Y_i mais seulement sa variance est appelé paramètre d'échelle du MLG.*

Remarque 2. *Le MLG de Nelder et Wedderburn correspond au cas où le prédicteur linéaire $(t_i\theta)$ est un scalaire; la fonction de lien est alors définie par $(\mu'\circ\phi)^{-1}$. Par contre, dans le modèle ci-dessus, il devient impossible de parler de fonction de lien.*

Pour montrer que (M_d) est un MLG, nous proposons le schéma suivant où θ désigne l'ensemble des paramètres du modèle *i.e.* :

$$\theta\in\mathbf{R}^{(d+1)(p+q+1)}\xrightarrow{t_{ijk}}\begin{pmatrix}\mu\\a_i\\b_j\\\lambda_1\\\alpha_{i,1}\\\beta_{j,1}\\\vdots\\\lambda_d\\\alpha_{i,d}\\\beta_{j,d}\end{pmatrix}\in\mathbf{R}^{3(d+1)}\xrightarrow{\phi}\gamma_{ijk}\in\mathbf{R}\xrightarrow{\mu}\frac{1}{2}(\gamma_{ijk})^2\in\mathbf{R}$$

avec $\gamma_{ijk}=\mu+a_i+b_j+\sum_{s=1}^{d}\lambda_s\alpha_{i,s}\beta_{j,s}$.

Nous aborderons dans la suite le modèle (M_d) à travers l'étude du modèle à un terme multiplicatif (M_1). Dans l'immédiat, tout en restant dans le MLG, nous étudions le comportement des estimateurs MV de θ lorsque le paramètre d'échelle ξ tend vers zéro.

2.2. Distribution de l'estimateur MV de θ lorsque le paramètre d'échelle ξ tend vers zéro

2.2.1 Notations et hypothèses

On considère les notations suivantes :

– Soit $\overline{\theta}$ la vraie valeur du paramètre θ.

– On note Q_n la mesure de probabilité de densité $\prod_{i=1}^{n}f_i(.,\overline{\theta},\xi)$.

– On notera $\mathcal{L}(T_n/Q_n)$ la distribution de la $v.a$ T_n image de la loi Q_n.

– On note $\phi_i = \phi \circ t_i$, $\mu_i = \mu \circ \phi_i$, $\Gamma_i(\theta) = \langle \mu''(\gamma_i) \circ \phi_i'(\theta), \phi_i'(\theta) \rangle$, $\phi'(\theta)$ désignant le vecteur des dérivées de ϕ par rapport aux composantes de θ, $\mu''(\gamma_i) = \mu''(\phi_i\theta)$ la matrice des dérivées secondes de $\mu(\phi_i\theta)$ par rapport aux composantes de θ et $\langle , \rangle$ le produit scalaire usuel; on pose alors :

$$\Gamma(\theta) = \sum_{i=1}^{n} \Gamma_i(\theta).$$

– On note avec une barre toute fonction prise en $\overline{\theta}$ par exemple $\overline{\Gamma} = \Gamma(\overline{\theta})$.

– On pose

$$\xi = \frac{1}{m} \quad \text{avec} \quad m \in \mathbf{N}^*$$

et on étudie le comportement de l'estimateur MV de θ lorsque m tend vers l'infini. Pour cela on considère le MLG défini dans (2.1) où $y = (y_1, \ldots, y_n)$ désigne un échantillon de taille n fixée. Le logarithme de la vraisemblance de y s'écrit :

$$L_n(y, \theta, m) = m \sum_{i=1}^{n} [y_i \gamma_i - \mu(\gamma_i) - \eta_i(y_i)] + \sum_{i=1}^{n} \psi(y_i, m).$$

Enfin, on suppose vérifiées les hypothèses (H) suivantes :

(H_1) : Les dérivées successives d'ordre trois de $L_n(y, \theta, m)$ existent et sont suffisamment petites en norme dans un voisinage de $\overline{\theta}$ défini par :

$$B(\overline{\theta}, \rho) = \{\theta \in \mathbf{R}^p; \|\theta - \overline{\theta}\| \leqslant \rho\}.$$

(H_2) : Pour tout ρ positif, toutes les dérivées jusqu'au troisième ordre de μ et ϕ sont bornées sur $B(\overline{\theta}, \rho)$.

2.2.2 Propriétés de l'estimateur MV $\widehat{\theta}_m$ de $\overline{\theta}$

Pour caractériser la loi limite de $[\sqrt{m}(\widehat{\theta}_m - \overline{\theta})/Q_n]$, on considère la $v.a$ $m^{-1/2} L_n'(Y, \theta, m)$ de $\mathbf{R}^p$ où L_n' est le vecteur des dérivées de L_n par rapport aux composantes de θ et on montre (cf. Dorkenoo [10]), les résultats suivants :

Proposition 2.2 *Sous les hypothèses (H), l'estimateur MV $\widehat{\theta}_m$ de $\overline{\theta}$ est tel que*

$$\mathcal{L}[\sqrt{m}(\widehat{\theta}_m - \overline{\theta})/Q_n] \xrightarrow{m \to \infty} \mathcal{L}(L)$$

où L est la distribution limite de $\overline{\Gamma}^{-1}[m^{-1/2}L'_n(y,\theta,m)]$ *avec en particulier*

$$\begin{cases} \mathbf{E}_{\overline{\theta},m}(L) = 0 \\ \mathbf{Var}_{\overline{\theta},m}(L) = \overline{\Gamma}^{-1} \end{cases} \qquad \square$$

La loi de L n'est pas nécessairement gaussienne puisqu'on n'a plus la condition de Liapounov du théorème central limite. Cependant, lorsqu'il s'agit d'observations issues de $v.a$ gaussienne, gaussienne inverse ou gamma, on vérifie que L est gaussienne.

3. Etude du modèle (M_d)

Nous abordons maintenant l'étude du modèle d'analyse de la variance avec interaction multiplicative à travers le modèle à un terme (M_1) et nous verrons dans quelle mesure les résultats du paragraphe précédent peuvent être utilisés.

3.1 Loi limite des paramètres multiplicatifs de (M_1)

Le modèle étudié est le suivant :

$$(M_1)\begin{cases} Y_{ijk} = \mu + a_i + b_j + \lambda\alpha_i\beta_j + \varepsilon_{ijk} \\ i = 1,\dots,p;\ j = 1,\dots,q \leqslant p;\ k = 1,\dots,r; \\ \sum_i a_i = \sum_i \alpha_i = \sum_j b_j = \sum_j \beta_j = 0 \\ \sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1;\ \lambda > 0; \\ \varepsilon_{ijk} \sim N(0,\sigma^2). \end{cases}$$

D'après Gabriel [12], l'estimation des paramètres multiplicatifs λ, (α_i), (β_j) peut se faire après celle de μ, (a_i) et (b_j). Ce qui revient à maximiser avec les notations classiques $z_{ij} = \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}$,

$$L_n(z,\theta,\sigma)^2 = -\frac{r}{2\sigma^2}\sum_{ij}(z_{ij} - \lambda\alpha_i\beta_j)^2.$$

Soit E_p la matrice carrée d'ordre p dont les éléments sont égaux à $(1/p)$ et soit I_p la matrice identité du même ordre. Si on note $l = (l_1,l_2,l_3,l_4)$ les multiplicateurs de Lagrange de l'estimation de $\theta = (\lambda,\alpha_i,\beta_j)$ sous les contraintes de (M_1) et $\widehat{\theta}_r$ l'estimateur MV de θ, on montre (cf. [10]) la

Proposition 3.1

$$\begin{bmatrix} \sqrt{r}(\widehat{\theta}_r - \overline{\theta}) \\ \frac{1}{\sqrt{r}}\,\widehat{l} \end{bmatrix} \xrightarrow{r\to\infty} N\left[\begin{pmatrix}0\\0\end{pmatrix};\sigma^2\begin{pmatrix}\Delta & 0\\ 0 & R\end{pmatrix}\right]$$

avec

$$\Delta = \frac{1}{\lambda^2}\begin{bmatrix}\lambda^2 & 0 & 0\\ 0 & I_p - E_p - \alpha\alpha' & 0\\ 0 & 0 & I_q - E_q - \beta\beta'\end{bmatrix} \quad et\ R = \begin{pmatrix}\frac{\lambda^2}{p} & 0 & 0 & 0\\ 0 & \frac{\lambda^2}{q} & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

□

D'après les résultats sur le MLG lorsque le paramètre d'échelle ξ tend vers zéro, on déduit dans le modèle (M_1) avec une observation par cellule $(r = 1)$, la loi limite des paramètres multiplicatifs lorsque σ^2 tend vers zéro. D'où le résultat suivant :

Corollaire 3.2

$$\sigma^{-1}(\widehat{\theta} - \overline{\theta}) \xrightarrow{\sigma\to 0} N(0,\Delta)$$

□

Remarque 3. *Des résultats similaires sont proposés dans [7], et [14] par différentes autres approches.*

L'extension au modèle (M_d), avec $\theta = (\lambda_1, \alpha_{i,1}, \beta_{j,1}, \ldots, \lambda_d, \alpha_{i,d}, \beta_{j,d})$, est donnée par le résultat suivant :

Proposition 3.3

$$\sqrt{r}(\widehat{\theta}_r - \overline{\theta}) \xrightarrow{r\to\infty} N\left[0;\sigma^2\begin{pmatrix}\Delta_{11} & \cdots & \Delta_{1,d}\\ \vdots & \ddots & \vdots\\ \Delta_{d,1} & \cdots & \Delta_{d,d}\end{pmatrix}\right]$$

avec

$$\Delta_{i,i} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\lambda_i^2}[I_p - E_p - \alpha_i\alpha_i'] & 2\lambda_i \sum_{k\neq i} \frac{\lambda_k\alpha_k\beta_k'}{(\lambda_k^2 - \lambda_i^2)^2} \\ & + \sum_{k\neq i} \frac{\lambda_k^2(3\lambda_i^2 - \lambda_k^2)\alpha_k\alpha_k'}{\lambda_i^2(\lambda_k^2 - \lambda_i^2)^2} & \\ 0 & 2\lambda_i \sum_{k\neq i} \frac{\lambda_k\beta_k\alpha_k'}{(\lambda_k^2 - \lambda_i^2)^2} & \frac{1}{\lambda_i^2}[I_q - E_q - \beta_i\beta_i'] \\ & & + \sum_{k\neq i} \frac{\lambda_k^2(3\lambda_i^2 - \lambda_k^2)\beta_k\beta_k'}{\lambda_i^2(\lambda_k^2 - \lambda_i^2)^2} \end{pmatrix}$$

et pour tout $i \neq j$,

$$\Delta_{i,j} = \Delta_{j,i}' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{-(\lambda_i^2 + \lambda_j^2)\alpha_j\alpha_i'}{(\lambda_i^2 - \lambda_j^2)^2} & \frac{-2\lambda_i\lambda_j\alpha_j\beta_i'}{(\lambda_i^2 - \lambda_j^2)^2} \\ 0 & \frac{-2\lambda_i\lambda_j\beta_j\alpha_i'}{(\lambda_i^2 - \lambda_j^2)^2} & \frac{-(\lambda_i^2 + \lambda_j^2)\beta_j\beta_i'}{(\lambda_i^2 - \lambda_j^2)^2} \end{pmatrix} \qquad \square$$

Nous allons utiliser ces résultats pour résoudre un problème de choix de modèle en analyse de la variance à deux facteurs avec interaction.

3.2 Choix du modèle (M_d) en vue de la prédiction

Dans le modèle d'analyse de la variance à deux facteurs avec interaction

$$(M) \begin{cases} Y_{ijk} = \mu + a_i + b_j + \eta_{ij} + \varepsilon_{ijk} \\ \varepsilon_{ijk} \sim N(0, \sigma^2) \end{cases}$$

on recherche le modèle $(M_d)_{d\leqslant(q-1)}$ qui prédit le mieux au sens de l'erreur quadratique moyenne de prédiction $(EQMP)$ les pq $v.a$:

$$Y_{ij}^0 = \mu + a_i + b_j + \eta_{ij} + \varepsilon_{ij}^0$$

indépendantes des Y_{ijk}, les ε_{ij}^0 étant centrées, indépendantes et de même variance σ^2/r. Si on note P_{ij}^d, le prédicteur de Y_{ij}^0 associé à (M_d), considéré comme sous-modèle de (M), on a :

$$P_{ij}^d = \widehat{\mu} + \widehat{a}_i + \widehat{b}_j + \sum_{s=1}^{d} \widehat{\lambda}_s\widehat{\alpha}_{i,s}\widehat{\beta}_{j,s}.$$

Avec les notations

$$\begin{cases} \widehat{\eta}_{ij}^{d} = \sum_{s=1}^{d} \widehat{\lambda}_s \widehat{\alpha}_{i,s} \widehat{\beta}_{j,s} & \eta_{ij}^{d} = \sum_{s=1}^{d} \lambda_s \alpha_{i,s} \beta_{j,s} \\ EQMP_{ij}^{d} = \mathbf{E}(P_{ij}^{d} - Y_{ij}^{0})^2 & EQMP_d = \sum_{ij} EQMP_{ij}^{d} \end{cases}$$

on montre que

$$EQMP_d = \frac{\sigma^2}{r}(pq + p + q - 1) + \sum_{ij} \mathrm{Var}(\widehat{\eta}_{ij}^{d}) + \sum_{ij} (\eta^{ij} - \eta_{ij})^2.$$

Lorsque r tend vers l'infini, on a $\eta_{ij}^{d} \simeq \mathbf{E}(\sum_{s=1}^{d} \widehat{\lambda}_s \widehat{\alpha}_{i,s} \widehat{\beta}_{j,s})$ et donc

$$EQPM_d = [pq + (d+1)(p+q-d-1)]\frac{\sigma^2}{r} + \sum_{s=d+1}^{q-1} \lambda_s^2 + 0\left(\frac{1}{r}\right).$$

Enfin, avec

$$\sum_{s=d+1}^{q-1} \lambda_s^2 = \mathbf{E}\left[\sum_{s=d+1}^{q-1} \widehat{\lambda}_s^2 - \frac{\widehat{\sigma}^2}{r}[pq - (d+1)(p+q-d-1)]\right]$$

on a la

Proposition 3.4 *Lorsque r tend vers l'infini, un estimateur convergent de l'$EQMP_d$ est donné par :*

$$C_d = \sum_{s=d+1}^{q-1} \widehat{\lambda}_s^2 + 2\frac{\widehat{\sigma}^2}{r}(d+1)(p+q-d-1)$$

où $\widehat{\sigma}^2$ désigne l'estimateur usuel de σ^2 dans le modèle (M) □

Remarque 4. *D'après les résultas obtenus dans le MLG lorsque le paramètre d'échelle $\zeta = \sigma^2$ tend vers zéro, le critère C_d reste valable pour de «faibles valeurs» de $(\sigma^2/\lambda_s^2)_{s=1,\dots,(q-1)}$.*

4. Application au choix de modèles

Nous proposons maintenant quelques exemples d'utilisation de la quantité C_d pour le choix de modèles en comparaison avec des procédures de tests d'hypothèses proposées dans la littérature ([17] [18] [26]). Ces procédures consistent à donner sous l'hypothèse du modèle additif (M_0) suivant

$$(M_0)\begin{cases} Y_{ij} = \mu + a_i + b_j + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0,\sigma^2) \quad \text{i.i.d. où } \sigma^2 \text{ est supposée connue.} \end{cases}$$

et pour des couples (p, q) donnés, une approximation des quantiles de statistiques de test dont les densités sont relativement complexes.

D'autre part, dans (M_d), le problème fréquemment rencontré est l'estimation de σ^2. Nous prenons, dans les exemples qui suivent, l'estimateur MV «corrigé» en tenant compte du degré de liberté de la partie résiduelle *i.e.* :

$$\widehat{\sigma}^2 = \frac{\sum_{s=d+1}^{q-1} \widehat{\lambda}_s^2}{(p-d-1)(q-d-1)}.$$

La seconde partie des applications est consacrée aux simulations. Elle montre les résultats de notre critère sur des tableaux pour lesquels d est connu.

4.1 Exemple 1

On considère dans la table 1 le jeu de données de Hegemann et Johnson [15] pour tester la dimension du modèle d'analyse de la variance. Dans un premier temps, nous avons conçu un algorithme itératif pour l'estimation des paramètres multiplicatifs de (M_d) avec le logiciel $GLIM$ que nous ne présentons pas ici (cf. [10] ou de Falguerolles et Francis [11]). Pour chacun des trois termes nous avons les estimations suivantes :

$$\widehat{\lambda}_1^2 = 366 \quad \widehat{\lambda}_2^2 = 29.9 \quad \widehat{\lambda}_3^2 = 11.5.$$

TABLE 1

(q,p)	1	2	3	4	5
1	28.2	29.3	33.7	41.2	50.9
2	23.5	24.8	24.1	34.7	32.8
3	17.4	15.2	17.8	14.7	16.6
4	10.1	11.5	15.6	9.9	4.7

On remarque déjà une part importante au premier terme d'interaction, soit 90 %. Nous avons impliqué notre critère pour le choix de modèles et nous obtenons :

$$C_0 = 518 \quad C_1 = 234 \quad C_2 = 260 \quad C_3 = 276.$$

D'après notre critère, le modèle (M_1) se présente comme le meilleur pour la prédiction. Il est à noter que ce choix n'est pas en contradiction avec les résultats de Hegemann et Johnson qui retiennent le même modèle par leur procédure de test.

4.2 Exemple 2

On considère dans la table 2 le jeu de données de Krishnaiah et Yochmowitz [18] sur des singes. Ce sont des mesures de cholestérol de 19 singes mâles sur trois périodes en 1977 et sur trois autres en 1978. L'étude consiste à analyser l'effet *période* $\times$ *sujet* et à choisir le modèle d'interaction adéquat pour la prédiction. Dans cet exemple, l'estimation des λ_s^2 donne :

$$\widehat{\lambda}_1^2 = 19519 \quad \widehat{\lambda}_2^2 = 5263 \quad \widehat{\lambda}_3^2 = 2185 \quad \widehat{\lambda}_4^2 = 1668 \quad \widehat{\lambda}_5^2 = 1256$$

TABLE 2

Périodes Sujets	**771**	**772**	**773**	**781**	**782**	**783**
1	125	105	106	107	130	158
2	122	106	93	97	126	126
3	116	84	89	118	129	130
4	111	149	73	101	130	148
5	120	88	104	116	124	173
6	127	231	139	109	138	164
7	135	94	142	98	119	148
8	130	103	127	124	132	149
9	170	120	125	173	160	196
10	132	105	132	117	136	158
11	121	149	104	107	94	120
12	108	76	108	112	116	132
13	134	75	112	107	113	148
14	105	128	141	108	135	143
15	143	119	114	118	153	145
16	110	86	99	102	100	117
17	119	91	105	123	121	149
18	124	98	118	103	102	127
19	107	99	98	77	110	125

et les valeurs du critère sont, à une constante multiplicative près (10^3) :

$$C_0 = 372 \qquad C_1 = 244 \qquad C_2 = 252 \qquad C_3 = 285.$$

Comme précédemment, nous retenons le même modèle que Krishnaiah et Yochmowitz à savoir le modèle (M_1).

4.3 Simulations

On résume dans les tables suivantes les résultats de simulations effectuées avec le logiciel de calcul matricel *GAUSS* où d^* désigne la vraie valeur du nombre de termes multiplicatifs et d le choix retenu par le critère C_d. Ces résultats donnent pour différentes valeurs de (λ^2/σ^2), le pourcentage d'obtention des modèles (M_0), (M_1), (M_2) et (M_3). Pour chaque couple (p, q), on a effectué 1000 simulations, chacune consistant en un jeu de données comprenant 50 observations par cellules et $\sigma^2 = 10$.

Afin de mesurer l'influence du second terme d'interaction, les résultats de la table 5 ont été obtenus en fixant $\lambda_1^2 = \sigma^2 = 10$.

Les résultats présentés ci-dessous ne sont qu'une partie des simulations effectuées pour le choix de modèles. Il nous semble donc important de signaler que les résultats obtenus avec $r = 10$ sont assez voisins de ceux présentés ici. Sur l'ensemble de nos résultats, en dehors de l'influence du nombre d'observations par cellule, nous remarquons que le critère C_d fournit des résultats d'autant meilleurs que p est grand.

TABLE 3

(p,q)	d^*	(λ^2/σ^2)	$d = 0$	$d = 1$	$d = 2$	$d = 3$
(5,4)	0	0	80 %	20 %	-	-
(10,4)	0	0	93 %	7 %	-	-
(15,4)	0	0	98 %	2 %	-	-
(20,4)	0	0	99 %	1 %	-	-

TABLE 4

(p,q)	d^*	(λ^2/σ^2)	$d=0$	$d=1$	$d=2$	$d=3$
(5,4)	1	0.1	50 %	45 %	5 %	-
(5,4)	1	1	-	86 %	11 %	2 %
(5,4)	1	2	-	88 %	12 %	-
(10,4)	1	0.1	74 %	23 %	3 %	-
(10,4)	1	1	1 %	85 %	14 %	-
(10,4)	1	2	-	89 %	11 %	-
(15,4)	1	0.1	84 %	16 %	-	-
(15,4)	1	1	-	96 %	4 %	-
(15,4)	1	2	-	97 %	3 %	-
(20,4)	1	0.1	93 %	7 %	-	-
(20,4)	1	1	2 %	98 %	-	-
(20,4)	1	2	-	100 %	-	-

TABLE 5

(p,q)	d^*	(λ_2^2/σ^2)	$d=0$	$d=1$	$d=2$	$d=3$
(5,4)	2	0.1	-	53 %	42 %	5 %
(5,4)	2	0.5	-	-	86 %	14 %
(5,4)	2	1	-	-	92 %	8 %
(10,4)	2	0.1	-	74 %	25 %	1 %
(10,4)	2	0.5	-	10 %	85 %	5 %
(10,4)	2	1	-	-	96 %	4 %
(15,4)	2	0.1	-	83 %	17 %	-
(15,4)	2	0.5	-	24 %	74 %	2 %
(15,4)	2	1	-	1 %	99 %	-
(20,4)	2	0.1	2 %	92 %	6 %	-
(20,4)	2	0.5	-	45 %	54 %	1 %
(20,4)	2	1	-	4 %	96 %	-

5. Conclusion

La résolution de problèmes concrets, comme le choix de modèles dans une analyse de la variance avec une observation par cellule, pose souvent la question de la validité des résultats asymptotiques. Sans perdre de vue la puissance de ces derniers, nous avons réexaminé les hypothèses de la théorie asymptotique pour proposer des solutions à ces problèmes.

Tout en répondant à un besoin pratique, notre étude sur le paramètre d'échelle du MLG reste relativement pauvre quant à son champ application. Enfin, au delà du modèle (M_d) d'analyse de la variance, l'introduction du modèle linéaire généralisé au second paragraphe permet l'étude asymptotique dans nombre d'autres modèles et sans doute une ouverture à d'intéressantes autres investigations.

Nous remercions les rapporteurs de cet article et J.-B. Denis pour leurs nombreuses remarques qui ont permis d'améliorer ce travail.

Bibliographie

[1] AITCHISON J., SILVEY S.D. (1958). *Maximum likelihood estimation of parameters subject to restrictions*. Annals of Math. Stat. 29, pp. 813

[2] ALLEN D.M. (1971). *Mean square error of prediction as criterion selecting variables*. Technometrics 13, pp. 469

[3] ALLEN D.M. (1974). *The relationship between variable selection and data augmentation and method of prediction.* Technometrics 16, pp. 125

[4] BOIK R.J. (1986). *Testing the rank of a matrix with applications to the analysis of interaction in ANOVA*. J.A.S.A. 81, pp. 243

[5] BOIK R.J. (1989). *Reduced-rank models for interaction in unequally replicated two-way classifications*. Journal of Multivariate Analysis 28, pp. 69

[6] BRADU D., GABRIEL K.R. (1974). *Simultaneous statistical inference on interactions in two-way analysis of variance.* J.A.S.A. 68, pp. 428

[7] CHADOEUF J., DENIS J.B. (1991). *Asymptotic variances for the multiplicative interaction model.* Journal of Applied Statistics, Vol. 18, N° 3, pp. 331

[8] CORSTEN L.C.A., VAN EIJSBERGEN A.C. (1972). *Multiplicative effects in two-way analysis of variance*. Statistica Neerlandica 26, pp. 61

[9] DENIS J.-B (1991). *Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes*. R.S.A. Vol. 39 N° 2, pp. 5-24

[10] DORKENOO K.M.M. (1992). *Etude de modèles avec interaction multiplicative en analyse de la variance*. Thèse N.R. Toulouse-France

[11] FALGUEROLLES A. de, FRANCIS B.(1992). *Algorithmic Approaches for Fitting Bilinear Models* Computational Statistics, Physica-Verlag, pp. 77-82

[12] GABRIEL K.R. (1978). *Least squares approximation of matrices by additive and multiplicative models*. J.R.S.S. série B, 40, pp. 186

[13] GOLLOB H.F. (1968) *A statistical model wich combines features of factor analysis and anova techniques.* Psychometrika 33, pp. 73

[14] GOODMAN L.A., HABERMAN S. (1990). *The analysis of non-additivity in two-way analysis of variance.* J.A.S.A. 85, pp. 139

[15] HEGEMANN V.J., JOHNSON D.E. (1976) *On analyzing two-way analysis of variance data with interaction.* Technometrics 18, pp. 273

[16] JOHNSON D.E., GRAYBILL F.A. (1972). *On analysis of a two-may model with interaction and no replication.* J.A.S.A. 67, pp. 862

[17] JOHNSON D.E. (1976). *Some new multiple comparison procedures for two-way anova model with interaction.* Biometrics 32, pp. 929

[18] KRISHNAIAH P.R., YOCHMOWITZ M.G. (1980). *Inference of interaction in two-way classification model.* Handbook of Statistics, Vol. 1, pp. 973

[19] MALLOWS C.L. (1973). *Some comments on C_p.* Technometrics 15, pp. 661

[20] MANDEL J. (1969). *The partitioning of interaction in analysis of variance.* Journal of Research – National Bureau of Standard, B.73

[21] MANDEL J. (1970). *Distribution of eigenvalues of covariance matrices of residuals in analysis of variance.* Journal of Research – National Bureau of Standard, B.74, pp. 149

[22] MANDEL J. (1971). *A new analysis of variance model for non-additive data.* Technometrics 13, pp. 1

[23] MATHIEU J.R. (1981). *Tests of χ^2 in the generalized linear model.* Statistics Vol. 12, 4, pp. 509

[24] NELDER J.A., WEDDERBURN R.W.M. (1972). *Generalized linear models.* J.R.S.S. série A, 135, pp. 370

[25] ROBERT C. (1982). *Propriétés optimales de certains estimateurs d'interaction en analyse de la variance.* Thèse 3^e cycle Grenoble, France

[26] SCHUURMANN F.J., KRISCHNAIAH P.R., CHATTOPADHYAY (1973). *On the distributions of the ratios of the extreme roots to the trace of the Wishart matrix.* Journal of Multivariate Analysis 3, pp. 445

[27] TUKEY J.W. (1949). *One degree of freedom for non-additivity.* Biometrics Vol.5, pp. 232

[28] WILLIAMS E.J. (1952). *The interpretation of interactions in factorials experiments.* Biometrika 39, pp. 65