J.-R. MATHIEU

E. LAMBERT

Étude d'un test de l'identité des marges d'un tableau de corrélation

Revue de statistique appliquée, tome 18, n° 1 (1970), p. 5-19

<http://www.numdam.org/item?id=RSA_1970__18_1_5_0>

ÉTUDE D'UN TEST DE L'IDENTITÉ DES MARGES D'UN TABLEAU DE CORRELATION

J.-R. MATHIEU et E. LAMBERT

Laboratoire de statistique, Faculté des Sciences, Toulouse

INTRODUCTION

Soit θ un paramètre à m dimensions : $\theta = [\theta_1, \theta_2, \ldots \theta_m]'$, prenant ses valeurs dans une partie ω_0 non vide de R^m. Nous notons H_0 l'hypothèse triviale : $\theta \in \omega_0$, et nous considérons les deux hypothèses emboitées H_1 et H_2 ;

$H_1 : \theta \in \omega_1 \subset \omega_0$; $H_2 : \theta \in \omega_2 \subset \omega_1 \subset \omega_0$, ω_1 et ω_2 étant non vides.

Un test restreint de H_2 contre $(H_1 - H_2)$ est un test de l'hypothèse nulle : $\theta \in \omega_2$, contre l'hypothèse alternative : $\theta \in \omega_1 - \omega_2$.

Plusieurs auteurs ont construit et étudié les propriétés des tests restreints : en particulier, d'après AITCHISON [1], dans la mesure où un test de H_1 contre $(H_0 - H_1)$, c'est-à-dire en prenant pour alternative : $\theta \in \omega_0 - \omega_1$, a conduit à accepter H_1, alors le test restreint de H_2 contre $(H_1 - H_2)$ est plus puissant que ne l'est, pour les alternatives : $\theta \in \omega_1 - \omega_2$, le test de H_2 contre $(H_0 - H_2)$. FIX, HODGES et LEHMANN [3] ont construit des tests restreints du χ^2.

AITCHINSON [1] a introduit la notion d'hypothèses séparables : supposons maintenant que H_1 et H_2 ne sont plus emboitées et sont telles que ω_1 et ω_2 aient une intersection non vide ; notons $H_1 \cap H_2$, l'hypothèse : $\theta \in \omega_1 \cap \omega_2$ et considérons les tests restreints de $H_1 \cap H_2$ contre $(H_1 - H_1 \cap H_2)$ et de $H_1 \cap H_2$ contre $(H_2 - H_1 \cap H_2)$: on dit que H_1 et H_2 sont séparables pour les tests utilisés, si ces deux tests restreints ont respectivement même région critique que les tests de H_2 contre $(H_0 - H_2)$ et de H_1 contre $(H_0 - H_1)$.

Nous savons que les deux hypothèses d'indépendance (H) et d'identité des marges (I.M.) d'un tableau de corrélation k x k sont séparables pour les tests du χ^2 ; nous avons voulu utiliser cette propriété pour construire un test de I.M. contre H_0 - I.M. à l'aide de la statistique $\chi^2_{H \cap I.M.} - \chi^2_H$, statistique du test restreint de H $\cap$ I.M. contre H - H $\cap$ I.M. [5]. Pour déterminer la région critique de seuil α de ce test, nous avons établi le résultat suivant concernant la distribution de $\chi^2_{H_1 \cap H_2} - \chi^2_{H_2}$, H_1 et H_2 étant deux hypothèses quelconques [4] : $\chi^2_{H_1 \cap H_2} - \chi^2_{H_2}$ converge en loi, lorsque la taille n de l'échantillon tend vers l'infini, vers une variable aléatoire de χ^2 centré, ceci étant vrai non seulement sous H_2 mais également sous certaines alternatives de H_2. Par suite, si H_1 est l'hypothèse I.M. et si H_2 est l'hypothèse H, le test que nous proposons pour éprouver I.M. contre (H_0 - I.M.) a asymptotiquement pour région critique de seuil α, la partie de l'espace R^n des observations où $\chi^2_{H \cap I.M.} - \chi^2_H$ est supérieure ou égale à χ^2_α, valeur limite, au seuil

α, du χ^2 centré à k - 1 degrés de liberté ; le seuil α est alors conservé, même lorsqu'il n'y a pas indépendance.

Afin d'étudier la puissance de ce test, nous étudions la convergence en loi de $\chi^2_{H\cap I.M.} - \chi^2_H$, sous les alternatives de I.M. Nous montrons d'abord (I) que, sous certaines alternatives de H_1, $\chi^2_{H_1\cap H_2} - \chi^2_{H_2}$ converge en loi vers une variable aléatoire de χ^2 non centré, tant sous H_2 que sous certaines alternatives de H_2. Ce résultat, appliqué au cas particulier d'un tableau de corrélation (II), permet donc de déterminer la puissance du test de I.M. contre (H_o - I.M.), lorsque ce test est effectué à l'aide de la statistique $\chi^2_{H\cap I.M.} - \chi^2_H$.

D'un po.nt de vue pratique, il est important de savoir dans quelle mesure la technique que nous proposons pour éprouver I.M., conserve, pour des tailles finies de l'échantillon, les propriétés qui découlent du comportement asymptotique de la statistique du test. Dans cet esprit, nous avons simulé sur ordinateur (III) des tableaux de corrélation pour différentes valeurs de la taille de l'échantillon. Nous avons, en outre, comparé les résultats obtenus par le test proposé ici et par le test construit par BHAPKAR [2] pour éprouver I.M.

I - CONVERGENCE EN LOI DE $\chi^2_{H_1\cap H_2} - \chi^2_{H_2}$ SOUS LES ALTERNATIVES DE H_1.

Etant donné une loi multinomiale à K catégories, de paramètres p_i et n, nous notons $q_i = ni/n$ la fréquence relative de la i-ième catégorie ; p_i est fonction d'un paramètre θ à m dimensions.

Soit H_2, l'hypothèse : $p_i = f_i(\theta)$,

f_i étant linéaire par rapport à θ_r ($1 \leq r \leq l < m$) et possédant des dérivées partielles continues jusqu'à l'ordre 2 par rapport à θ_r ($l < r \leq m$).

Nous supposons H_2 fausse, les vraies valeurs p_i^o de p_i vérifiant :

$$p_i^o = f_i(\theta^o) + \frac{d_i^o}{\sqrt{n}} \quad \text{avec} \quad \sum_{i=1}^{K} d_i^o = 0 \tag{1}$$

L'hypothèse H_1, que les d_i^o soient tous nuls ou non, c'est-à-dire, sous H_2 comme sous les alternatives de H_2, se formule :

$$\theta_r^o = \dot{\theta}_r^o \qquad 1 \leq r \leq l \quad \text{où } \dot{\theta}_r^o \text{ est donné.}$$

Nous considérons les alternatives suivantes de H_1 :

$$\theta_r^o = \dot{\theta}_r^o + \frac{\varepsilon_r^o}{\sqrt{n}} \quad 1 \leq r \leq l \tag{2}$$

Par suite, sous $H_1 \cap H_2$: $p_i = f_i(\dot{\theta}^o \quad \theta_{l+1}, \theta_{l+2}, \ldots, \theta_m)$

avec

$$\dot{\theta}^o = [\dot{\theta}_1^o , \dot{\theta}_2^o , \ldots , \dot{\theta}_l^o]'$$

Nous étudions le comportement asymptotique de $\chi^2_{H_1\cap H_2} - \chi^2_{H_2}$, lorsque p_i^o et θ_r^o vérifient (1) et (2), avec :

$$\chi^2_{H_1 \cap H_2} = n \sum_{i=1}^{K} \frac{[q_i - f_i(\dot{\theta}^o, \tilde{\theta_1}(q))]^2}{f_i(\dot{\theta}^o, \tilde{\theta_1}(q))} \quad \text{et} \quad \chi^2_{H_2} = n \sum_{i=1}^{K} \frac{[q_i - f_i(\tilde{\theta_2}(q))]^2}{f_i(\tilde{\theta_2}(q))}$$

où

$$\tilde{\theta_1}(q) = [\tilde{\theta}_{1,l+1}(q), \tilde{\theta}_{1,l+2}(q), \ldots, \tilde{\theta}_{1,m}(q)]'$$

et

$$\tilde{\theta_2}(q) = [\tilde{\theta}_{2,1}(q), \tilde{\theta}_{2,2}(q), \ldots, \tilde{\theta}_{2,m}(q)]',$$

$$\tilde{\theta}_{1,r}(q) \; (l < r \leq m) \quad \text{et} \quad \tilde{\theta}_{2,r}(q) \; (1 \leq r \leq m)$$

étant les estimateurs de θ_r^o, respectivement sous $H_1 \cap H_2$ et sous H_2.

Pour étudier la convergence en loi de $\chi^2_{H_1 \cap H_2} - \chi^2_{H_2}$, nous reprenons la démarche suivie par NEYMANN [6].

Soit :

$$\chi_o^2 = n \sum_{i=1}^{K} \frac{(q_i - \Phi_i)^2}{\dot{f}_i^o} \quad \text{où} \quad \Phi_i = f_i^o + \sum_{r=1}^{m} \dot{f}_{i,r}^o (\theta_r - \theta_r^o)$$

avec

$$\dot{f}_i^o = f_i(\dot{\theta}_1^o, \dot{\theta}_2^o, \ldots, \dot{\theta}_l^o, \theta_{l+1}^o, \theta_{l+2}^o, \ldots, \theta_m^o)$$

$$f_i^o = f_i(\theta_1^o, \theta_2^o, \ldots, \theta_m^o)$$

et

$$\dot{f}_{i,r}^o = \left(\frac{\partial f_i}{\partial \theta_r}\right)_{\theta = (\dot{\theta}_1^o, \dot{\theta}_2^o, \ldots, \dot{\theta}_l^o, \theta_{l+1}^o, \theta_{l+2}^o, \ldots, \theta_m^o)}$$

Nous minimisons χ_o^2 successivement sous H_2 et sous $H_1 \cap H_2$.

Soit Q_2 le minimum de χ_o^2 sous H_2, obtenu en remplaçant θ_r par $\overline{\theta}_{2,r}(q)$, estimateur de θ_r^o obtenu en minimisant χ_o^2. On détermine $\overline{\theta}_{2,r}(q)$ en résolvant le système :

$$\left[\frac{\partial \chi_o^2}{\partial \theta_s}\right]_{\theta = \bar{\theta}_2(q)} = 0 \qquad s = 1, 2, \ldots, m.$$

On a alors :

$$\bar{\theta}_{2,r}(q) = \theta_r^o + \frac{[(L'L)^{-1} L'\Delta]_r}{\sqrt{n}}$$

avec

$$[L]_{i,j} = \frac{\dot{f}_{i,j}^o}{\sqrt{\dot{f}_i^o}} \quad \text{et} \quad [\Delta]_j = \frac{\sqrt{n}\,(q_j - f_j^o)}{\sqrt{\dot{f}_j^o}}$$

On obtient immédiatement : $Q_2 = \Delta'[I - L(L'L)^{-1}L']\Delta$ où I est la matrice unité.

De même, soit Q_1 le minimum de χ_o^2 sous $H_1 \cap H_2$, c'est-à-dire en donnant à θ_r $(1 \leq r \leq l)$ la valeur $\dot{\theta}_r^o$. On a alors, en raison de la linéarité

de $f_i(\theta)$ par rapport à $\theta_r (1 \leq r \leq l)$, $\Phi_i = \dot{f}_i^o + \sum_{r=l+1}^{m} \dot{f}_{ir}^o (\theta_r - \theta_r^o)$.

Q_1 est obtenu en remplaçant θ_r $(l < r \leq m)$ par les solutions $\bar{\theta}_{1,r}(q)$ du système :

$$\left[\frac{\partial \chi_o^2}{\partial \theta_s}\right]_{\theta = [\dot{\theta}^o, \bar{\theta}_1(q)]} = \qquad s = l+1,\ l+2\ \ldots,\ m.$$

On obtient :

$$Q_1 = \Delta_1'[I - L_1(L_1'L_1)^{-1}L_1']\Delta_1$$

où L_1 est la matrice L privée de ses l premières colonnes et où

$$[\Delta_1]_j = \frac{\sqrt{n}(q_j - \dot{f}_j^o)}{\sqrt{\dot{f}_j^o}}$$

D'autre part, les estimateurs $\tilde{\theta}_{2,r}(q)$ vérifient :

$$\left(\frac{\partial \chi^2}{\partial \theta_s}\right)_{\theta = \tilde{\theta}_2(q)} = 0 \qquad s = 1, 2, \ldots, m \quad \text{avec} \quad \chi^2 = \sum_{i=1}^{K} \frac{[q_i - f_i(\theta)]^2}{f_i(\theta)}$$

$\tilde{\theta}_{2,r}(q)$ est donc tel que $\tilde{\theta}_{2,r}(f_i^o) = \theta_r^o$; de plus si $\tilde{\theta}_{2,r}(q)$ possède des dérivés partielles jusqu'à l'ordre 2, ces dérivées partielles sont continues ; on montre alors, en résolvant le système :

$$\left\{\frac{\partial .}{\partial q_j}\left[\left(\frac{\partial \chi^2}{\partial \theta_s}\right)_{\theta = \tilde{\theta}_2(q)}\right]\right\}_{q = f^o} = 0,$$

que

$$\left[\frac{\partial \tilde{\theta}_{2,r}(q)}{\partial q_j}\right]_{q = f^o} = \frac{\partial \bar{\theta}_{2,r}}{\partial q_j} + 0\left(\frac{1}{\sqrt{n}}\right)$$

On a donc, en effectuant un développement de TAYLOR, limité au premier ordre, de $\tilde{\theta}_{2,r}(q)$ au voisinage de θ_r^o :

$$\tilde{\theta}_{2,r}(q) - \theta_r^o = \left[\frac{(L'L)^{-1}\Delta}{\sqrt{n}}\right]_r + \gamma_{r,n}$$

où $n^p\,\gamma_{r,n}$ converge en probabilité vers zéro si $p < 1$.

En effectuant un développement de TAYLOR, limité au premier ordre, de $f_i(\tilde{\theta}_2(q))$, on obtient :

$$f_i(\tilde{\theta}_2(q)) - f_i(\theta^o) = \sum_{j=1}^{m} \dot{f}^o_{i,j} \frac{[(L'L)^{-1} L'\Delta]_j}{\sqrt{n}} + \delta_{i,n},$$

où $n^p \delta_{i,n}$ converge en probabilité vers zéro si $p < 1$.

En outre

$$n \frac{[q_i - f_i(\tilde{\theta}_2)]^2}{f_i(\tilde{\theta}_2)} = n \frac{[q_i - f_i(\tilde{\theta}_2)]^2}{\dot{f}^o_i} + \lambda_{i,n},$$

où $n^p \lambda_{i,n}$ converge en probabilité vers zéro si $p < \frac{1}{2}$.

Par suite

$$\chi^2_{H_2} = \Delta' [I - L(L'L)^{-1} L'] \Delta + \varepsilon_n$$

où $n^p \varepsilon_n$ converge en probabilité vers zéro si $p < \frac{1}{2}$.

On mène le calcul de la même façon pour $\chi^2_{H_1 \cap H_2}$ et on obtient :

$$\chi^2_{H_1 \cap H_2} = \Delta'_1 [I - L_1(L'_1 L_1)^{-1} L'_1]\Delta_1 + \xi_n$$

où $n^p \xi_n$ converge en probabilité vers zéro si $p < \frac{1}{2}$.

On a donc

$$\chi^2_{H_1 \cap H_2} - \chi^2_{H_2} = Q_1 - Q_2 + \eta_n$$

où $n^p \eta_n$ converge en probabilité vers zéro si $p < \frac{1}{2}$.

Pour déterminer la densité asymptotique de $Q_1 - Q_2$, nous posons :

$$X = [X_1, \ldots, X_k]' \quad \text{avec} \quad X_i = \frac{(q_i - p_i^o)\sqrt{n}}{\sqrt{\dot{f}^o_i}}$$

X converge en loi vers une variable aléatoire multinormale de moyenne nulle et de matrice des variances et covariances : $\Lambda = I - f \times f'$

$$\text{avec } f = [\sqrt{\dot{f}^o_1}, \sqrt{\dot{f}^o_2}, \ldots, \sqrt{\dot{f}^o_K}]'$$

On a alors

$$\chi^2_o = \sum_{i=1}^{K} (X_i - \alpha_i)^2,$$

où α_i est une fonction linéaire des paramètres $\theta_r - \theta_r^o$ $(1 \leq r \leq m)$ et d_i^o ; α_i est donc nul si $d_i^o = 0$ et $\theta_r = \theta_r^o$.

A l'aide d'un changement de variable [6], on obtient :

$$\chi_o^2 = \sum_{i=1}^{K-s-m} Z_i^2 + \sum_{j=1}^{s+m} (Z_{K-s-m+j} - \sum_{v=1}^{j} t_{jv}\mu_v)^2 \qquad (3)$$

où t_{jv} est une constante indépendante de θ_r et d_i^o ;

où $\mu_1, \mu_2, \ldots, \mu_s$ constituent un système de paramètres indépendants parmi les K paramètres d_i^o ; et où $\mu_{s+r} = \theta_r - \theta_r^o$ $(1 \leq r \leq m)$.

Les variables aléatoires Z_i ont pour densité conjointe

$$g(z) = C \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{K+s+m} z_i^2\right)\right]$$

Q_2 est alors le minimum obtenu en donnant θ_r $(1 \leq r \leq m)$, donc à μ_v $(s < v \leq s + m)$, des valeurs telles que les m derniers termes du second membre de (3) s'annulent, les paramètres d_i^o ayant alors des valeurs non toutes nulles ; on a donc :

$$Q^2 = \sum_{i=1}^{K-s-m} Z_i^2 + \sum_{j=1}^{s} (Z_{K-s-m+j} - \sum_{v=1}^{j} t_{jv}\,\dot{\mu}_v)^2 \text{ les } \dot{\mu}_v \text{ étant non tous nuls.}$$

Q_2 converge donc en loi vers une variable aléatoire de χ^2 à K - m degrés de liberté et de paramètre d'excentricité λ_2

De même Q_1 est le minimum de χ_o^2 obtenu en donnant à μ_v $(s < v \leq s + l)$ les valeurs $\dot{\theta}_r^o - \theta_r^o$ et à μ_v $(s + l < v \leq s + m)$ des valeurs telles que les $m - l$ derniers termes du second membre de (3) s'annulent ; on a donc :

$$Q_1 = \sum_{i=1}^{K-s-m} Z_i^2 + \sum_{j=1}^{s+l} (Z_{K-s-m+j} - \sum_{v=1}^{j} t_{jv}\,\dot{\mu}_v)^2$$

Q_1 converge donc en loi vers une variable aléatoire de χ^2 à $K - m + l$ degrés de liberté et de paramètre d'excentricité λ_1

On a donc :

$$Q_1 - Q_2 = Q_r = \sum_{j=s+1}^{s+l} (Z_{K-s-m+j} - \sum_{v=1}^{j} t_{jv}\,\dot{\mu}_v)^2$$

qui converge en loi vers une variable aléatoire de χ^2, indépendant de Q_2, à l degrés de liberté et de paramètre d'excentricité λ ; on a :

$$E[Q_1] = E[Q_2] + E[Q_r] \text{ soit } K - m + l + \lambda_1 = K - m + \lambda_2 + l + \lambda$$

$$\text{soit } \lambda = \lambda_1 - \lambda_2$$

II - APPLICATION A UN TABLEAU DE CORRELATION k x k

Soit p_{ij} la probabilité relative à la cellule (i, j) ; i = 1, 2, ..., k ; j = 1, 2, ..., k.

Posons :

$$p_{i.} = \sum_{j=1}^{k} p_{ij} \quad \text{et} \quad p_{.j} = \sum_{i=1}^{k} p_{ij}, \quad \text{avec} \quad \sum_{i=1}^{k} p_{i.} = \sum_{j=1}^{k} p_{.j} = 1$$

Posons :

$$\theta_r \doteq p_{.r} - p_{r.} \quad , \quad r = 1, \ldots, k-1,$$

et

$$\theta_{k-1+r} = p_{r.} \quad , \quad r = 1, \ldots, k-1.$$

Soit H_2 l'hypothèse H d'indépendance : $p_{ij} = p_{i.}\ p_{.j} \forall_i$ et j.

Donc, sous H : $p_{ij} = h_{ij}(\theta)$, où les fonctions h_{ij} jouent le rôle des fonctions f_i introduites au paragraphe précédent : on vérifie immédiatement que les fonctions $h_{ij}(\theta)$ vérifient les conditions imposées aux fonctions $f_i(\theta)$.

Nous supposons donc que H est fausse et que :

$$p_{ij} = h_{ij}(\theta^o) + \frac{d_{ij}^o}{\sqrt{n}} \quad \text{avec} \quad d_{i.}^o = \sum_{j=1}^{k} d_{ij}^o = 0 \quad \text{et} \quad d_{.j}^o = \sum_{i=1}^{k} d_{ij}^o = 0$$

H_1 est l'hypothèse I.M. d'identité des marges : $p_{i.} = p_{.i}\ \forall_i$

soit

$$\dot{\theta}_r^o = 0 \qquad r = 1, 2, \ldots, k-1$$

Nous supposons que H est fausse et que θ_r^o vérifie :

$$\theta_r^o = \dot{\theta}_r^o + \frac{\varepsilon_r^o}{\sqrt{n}} = \frac{\varepsilon_r^o}{\sqrt{n}} \quad , \quad r = 1, 2, \ldots, k-1, \quad \text{avec} \quad \sum_{r=1}^{k} \varepsilon_r^o = 0 \tag{4}$$

Nous savons que H et I.M. sont séparables pour les tests du χ^2 et, que $\chi^2_{H \cap I.M.} - \chi^2_H$ converge en loi, lorsque I.M. est vraie, vers une variable aléatoire de χ^2 à k - 1 degrés de liberté, avec :

$$\chi^2_{H \cap I.M.} = \sum_{i,j} \frac{\left[x_{ij} - \frac{(x_{i.} + x_{.i})(x_{j.} + x_{.j})}{4n} \right]^2}{\frac{(x_{i.} + x_{.i})(x_{j.} + x_{.j})}{4n}}$$

et

$$\chi^2_H = \sum_{i,j} \frac{\left[x_{ij} - \frac{x_{i.}\ x_{.j}}{n} \right]^2}{\frac{x_{i.}\ x_{.j}}{n}}$$

où x_{ij} est le nombre des observations dans la cellule (i, j) et où

$$x_{i.} = \sum_{j=1}^{k} x_{ij} \qquad \text{et} \qquad x_{.j} = \sum_{i=1}^{k} x_{ij}$$

D'après le résultats de I, $\chi^2_{H \cap I.M.} - \chi^2_H$ converge en loi, lorsque I.M. est fausse, c'est-à-dire, lorsqu'il existe r tel que ε^o_r soit non nul, vers une variable aléatoire de χ^2 à k - 1 degrés de liberté, de paramètre d'excentricité :

$$\lambda = \lambda_1 - \lambda_2 = \frac{1}{2} \sum_{r=1}^{k} \frac{\varepsilon^{o2}_r}{p^o_{r.}} \tag{5}$$

Ce résultat constitue donc une étude de la puissance du test de I.M., contre H_o - I.M., lorsque I.M. est éprouvée par la statistique

$$\chi^2_{H \cap I.M.} - \chi^2_H$$

III - SIMULATION DE TABLEAUX DE CORRELATION

D'après la convergence en loi de $\chi^2_{H \cap I.M.} - \chi^2_H$, lorsque I.M. est vraie, la région critique de seuil α du test est donc asymptotiquement la région de l'espace des observations où $\chi^2_{H \cap I.M.} - \chi^2_H$ est supérieur ou égal à la valeur χ^2_α telle que si X est une variable de χ^2 à k - 1 degrés de liberté, $\Pr[X \geq \chi^2_\alpha] = \alpha$. Pour des tailles finies de l'échantillon, la région critique est encore celle qui vient d'être définie, mais nous ignorons si elle correspond toujours à un seuil α. Dans le but de répondre à cette question, nous avons simulé sur ordinateur I.B.M. 7044 des tableaux de corrélation k x k en fixant n, $p^o_{i.}$, $p^o_{.j}$ et d^o_{ij} ; on a alors :

$$p^o_{ij} = p^o_{i.} \, p^o_{.j} + \frac{d^o_{ij}}{\sqrt{n}} \tag{6}$$

(d^o_{ij} nul pour tout (i, j) entraîne l'indépendance).

Une fois déterminé p^o_{ij}, le processus de simulation se ramène à l'observation de la variable aléatoire uniformément répartie sur [0,1], l'intervalle [0,1] étant découpé en intervalles de longueur p^o_{ij} ; du nombre d'observations dans chacun de ces intervalles, on déduit le tableau de corrélation observé. Pour chaque tableau observé, nous avons calculé :

$$\chi^2_{H \cap I.M.} \; \chi^2_H , \qquad \text{puis} \qquad \chi^2_{H \cap I.M.} - \chi^2_H .$$

Pour étudier le comportement, pour n fini, du seuil α, nous avons, pour chaque valeur de k (k = 3, 4, 5, 6), fixé les valeurs de $p^o_{i.}$ et pris $p^o_{.j}$ égal à $p^o_{j.}$, puis déterminé deux ensembles de valeurs d^o_{ij} : un premier ensemble vérifiant H, c'est-à-dire avec les d^o_{ij} tous nuls et un second ensemble de valeurs d^o_{ij} non toutes nulles. Pour chaque valeur des paramètres k, $p^o_{i.}$, et d^o_{ij}, nous avons fait varier n, déterminé, pour chaque valeur de n, p^o_{ij} suivant (6) et simulé pour ces valeurs de p^o_{ij}, 100 tableaux de corrélation ; pour chaque tableau observé, nous avons comparé la valeur observée de $\chi^2_{H \cap I.M.} - \chi^2_H$ à la valeur χ^2_α (α = 5 % ou α = 1 %) ; nous avons donc soit accepté soit rejeté I.M., alors que I.M. est vraie. Les résultats obtenus, nombre de rejets de I.M. pour 100 tableaux observés, sont présentés dans le tableau 1 (colonnes (1)).

Pour chacun de ces tableaux observés, nous avons également calculé la valeur observée de la statistique proposée par BHAPKAR pour éprouver I.M. [2] :

$$\chi^2_{I.M.} = d' W^{-1} d$$

où

$$[d]_i = x_{i.} - x_{.i} \qquad i = 1, 2, \ldots, k-1$$

et

$$[W]_{ij} = \left[\delta_{ij} (x_{i.} + x_{.i}) - x_{ij} - x_{ji} - \frac{d_i d_j}{n} \right] \begin{array}{l} i = 1, 2, \ldots, k-1 ; \\ j = 1, 2, \ldots, k-1 . \end{array}$$

δ_{ij} étant le symbole de KRONECKER.

Cette statistique est asymptotiquement distribuée comme un χ^2 à k - 1 degrés de liberté, centré si I.M. est vraie. Les résultats obtenus par ce test, nombre de rejets à tort de I.M., sont présentés dans le tableau 1 (colonnes (2)).

Nous avons voulu compléter cette étude numérique par une étude pour n fini, de la puissance de ces deux tests de I.M. Nous avons donc, pour chaque valeur de k, fixé les valeurs de $p^o_{i.}$, les valeurs ε^o_r non nulles, ce qui implique que I.M. est fausse, puis deux systèmes de valeurs d^o_{ij} ; et nous avons déterminé, pour chaque valeur de n, les valeurs $p^o_{.j}$ satisfaisant (4) et les valeurs de p^o_{ij} satisfaisant (6). Pour chaque valeur des paramètres, nous avons simulé 100 tableaux et noté le nombre de fois où l'observation, d'une part de $\chi^2_{H \cap I.M.} - \chi^2_H$, d'autre part, de la statistique de BHAPKAR, a conduit à rejeter I.M., alors que I.M. est fausse, α ayant successivement les valeurs 5 % et 1 %.

Ces résultats sont présentés dans les tableaux 2, 3, 4, 5.

Pour chaque valeur de k, nous avons choisi quatre ensembles de valeurs ε^o_r :

$$\varepsilon^o_j = [\varepsilon^o_{j,1}, \varepsilon^o_{j,2}, \ldots, \varepsilon^o_{j,k}] \; (j = 1, 2, 3, 4).$$

tels que

$$\varepsilon^o_{1,r} = 2 \, \varepsilon^o_{2,r} = 5 \, \varepsilon^o_{3,r} = 10 \, \varepsilon^o_{4,r} \text{ pour tout } r.$$

On a donc d'après (5) :

$$\lambda_1 = 4 \, \lambda_2 = 25 \, \lambda_3 = 100 \, \lambda_4$$

Lorsque les d^o_{ij} sont non tous nuls, nous nous sommes efforcés de prendre des valeurs telles que le test de l'hypothèse H par la statistique χ^2_H ait une puissance suffisamment grande : sous l'hypothèse I.M., l'hypothèse H, alors qu'elle est fausse, a été rejetée au moins 33 fois sur 100 au seuil de 1 % et au moins 52 fois sur 100 au seuil de 5 % ; lorsque I.M. est fausse, l'hypothèse H a été rejetée, à juste titre, au moins 68 fois sur 100 au seuil 1 % et au moins 84 fois sur 100 au seuil 5 %.

De ces résultats numériques, il semble ressortir que le test que nous proposons pour éprouver I.M., a, pour n fini, une erreur de première espèce qui n'est supérieure ni au seuil asymptotique α, ni à l'erreur de première espèce du test de BHAPKAR. En ce qui concerne la puissance, il semble que lorsque H est vérifié, la puissance ne soit différente ni du résultat asymptotique ni de la puissance du test de BHAPKAR. Lorsqu'il n'y a pas indépendance, la puissance du test de I.M. par $\chi^2_{H \cap I.M.} - \chi^2_H$ semble supérieur à celle du test de BHAPKAR sauf pour $k = 3$ et $\varepsilon = \varepsilon_4$ où le test de BHAPKAR est plus puissant ; il semble toutefois que dans ce cas, l'erreur de première espèce du test de BHAPKAR soit plus grande que celle du test proposé ici.

Ce test de I.M. contre (H_0 - I.M.) représente une application de la théorie des tests restreints et des hypothèses séparables ; et il est intéressant de noter que, dans ce cas particulier, les résultats numériques permettent de penser que ce test concerve, pour n fini, ses propriétés asymptotiques.

En outre, ces résultats sont comparables à ceux qui sont obtenus par un autre test, (celui de BHAPKAR) dont la pratique nécessite l'inversion d'une matrice qui peut être singulière, ou simplement, s'avérer singulière lorsque l'on recherche son inverse en double précision (10^{-5}), par les méthodes classiques.

Nous remercions le comité de rédaction de la revue pour les remarques et suggestions concernant cet article.

BIBLIOGRAPHIE

[1] J. AITCHISON (1962) - J.R. Stat. Soc. B. 24 - p. 234 - 250.

[2] V.P. BHAPKAR (1966) - J. Amer. Stat. Ass. 61 - p. 228 - 235.

[3] FIX - HODGES - LEHMANN (1959) - The Harald Cramer Volume p.92 - 107 - Wiley : New-York.

[4] J.R. MATHIEU (1968) - C.R. Acad. Sc. t. 267 A : p. 701 - 703.

[5] J.R. MATHIEU - E. LAMBERT (1968) - C.R. Acad. Sc. t. 267 A : 832 - 834.

[6] J. NEYMANN (1949) - Contribution to the theory of χ^2 tests (Proceedings of the first Berkeley Symposium) : p. 239 - 278.

C.T. Ireland - H.H.K.O. - S. Kullback (1969) - J. Amer. Stat. Ass. 64 - p. 1323-1341

Tableau 1

Erreur de première espèce

n	seuil asymptotique α	k = 3				k = 4				k = 5				k = 6			
		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
100	5 %	10	8	1	5	9	11	8	11	6	8	5	5	-	-	-	-
	1 %	1	2	0	0	2	3	1	1	0	3	2	0	-	-	-	-
200	5 %	4	6	0	4	4	7	3	5	5	4	4	8	2	4	4	3
	1 %	1	2	0	2	1	1	0	0	1	1	3	3	0	1	0	1
300	5 %	3	2	2	10	3	3	2	8	5	5	4	6	6	5	5	9
	1 %	0	0	0	1	1	0	0	1	2	1	2	2	2	1	0	2
400	5 %	4	4	2	6	3	4	1	6	7	8	3	4	5	6	3	4
	1 %	1	1	1	2	2	1	0	0	0	4	1	0	1	2	1	2
500	5 %	6	7	1	2	5	5	2	5	5	5	8	12	2	0	2	7
	1 %	1	1	0	1	1	3	0	0	2	2	0	2	0	0	0	0
600	5 %	7	6	2	4	9	6	10	13	6	8	5	6	5	6	3	3
	1 %	1	0	0	1	2	3	1	1	3	3	2	0	1	1	1	0
700	5 %	13	13	1	3	5	6	4	4	9	8	1	4	10	9	2	3
	1 %	2	2	0	0	0	0	1	1	4	4	0	1	2	1	0	0
800	5 %	6	7	1	4	6	6	6	11	4	5	0	2	4	6	4	7
	1 %	0	0	0	1	1	0	2	3	1	1	0	0	2	2	1	3
900	5 %	8	9	3	9	4	4	3	4	6	6	1	2	2	2	3	6
	1 %	2	1	0	4	2	1	0	0	0	0	0	1	0	0	0	1
1 000	5 %	4	3	3	5	7	7	2	3	6	7	2	3	3	3	6	10
	1 %	1	1	0	2	3	3	0	1	2	3	2	2	1	0	0	0

Colonnes (1) : Pour 100 tableaux observés nbre de tableaux où l'observation de $\chi^2_{H \cap I.M.} - \chi^2_H$ à conduit à rejeter I.M. à tort.

Colonnes (2) : Pour 100 tableaux observés nbre de tableaux où l'observation de la statistique de BHAPKAR à conduit à rejeter I.M. à tort.

Tableau 2

$k = 3 \; ; \; \varepsilon_1 = [1 - 4, 3]$

n	seuil asymptotique α	ε_1 ; $\lambda_1 = 40{,}166$				ε_2 ; $\lambda_2 = 10{,}04$				ε_3 ; $\lambda_3 = 1{,}60$				ε_4 ; $\lambda_4 = 0{,}40$			
		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
100	5 %	100	100	100	100	70	78	98	72	17	22	45	19	7	5	2	16
	1 %	100	100	100	100	51	53	92	43	7	6	29	3	3	4	0	6
200	5 %	100	100	100	100	88	89	86	69	29	27	34	11	3	4	2	9
	1 %	100	100	100	100	54	52	79	53	8	10	18	2	2	2	1	2
300	5 %	100	100	100	100	88	87	88	74	23	24	38	16	6	7	2	12
	1 %	99	99	100	100	68	72	81	51	4	5	23	6	1	2	1	3
400	5 %	100	100	100	100	85	83	95	80	14	14	45	29	4	5	0	6
	1 %	100	100	100	100	58	61	85	56	5	4	28	12	0	0	0	0
500	5 %	100	100	100	100	74	75	88	75	17	18	34	17	10	10	2	9
	1 %	100	100	100	100	56	56	75	46	6	6	21	1	3	3	2	3
600	5 %	100	100	100	100	83	86	86	76	13	14	32	16	4	3	0	6
	1 %	100	100	100	100	55	55	76	51	2	2	19	4	0	0	0	1
700	5 %	100	100	100	100	77	77	86	75	21	23	33	20	10	10	4	13
	1 %	100	100	100	100	46	46	75	59	5	6	18	7	4	4	1	3
800	5 %	100	100	100	100	79	80	89	75	20	19	38	23	12	10	3	8
	1 %	100	100	100	100	52	52	81	54	6	6	20	8	3	3	1	3
900	5 %	100	100	100	100	84	82	87	70	17	17	29	16	12	10	5	8
	1 %	100	100	100	100	56	59	69	45	0	1	13	5	2	2	1	3
1 000	5 %	100	100	100	100	76	79	80	69	15	15	34	21	7	7	3	6
	1 %	100	100	100	99	58	58	64	45	4	4	14	6	3	2	1	2

Colonnes (1) : Pour 100 tableaux observés, nombre de tableaux où l'observation de $\chi^2_{H\cap I.M.} - \chi^2_H$ a conduit à rejeter I.M. à justre titre.

Colonnes (2) : Pour 100 tableaux observés, nombre de tableaux où l'observation de la statistique de BHAPKAR a conduit à rejeter I.M. à juste titre.

Tableau 3

$k = 4 \; ; \; \varepsilon_1 = [4 \; ; \; 0,5 \; ; \; -2,5 \; ; \; -2]$

n	seuil asymptotique α	$\varepsilon_1 ; \lambda_1 = 96,04$				$\varepsilon_2 ; \lambda_2 = 24,01$				$\varepsilon_3 ; \lambda_3 = 3,84$				$\varepsilon_4 ; \lambda_4 = 0,96$			
		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
100	5 %	100	100	100	100	93	97	100	97	35	31	61	34	9	9	39	17
	1 %	100	100	100	100	79	82	94	91	22	21	38	14	1	4	27	5
200	5 %	100	100	100	100	99	98	100	99	23	30	63	40	7	9	37	12
	1 %	100	100	100	100	90	90	98	94	13	15	34	23	1	4	21	3
300	5 %	100	100	100	100	92	94	100	95	25	28	57	36	8	8	31	9
	1 %	100	100	100	100	78	79	96	95	14	17	38	20	3	3	18	3
400	5 %	100	100	100	100	96	99	99	99	27	28	67	43	11	10	28	13
	1 %	100	100	100	100	88	87	97	91	14	16	44	22	3	4	14	4
500	5 %	100	100	100	100	95	96	100	97	30	33	45	26	10	11	31	11
	1 %	100	100	100	100	87	87	98	93	16	16	28	10	2	2	17	7
600	5 %	100	100	100	100	96	95	100	99	31	31	49	26	9	11	37	19
	1 %	100	100	100	100	88	89	97	92	12	14	21	14	1	2	22	3
700	5 %	100	100	100	100	98	98	100	99	33	31	56	35	14	15	31	15
	1 %	100	100	100	100	87	88	100	95	17	17	34	18	5	4	13	5
800	5 %	100	100	100	100	98	98	100	99	27	25	62	43	8	7	29	7
	1 %	100	100	100	100	92	90	97	91	11	13	42	26	3	2	13	3
900	5 %	100	100	100	100	99	99	100	100	21	23	49	33	9	10	34	13
	1 %	100	100	100	100	90	91	99	97	7	7	27	13	4	5	11	6
1 000	5 %	100	100	100	100	92	93	99	98	26	24	54	37	10	12	28	18
	1 %	100	100	100	100	81	83	97	96	11	11	33	15	2	3	14	2

Tableau 4

$k = 5 \; ; \; \varepsilon_1 = [-0,5 \; ; \; -3 \; ; \; 0,5 \; ; \; 5 \; ; \; -2]$

		$\varepsilon_1 \; ; \; \lambda_1 = 145,81$				$\varepsilon_2 \; ; \; \lambda_2 = 36,45$				$\varepsilon_3 \; ; \; \lambda_3 = 5,83$				$\varepsilon_4 = \lambda_4 = 1,45$			
n	seuil asymptotique α	$d_{ij} = 0$		$d_{ij} \neq 0$		$d_{ij} = 0$		$d_{ij} \neq 0$		$d_{ij} = 0$		$d_{ij} \neq 0$		$d_{ij} = 0$		$d_{ij} \neq 0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(1)
100	5 %	100	100	100	100	98	99	97	95	36	38	38	36	10	16	35	16
	1 %	100	100	100	100	93	97	93	86	16	20	25	16	3	5	12	7
200	5 %	100	100	100	100	100	100	98	96	42	44	49	36	9	6	29	17
	1 %	100	100	100	100	95	96	92	90	18	20	39	17	3	1	15	6
300	5 %	100	100	100	100	100	99	98	98	36	40	46	36	17	16	15	12
	1 %	100	100	100	100	96	96	94	93	13	15	19	13	4	4	6	3
400	5 %	100	100	100	100	99	100	100	99	39	39	45	38	10	11	22	18
	1 %	100	100	100	100	98	97	96	94	15	16	29	21	2	2	10	3
500	5 %	100	100	100	100	99	99	100	100	36	41	46	36	9	12	17	12
	1 %	100	100	100	100	96	96	99	96	20	20	22	17	2	3	7	1
600	5 %	100	100	100	100	99	99	100	98	46	44	48	38	17	16	21	12
	1 %	100	100	100	100	98	98	99	96	22	23	25	23	3	4	8	4
700	5 %	100	100	100	100	100	100	100	100	38	41	52	47	14	18	15	9
	1 %	100	100	100	100	98	97	99	98	20	23	29	23	3	2	4	3
800	5 %	100	100	100	100	99	99	100	98	40	40	49	42	16	13	24	20
	1 %	100	100	100	100	97	97	98	96	17	20	25	23	2	4	9	3
900	5 %	100	100	100	100	99	99	98	97	42	40	45	44	11	12	21	15
	1 %	100	100	100	100	95	96	94	95	22	23	29	28	4	6	8	3
1 000	5 %	100	100	100	100	99	99	100	100	52	50	43	38	14	15	14	7
	1 %	100	100	100	100	98	98	97	96	24	24	18	18	2	0	2	1

Tableau 5

$k = 6 \; ; \; \varepsilon_1 = [3,2 \; ; \; -2,8 \; ; \; -1,2 \; ; \; 0,88 \; ; \; 1,52 \; ; \; -1,6]$

n	seuil asymptotique α	$\varepsilon_1 ; \lambda_1 = 147,96$				$\varepsilon_2 ; \lambda_2 = 36,99$				$\varepsilon_3 ; \lambda_3 = 5,91$				$\varepsilon_4 ; \lambda_4 = 1,47$			
		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$		$d^o_{ij} = 0$		$d^o_{ij} \neq 0$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
200	5 %	100	100	100	100	99	98	97	96	27	35	34	29	17	9	26	11
	1 %	100	100	100	100	87	92	96	87	10	10	20	9	3	2	9	6
300	5 %	100	100	100	100	95	96	100	99	27	30	35	27	17	13	24	14
	1 %	100	100	100	100	89	87	97	92	15	12	16	9	5	5	12	4
400	5 %	100	100	100	100	97	96	100	97	26	29	34	34	17	15	15	7
	1 %	100	100	100	100	92	93	95	92	15	14	18	14	4	3	8	0
500	5 %	100	100	100	100	94	94	100	99	32	32	33	35	12	14	22	11
	1 %	100	100	100	100	88	91	100	91	15	13	20	18	6	6	6	6
600	5 %	100	100	100	100	100	100	100	98	38	38	34	27	9	10	16	7
	1 %	100	100	100	100	95	97	97	95	18	18	15	11	1	2	6	1
700	5 %	100	100	100	100	98	98	100	100	35	35	44	44	9	11	14	14
	1 %	100	100	100	100	94	94	100	95	19	21	19	16	4	3	4	3
800	5 %	100	100	100	100	100	100	99	98	31	32	30	30	12	7	14	14
	1 %	100	100	100	100	95	95	99	97	16	16	13	15	2	1	6	2
900	5 %	100	100	100	100	99	98	100	100	48	43	32	29	11	8	9	8
	1 %	100	100	100	100	93	96	100	97	20	20	18	13	3	1	4	3
1 000	5 %	100	100	100	100	98	98	100	98	38	41	46	39	17	12	13	9
	1 %	100	100	100	100	90	92	98	95	19	21	23	21	4	3	6	3