Revue de statistique appliquée

Dembo Gadiaga
Rosaria Ignaccolo

Tests du χ^2 généralisés. Application au processus autorégressif

Revue de statistique appliquée, tome 53, n° 2 (2005), p. 67-84
<http://www.numdam.org/item?id=RSA_2005__53_2_67_0>

© Société française de statistique, 2005, 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/

Rev. Statistique Appliquée, 2005, **LIII** (2), 67-84

TESTS DU χ^2 GÉNÉRALISÉS APPLICATION AU PROCESSUS AUTORÉGRESSIF

Dembo GADIAGA[(1)], Rosaria IGNACCOLO[(2)]

(1) UFR/SEG & UFR/SEA (Département de Mathématique et Informatique)
Université de OUAGADOUGOU 03 B.P. 7164 Ouagadougou 03, Burkina Faso
Email : gdembo@hotmail.com
(2) Dipartimento di Statistica e Matematica Applicata
Università degli Studi di TORINO Italy
Email : ignaccolo@econ.unito.it

RÉSUMÉ

Le test du χ^2 classique fait partie d'une classe de tests d'adéquation basés sur des systèmes orthogonaux. Des considérations théoriques montrent que l'utilisation de systèmes orthogonaux lisses permet de construire des tests plus efficaces que le test usuel. Des simulations sont faites sur un processus autorégressif.

Mots-clés : *Estimations, tests d'adéquation et test du χ^2, processus autoregressif, simulations.*

ABSTRACT

The classical χ^2 test belongs to a family of goodness of fit tests based upon orthogonal systems; some theoretical results show that use of smooth orthogonal systems leads to more efficient tests than the PEARSON χ^2 test. We give some simulations using an autoregressive process.

Keywords : *Estimations, Goodness-of-fit tests, χ^2 test, autoregressive process, simulation.*

1. Introduction

Une manière naturelle de construire un test d'adéquation consiste à utiliser un estimateur de la loi dont on teste l'ajustement. Le calcul de la distance de la loi présumée à son estimée permettra d'apprécier la validité de l'hypothèse de départ.

Ainsi, le célèbre test du χ^2 est fondé sur un estimateur élémentaire de la densité : l'histogramme des fréquences. Or, il est bien connu que les estimateurs lisses de la densité approchent la loi des observations beaucoup mieux que l'histogramme.

On trouve des résultats théoriques sur les tests basés sur les estimateurs par projection («tests hilbertiens») dans Bosq (1980,1983,1989, 2000), Gadiaga (1983, 2003), Blacher (1985).

Dans [5] Bosq a montré en faisant des comparaisons empiriques (par simulations) que le test de Legendre est meilleur que le test du χ^2 si les variables aléatoires utilisées dans la statistique hilbertienne sont indépendantes et suivent la même loi.

Le présent travail est consacré au test statistique de l'hypothèse nulle : à savoir que les données suivent une loi fixée que nous noterons par μ. Pour ce faire, nous nous plaçons dans le cas où les observations forment un processus stationnaire et mélangeant. Après quelques rappels sur les tests hilbertiens en dimension finie, nous dégageons la loi limite de la statistique considérée. Les éléments de la matrice des variances covariances sont estimés, enfin des simulations sont faites.

2. Les données du problème

Considérons une suite de v.a. $(X_n)_{n\geqslant 1}$ définie sur un espace probabilisé (Ω, A, P) à valeurs dans un espace mesurable (E, B) où E est l'espace des observations, B une tribu sur E et soit μ une probabilité sur (E, B); supposons que l'espace de Hilbert $L^2(\mu)$ est séparable.

En outre supposons que la suite $(X_n)_{n\geqslant 1}$ est :

1) strictement stationnaire,
2) fortement mélangeante.

Nous entendons par processus strictement stationnaire, une suite $(X_n)_{n\geqslant 1}$ telle que : la distribution des vecteurs aléatoires $(X_m, \dots, X_{m+n})$ soit indépendante de m, pour tout m de $\mathbb{N}^*$, voir [8] p.94.

Comme [16] p.156, ou encore [13], nous appelons processus fortement mélangeant, de coefficient de mélangeance α (on dira α-fortement mélangeant) un processus vérifiant :

$$\forall\, n \in \mathbb{N}^*, \forall\, A \in M_1^t, \forall\, B \in M_{t+n}^{\infty} \quad |P(A \cap B) - P(A)P(B)| \leqslant \alpha(n)$$

où M_1^t désigne la σ-algèbre engendrée par $X_1, \dots, X_t$ et M_{t+n}^{∞} celle qui l'est par les variables $X_{t+n}, \dots$; tandis que $\lim\limits_{n\to+\infty} \alpha(n) = 0$, où $\alpha(n)$ est une suite décroissante de nombres réels positifs.

Soit $(e_1, \dots, e_k)$ avec $e_1 \equiv 1$ une base orthonormale d'un sous-espace Hilbertien E_k $(k \geqslant 2)$ de $L^2(\mu)$, nous désignerons par $K(\cdot, \cdot)$ le noyau reproduisant de l'espace E_k engendré par $(e_1, \dots, e_k)$; alors $K(x, t)$ s'écrit :

$$K(x,t) = \sum_{j=1}^{k} e_j(x)e_j(t) \qquad \text{avec } (x,t) \in E \times E$$

Considérons maintenant la statistique à valeurs dans $L^2(\mu)$ définie par :

$$T_n(\cdot) = \sum_{i=1}^{n} K(X_i, \cdot) = n \left[1 + \sum_{j=2}^{k} \widehat{a}_{j_n} e_j(\cdot) \right]$$

où

$$\widehat{a}_{j_n} = \frac{1}{n}\sum_{i=1}^{n} e_j(X_i)\,; \qquad j = 2, \ldots, k.$$

Posons $Sn(\cdot) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} K(X_i, \cdot) - 1\right]$

L'objet de ce travail consiste à tester l'hypothèse H_0 «la loi commune des X_i est μ avec $\forall\, j = 2, \ldots, k$: $\int_E e_j \mathrm{d}\,\mu = 0$».

Contre l'hypothèse alternative H_1 : «la loi commune des X_i est ν, tel qu'il existe au moins un $j_0 \in \{2, \ldots, k\}$: $\int_E e_{j_0} \mathrm{d}\,\nu \neq 0$».

Sous H_0, la statistique $S_n(\cdot)$ définie ci-dessus est égale à :

$$S_n(\cdot) = \frac{1}{\sqrt{n}}\,[T_n(\cdot) - E[T_n(\cdot)]]$$

En effet,

$$\forall\, t \in E,\ E[T_n(t)] = E\left[\sum_{i=1}^{n} K(X_i, t)\right]$$

Or

$$E[T_n(t)] = \int_E \sum_{i=1}^{n} K(x_i, t)\mathrm{d}\,\mu(x) \text{ sous } H_0$$

$$= \int_E nK(x, t)\mathrm{d}\,\mu(x)$$

sous H_0 (car les X_i sont de même loi)

$$\int_E K(x, t)\mathrm{d}\,\mu(x) = \langle K(\cdot, t), 1(\cdot)\rangle = 1$$

Par suite

$$E[T_n(t)] = n$$

et donc ,

$$S_n(\cdot) = \frac{1}{\sqrt{n}}[T_n(\cdot) - E[T_n(\cdot)]]$$

$$= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} K(Xi, 1) - 1\right]$$

Remarques. –

1) $E[e_j(X_i)] = 0$ pour tout $j = 2, \ldots, k$. En effet, $E[e_j(X_i)] = \int_E e_j(x) \mathrm{d}\,\mu(x) = \int_E 1 \cdot e_j(x) \mathrm{d}\,\mu(x) = 0$; $j = 2, \ldots, k$ car $e_1 = 1, e_2, \ldots, e_k$ constituent un système orthonormal .
2) l'alternative définie ci-dessus paraît artificielle mais elle peut s'écrire d'une façon naturelle dans les cas usuels. Si on se propose de tester $f = 1$ contre $f = 1 + \sum_{i=1}^{q} \alpha_i f_i$ avec au moins un $\alpha_i \neq 0$ pour $i > 1$ (les hypothèses sont écrites avec des densités), où les densités f_i sont dans $L^2(\mu)$, il suffit alors de construire une base orthonormale de l'espace vectoriel engendré par $(1, f_1, \ldots, f_q)$ pour obtenir un test de H_0 contre H_1.

   Sous certaines conditions assez générales, nous avons obtenu dans [9] p.17-18 la loi limite de $||Sn(\cdot)||^2$. Nous avons en particulier obtenu

PROPOSITION 1. – *Sous les conditions suivantes :*

1) La loi des X_i est μ

2) $\sup_{x \in E} \sup_{j \in \{2,\ldots,k\}} |e_j(x)| = M < +\infty$

3) $\sum_{i \geqslant 1}^{\infty} \alpha(i) < +\infty$

alors [1] $||Sn(\cdot)||^2 \xrightarrow{L} Q = \sum_{j=2}^{k} Z_j^2$ *où Q est une variable aléatoire dont la fonction caractéristique est donnée par* $[\det(I_{k-1} - 2it\Sigma)]^{1/2}$*; I_{k-1} est la matrice identité dans $\mathbb{R}^{k-1}$ et $\Sigma = (\sigma_{jj'})$, $2 \leqslant j, j' \leqslant k$ avec $\sigma_j^2 < +\infty$ et $|\sigma_{jj'}| < +\infty$ est definie par*

$$\sigma_{jj'} = \begin{cases} \sigma_j^2 = 1 + 2 \sum_{\nu=1}^{\infty} E[e_j(x_1) e_j(x_{\nu+1})], & \textit{si } j = j' \\ \sigma_{jj'} = \sum_{\nu=1}^{\infty} E[e_j(x_1) e_{j'}(x_{\nu+1}) + e_j(x_{\nu+1}) e_{j'}(x_1)], & \textit{si } j \neq j' \end{cases}$$

PROPOSITION 2. –

1) si la loi de probabilité des X_n est ν telle qu'il existe $j \in \{2, \ldots, k\}$ pour lequel e_j soit ν-intégrale et d'intégrale non nulle

2) si $\sup_{x \in E} \sup_{j \in \{2,\ldots,k\}} |e_j(x)| = M < +\infty$

[1] L désigne la convergence en loi

3) si $\sum_{i=1}^{\infty} \sqrt{\alpha(i)} < +\infty$

alors

$$\exists\, j_0 \in \{2, \ldots, k\} : \ \frac{1}{n} \sum_{i=1}^{n} e_{j_0}(X_i) \xrightarrow[n\to+\infty]{\text{p.s.}} E[e_{j_0}(X_1)] \neq 0$$

de plus $||Sn(\cdot)||^2 \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty$

Remarques. –

1) on peut donc définir un test de H_0 contre H_1 en se donnant une suite (w_n) de nombres réels positifs en choisissant comme région critique $\{(x_1, \ldots, x_n) \ : \ ||Sn(\cdot)||^2 > w_n\}$
2) le test du χ^2 comme test hilbertien.
   Le test du χ^2 est un cas particulier du test précédent; en effet, on peut le construire de la façon suivante : on se donne $\{A_1, A_2, \ldots, A_k\}$ une partition de E telle que $p_j = \mu(A_j) > 0; j = 1, \ldots, k$.
   Alors $\{f_j = p_j^{-1/2} \mathbb{1}_{A_j} \ ; \ j = 1, 2, \ldots, k\}$ est un système orthonormal dans $L^2(\mu)$ qui engendre un espace vectoriel de dimension k.
   Le noyau de cet espace est donné par la formule

$$K(x,t) = \sum_{j=1}^{k} p_j^{-1} \mathbb{1}_{A_j}(x) \mathbb{1}_{A_j}(t) \ ; \ (x,t) \in E \times E$$

on a donc

$$\begin{aligned}\sum_{i=1}^{n} K(x_i, \cdot) - n &= \sum_{i=1}^{n} \sum_{j=1}^{k} p_j^{-1} \mathbb{1}_{A_j}(X_i) \cdot \mathbb{1}_{A_j}(\cdot) - \sum_{j=1}^{k} (np_j^{1/2}) \mathbb{1}_{A_j}(\cdot) p_j^{-1/2} \\ &= \sum_{j=1}^{k} \left[(1/\sqrt{p_j}) \sum_{i=1}^{n} \mathbb{1}_{A_j}(X_i) - n\sqrt{p_j} \right] \mathbb{1}_{A_j}(\cdot)/\sqrt{p_j}\end{aligned}$$

d'où

$$\begin{aligned}||Sn(\cdot)||^2 &= n^{-1} || \sum_{i=1}^{n} K(X_i, \cdot) - n||^2 \\ &= n^{-1} \sum_{j=1}^{k} \left[\frac{1}{\sqrt{p_j}} \sum_{i=1}^{n} \mathbb{1}_{A_j}(X_i) - n\sqrt{p_j} \right]^2 \\ &= \sum_{j=1}^{k} \frac{1}{np_j} \left[\sum_{i=1}^{n} \mathbb{1}_{A_j}(X_i) - np_j \right]^2\end{aligned}$$

C'est bien la statistique utilisée dans le test du χ^2 .

3) dans la pratique les éléments de la matrice Σ étant inconnus, nous allons les estimer pour faire les simulations.

3. Estimation de $(\sigma_{jj'})$, $2 \leqslant j, j' \leqslant k$

3.1. Notations

Posons :

$$\varepsilon_\nu = \begin{cases} \frac{1}{2} & \text{si } \nu = 0 \\ 1 & \text{si } \nu > 0 \end{cases}$$

$$C_\nu^j = \frac{1}{n} \sum_{h=1}^{n-\nu} e_j(X_h) e_j(X_{h+\nu}) \ ; \ 0 \leqslant \nu \leqslant n-1$$

$$C_\nu^{jj'} = \frac{1}{n} \sum_{h=1}^{n-\nu} e_j(X_h) e_{j'}(X_{h+\nu}) \ ; \ 0 \leqslant \nu \leqslant n-1$$

$$C_\nu^{j'j} = \frac{1}{n} \sum_{h=1}^{n-\nu} e_{j'}(X_h) e_j(X_{h+\nu}) \ ; \ 0 \leqslant \nu \leqslant n-1$$

3.2. Proposition 3

Sous les conditions suivantes :

a) $\sup_{x \in E} \sup_{j \in \{2,\ldots,k\}} |e_j(x)| = M < +\infty$

b) $\sum_{i=1}^{\infty} \sqrt{\alpha(i)} < +\infty$

c) $\lim_{n \to -\infty} d_n = +\infty$ et $\sum_{n \geqslant 1} \frac{d_n^5}{n^2} < +\infty$

d) $\lim_{n \to -\infty} m_n = +\infty$ et $\sum_{n \geqslant 1} \frac{m_n^5}{n^2} < +\infty$

alors

1) $\widehat{\sigma}_j^2 = 2 \sum_{\nu=0}^{d_n} \varepsilon_\nu C_\nu^j$ est un estimateur convergent presque sûrement et en moyenne quadratique de σ_j^2. De plus il est sans biais.

2) $\widehat{\sigma}_{jj'} = \sum_{\nu=1}^{m_n}(C_\nu^{jj'} + C_\nu^{j'j})$ est un estimateur convergent presque sûrement et asymptotiquement sans biais de $\sigma_{jj'}$.

Démonstration. – Ces résultats sont tout à fait analogues à ceux obtenus par Anderson (p.508-518), Grenander et Rosenblatt (p.145-153) et Carbon M. (p.65-68).

Remarques. –

a) pour l'estimateur de $\sigma_{jj'}$ nous avons utilisé le périodogramme « tronqué en d_n » pour σ_j^2 et celui « tronqué en m_n » pour $\sigma_{jj'}$ pour $j \neq j'$.

b) la matrice Σ a été donc estimée par $\widehat{\Sigma}_n$ où $\widehat{\Sigma}_n = (\widehat{\sigma}_{jj'})_{2\leqslant j,j'\leqslant k}$ et on a alors $\widehat{\Sigma}_n \xrightarrow[n\to+\infty]{\text{p.s.}} \Sigma$.

c) La matrice $\widehat{\Sigma}_n$ obtenue est symétrique car $\widehat{\sigma}_{jj'} = \widehat{\sigma}_{j'j}$

d) $\widehat{\Sigma}_n \xrightarrow[n\to+\infty]{\text{p.s.}} \Sigma$ et comme Σ est définie positive alors $\widehat{\Sigma}_n$ est presque sûrement asymptotiquement définie positive.

4. Approximation de la loi limite de $||S_n(\cdot)||^2$

4.1. Loi limite de $||S_n(\cdot)||^2$ en fonction des valeurs propres de Σ

La matrice Σ étant définie positive alors ses valeurs propres que nous noterons par λ_j sont telles que : pour tout $j \in \{2,\dots,k\},\ \lambda_j > 0$. Nous obtenons le résultat suivant :

PROPOSITION 4. – *Sous les conditions de la proposition 1, Σ étant définie positive :*
Alors

$$||Sn(\cdot)||^2 \xrightarrow[n\to+\infty]{L} \sum_{j=2}^{k} \lambda_j U_j^2$$

où les v.a. $Uj \sim N(0,1),\ j = 2,\dots,k$ sont normales centrées réduites et indépendantes.

Démonstration. – Elle est analogue à celle donnée dans [12] p.11-13.

Les valeurs propres λ_j étant inconnues, pour la réalisation de la simulation nous les estimons à partir de la matrice $\widehat{\Sigma}_n$.

4.2. Estimation des valeurs propres λ_j, $j = 2,\dots,k$

Nous utiliserons les résultats [6] p.102-104 pour l'estimation des λ_j, $j = 2,\dots,k$.

On y établit que si C est un opérateur de covariance et C_n son estimateur empirique; λ_j, $j \geqslant 1$ les valeurs propres de C c'est-à-dire $C(v_j) = \lambda_j v_j$, où $\{v_j, j \geqslant 1\}$ est un système orthonormal constitué par les vecteurs propres de C; alors les λ_{jn}, $j \geqslant 1$ qui sont des estimateurs empiriques de λ_j ,$j \geqslant 1$ sont données par :

$$C_n(v_{jn}) = \lambda_{jn} v_{jn}, \quad j \geqslant 1$$

Les λ_{jn}, $j \geqslant 1$ sont donc les valeurs propres de C_n associées aux vecteurs propres v_{jn}, $j \geqslant 1$, qui constituent un système orthonormal.

On montre en particulier que si :

$$\lambda_{1n} \geqslant \lambda_{2n} \geqslant \ldots \geqslant \lambda_{nn} \geqslant 0 = \lambda_{n+1,n} = \lambda_{n+2,n} = \ldots$$

alors, pour les opérateurs compacts, on a :

$$\sup_{j \geqslant 1} |\lambda_{jn} - \lambda_j| \leqslant ||C_n - C||$$

Nous appliquons ces résultats à notre situation.

En effet nous travaillons dans $\mathbb{R}^{k-1}$ (nous avons des vecteurs à $(k-1)$ dimensions) et les matrices $\widehat{\Sigma}_n$ et Σ sont bornées ($\forall\, j, j' = 2, \ldots, k$; $\sigma_j^2 < +\infty$ et $|\sigma_{jj'}| < +\infty$; $\widehat{\sigma}_{jj'}$ et $\widehat{\sigma}_j^2$ sont des estimateurs convergents de $\sigma_{jj'}$ et σ_j^2), on peut donc utiliser les résultats de [6].

Désignons par λ_j; $j = 2, \ldots, k$ les valeurs propres de la matrice Σ, comme Σ est supposée être définie positive alors les valeurs propres sont toutes positives $(\forall\, j, j' = 2, \ldots, k : 1\, \lambda_j > 0)$.

Soient λ_{jn} , $j = 2, \ldots, k$ les valeurs propres de $\widehat{\Sigma}_n$ ($\widehat{\Sigma}_n$ étant symétrique ses valeurs propres sont réelles). Par ailleurs, on a montré que $\widehat{\Sigma}_n \xrightarrow[n \to +\infty]{\text{p.s.}} \Sigma$, donc $\widehat{\Sigma}_n$ est presque sûrement asymptotiquement définie positive; l'on déduit que ses valeurs propres λ_{jn}, $j = 2, \ldots, k$ sont presque sûrement asymptotiquement positives.

Par conséquent : $\forall\, j = 2, \ldots, k$; $\exists\, N_j \in \mathbb{N}$, $\forall\, n \geqslant N_j$; $\lambda_{jn} \geqslant 0$ p.s. Posons

$$N = \max_{j=2,\ldots,k} (N_j).$$

Alors $\exists\, N \in \mathbb{N}, \forall\, j = 2, \ldots, k, \forall\, n \in \mathbb{N}, n \geqslant N, \lambda_{jn} \geqslant 0$p.s.

On déduit qu'à partir de ce rang N , on a :

$$\sup_{j=2,\ldots,k} |\lambda_{jn} - \lambda_j| \leqslant ||\widehat{\Sigma}_n - \Sigma||$$

Or $||\widehat{\Sigma}_n - \Sigma|| \xrightarrow[n \to +\infty]{\text{p.s.}} 0$.

Par suite $\sup |\lambda_{jn} - \lambda_j| \xrightarrow[n \to +\infty]{\text{p.s.}} 0$ et donc $\lambda_{jn} \xrightarrow[n \to +\infty]{\text{p.s.}} \lambda_j$.

D'où λ_{jn} est un estimateur convergent presque sûrement de λ_j pour tout $j = 2, \ldots, k$.

4.3. Approximation de la loi limite de $||S_n(\cdot)||^2$

$\widehat{\Sigma}_n$ étant symétrique, est diagonalisable et est semblable à la matrice $\widehat{D}_n$ définie par :

$$\widehat{D}_n = \begin{pmatrix} \lambda_{2n} & & \\ & \ddots & 0 \\ 0 & & \lambda_{kn} \end{pmatrix} \xrightarrow[n \to +\infty]{\text{p.s.}} D = \begin{pmatrix} \lambda_2 & & \\ & \ddots & 0 \\ 0 & & \lambda_k \end{pmatrix}$$

Remarquons que Σ est semblable à D.

Posons maintenant :

$$\forall\, j = 2, \ldots, k\,; \forall\, n \geqslant 1\,; \lambda^*_{jn} = \max(0, \lambda_{jn})$$

On a $\lambda^*_{jn} \xrightarrow[n \to +\infty]{\text{p.s.}} \lambda_j$ et $\forall\, n \geqslant 1;\ \lambda^*_{jn} \geqslant 0$ donc :

$$\widehat{D}^*_n = \begin{pmatrix} \lambda^*_{2n} & & \\ & \ddots & 0 \\ 0 & & \lambda^*_{kn} \end{pmatrix} \xrightarrow[n \to +\infty]{\text{p.s.}} D = \begin{pmatrix} \lambda_2 & & \\ & \ddots & 0 \\ 0 & & \lambda_k \end{pmatrix}$$

Nous utiliserons les estimateurs λ^*_{jn} pour approximer la loi limite de $||S_n(\cdot)||^2$.

PROPOSITION 5. – *Sous les conditions des propositions 1 et 3 et si* Σ *est définie positive, alors*

$$\sum_{j=2}^{k} \lambda^*_{jn} U_j^2 \xrightarrow[n \to +\infty]{\text{p.s.}} \sum_{j=2}^{k} \lambda_j U_j^2$$

où les U_j, $j = 2, \ldots, k$ *suivent des lois normales centrées réduites et indépendantes.*

Démonstration. – Puisque $\lambda^*_{jn} \xrightarrow[n \to +\infty]{\text{p.s.}} \lambda_j$ pour tout $j = 2, \ldots, k$ alors

$$\sum_{j=2}^{k} \lambda^*_{jn} U_j^2 \xrightarrow[n \to +\infty]{\text{p.s.}} \sum_{j=2}^{k} \lambda_j U_j^2$$

Remarque. – De la proposition ci-dessus on déduit donc que pour n assez grand la variable aléatoire $\widehat{U}_n^2 = \sum_{j=2}^{k} \lambda_{jn}^* U_j^2$ approxime presque sûrement la variable aléatoire $\sum_{j=2}^{k} \lambda_j U_j^2$ qui est la loi limite de $||S_n(\cdot)||^2$.

$\widehat{U}_n^2$ sera utilisée dans les simulations pour tester l'hypothèse nulle H_0.

5. Simulation

5.1. Introduction

On va réaliser une simulation sur un autorégressif d'ordre 1. On pose donc

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{1}$$

avec $|\rho| < 1$, $(\varepsilon_t)_{t \geqslant 1}$ étant une suite de variables aléatoires indépendantes et équidistribuées de loi normale $N(0,1)$.

On teste l'hypothèse : X_t suit la loi normale $N(0,\sigma^2)$, $\forall\, t$

$$(1) \Longrightarrow \sigma^2 = V(X_t) = \frac{V(\varepsilon_t)}{1-\rho^2}.$$

Comme ε_t suit la loi normale $N(0,1)$ alors $\sigma^2 = \dfrac{1}{1-\rho^2}$.

D'où $\rho = 0 \Longleftrightarrow \sigma^2 = 1$, c'est-à-dire X_t suit une loi normale $N(0,1)$.

L'hypothèse nulle H_0 peut s'écrire « $\rho = 0$ » ou encore « les variables aléatoires X_t sont indépendantes et de même loi $N(0,1)$ ».

On construit un processus générant N variables aléatoires indépendantes ε_t suivant la loi $N(0,1)$ et en utilisant l'équation (1) on calcule X_t, X_0 étant donné.

Nous avons choisi plusieurs valeurs de ρ pour générer le processus $(X_t)_{t\geqslant 1}$. La matrice $\widehat{\Sigma}_n$ étant fonction du système orthogonal considéré, nous avons choisi pour les calculs le système orthonormal constitué par les polynômes de Legendre.

À partir de $\widehat{\Sigma}_n$, on obtient les valeurs propres λ_{jn} et par suite λ_{jn}^*, $j = 2, \ldots, k$ ce qui permettra d'avoir $\widehat{U}_n^2$ en fonction de k.

Pour effectuer le test de Legendre (le test de Legendre s'obtient en remplaçant dans la statistique hilbertienne les e_j, $j = 1, \ldots, k$ par les polynômes de Legendre) on commence par amener les observations sur $[-1,1]$ en posant :

$$Y_i = 2F(X_i) - 1 \quad ; \quad i = 1, 2, \ldots, n$$

où F désigne la fonction de répartition de la loi des X_i.

Sous H_0 les Y_i suivent la loi uniforme U sur $[-1, 1]$.

Le système (e_j) est alors obtenu en orthonormalisant par rapport à U les polynômes $1, x, x^2, \ldots$ et on obtient pour tout j

$$e_j(x) = (1 + 2j)^{1/2} P_j(x) , \quad x \in [-1, 1]$$

où

$$P_j(x) = \frac{1}{2^j j!} \frac{\mathrm{d}^j}{\mathrm{d}\, x^j} (x^2 - 1)^j , \quad x \in [-1, 1]$$

Les quantités d_n et m_n utilisées dans l'estimation de la matrice Σ sont égales à $10 \cdot \log(n)$, donc dépendent seulement de la taille n de l'échantillon.

5.2. Test d'hypothèse

On va considérer 6 cas en choisissant des valeurs de ρ comme suit :

$$\rho = 0,20 \; ; \qquad \rho = 0,50 \; ; \qquad \rho = 0,70$$

$$\rho = -0,20 \; ; \quad \rho = -0,50 \; ; \quad \rho = -0,70$$

car de faibles valeurs de ρ ne seraient pas intéressantes ici pour le test d'adéquation, car on accepterait presque tout le temps l'hypothèse nulle.

On obtiendra à travers ces 6 cas la région critique du test ainsi que sa puissance empirique en fonction de la dimension k .

On se donne α le risque de première espèce. Cette donnée permet de déterminer la région de rejet de l'hypothèse nulle H_0 : «$\rho = 0$». Cette région critique est donnée par :

$$\{(x_1, \ldots, x_n) \ : \ ||Sn(\cdot)||^2 \geqslant W_n\}$$

W_n étant déterminé par

$$\alpha = P \left[\sum_{j=2}^{k} \lambda_{jn}^* U_j^2 \geqslant W_n / \rho = 0 \right]$$

puisque pour n assez grand, $\sum_{j=2}^{k} \lambda_j^* U_j^2$ approxime presque sûrement la variable aléatoire $\sum_{j=2}^{k} \lambda_j U_j^2$ qui est la loi limite de $||Sn(\cdot)||^2$.

Si β est la puissance du test, on a :

$$\beta = P_\nu \left[||Sn(\cdot)||^2 \geqslant W_n \right]$$

où $\nu \in H_1$: «$\rho \neq 0$».

Nous comparons ci-après pour différentes valeurs de ρ, la puissance du test de Legendre défini ci-dessus, avec celle du test du chi-2 classique de l'hypothèse nulle H_0 : X_t suit une loi normale $N(0,1)$ contre H_1 : X_t ne suit pas la loi $N(0,1)$.

Nous présentons ici les puissances empiriques du test obtenues comme pourcentage de rejet sur $m = 1\,000$ simulations. Les tableaux ci-dessous donnent les puissances empiriques pour $1\,000$ simulations, en faisant varier ν et ρ et pour $\alpha = 5$ %. Les colonnes représentent :

– k la dimension du noyau K
– χ^2 : test de Pearson pour des observations fortement mélangeantes
– leg : test avec les polynômes de Legendre pour des observations fortement mélangeantes.

Les tableaux 4.1 à 4.10 sont présentés en appendice comme illustration.

Nous poursuivrons l'étude des simulations pour l'évaluation de la puissance empirique du test en faisant varier la taille de l'échantillon et le niveau du test. Dans tous les tableaux, les valeurs en gras indiquent la puissance la plus grande lorsque k varie.

5.3. Commentaire des résultats

On constate de manière générale que, pour k fixé et $|\rho| \geqslant 0,5$, le test de Legendre a une puissance empirique plus grande que le test de Pearson.

On observe que la puissance maximale du test de Legendre est obtenue pour des valeurs de k plus petites que celle du test de Pearson.

a) Pour les petites valeurs de ρ ($\rho = \pm 0,2$) la puissance du test du χ^2 est meilleure que celle du test de Legendre, mais cette puissance est très faible, ce qui montre que le test n'est pas très bon. Ceci est dû au fait que la variance σ^2 de X_n est très proche de 1 $\left(\sigma^2 = \dfrac{1}{1-0,04} \cong 1,04\right)$ et donc l'hypothèse nulle est très proche de l'hypothèse alternative. Alors la puissance empirique tend vers le niveau quand n augmente et ainsi avec $n = 250$ (tableau 4) la puissance est plus petite que pour $n = 100$. Cette situation ne se présente pas pour $\rho = 0,5$ et $\rho = 0,7$, où la puissance empirique augmente beaucoup en passant de $n = 100$ à $n = 250$, soit pour le test de χ^2, soit pour le test de Legendre (voir tableaux 5 et 6, 8 et 9).

b) Pour les grandes valeurs de ρ ($\rho = \pm 0,5$; $\rho = \pm 0,7$) la puissance du test de Legendre est plus grande que celle de la puissance du test du χ^2. Cette puissance est maximale pour $k = 2,3,4$, alors que la puissance maximale du test du χ^2 est obtenue pour des valeurs de k plus grandes (pour $\rho = 0,7$ cette puissance maximale est obtenue pour $k = 10$).

   Il se dégage ainsi que la puissance maximale du test de Legendre est obtenue dès les premières valeurs de k.

c) En comparant les deux puissances de test, on observe que pour $\rho > 0$ ($\rho = 0,5$ et $\rho = 0,7$) le test du χ^2 ne se comporte mieux que le test de Legendre que pour des valeurs de $k \geqslant 9$. Sinon le test de Legendre est meilleur.

Il serait intéressant d'étudier la robustesse des tests hilbertiens usuels, quand on teste l'hypothèse de la non corrélation d'une suite de variables aléatoires. Une étude de ce type a été menée par R. Ignaccolo [14] pp. 106-114.

5.4. Appendice : Tableaux des puissances empiriques

TABLEAU 1

Puissances empiriques pour $\rho = 0,2$ et $n = 100$

k	χ^2	leg
2	**10,1**	**9,7**
3	8,9	8,3
4	8,1	7,2
5	8,4	6,5
6	7,8	6,5
7	7,4	6,2
8	6,2	6,2
9	6,2	5,9
10	6,0	5,3

TABLEAU 2

Puissances empiriques pour $\rho = -0,2$ et $n = 100$

k	χ^2	leg
2	**9,0**	**9,9**
3	8,9	8,0
4	8,8	7,1
5	7,4	7,3
6	7,7	7,3
7	6,7	6,8
8	5,6	6,8
9	6,0	5,4
10	6,1	4,2

TABLEAU 3

Puissances empiriques pour $\rho = 0,2$ *et* $n = 50$

k	χ^2	leg
2	12,1	13,8
3	12,3	**14,2**
4	**12,5**	12,2
5	**12,5**	11,8
6	10,9	10,7
7	10,2	9,9
8	10,2	9,1
9	8,0	8,1
10	7,9	7,5

TABLEAU 4

Puissances empiriques pour $\rho = 0,2$ *et* $n = 250$

k	χ^2	leg
2	**7,4**	**7,8**
3	7,3	6,7
4	6,2	6,3
5	6,2	6,3
6	6,9	5,6
7	5,2	5,7
8	6,1	4,9
9	5,2	4,9
10	5,3	4,8

TABLEAU 5

Puissances empiriques pour $\rho = 0,5$ et $n = 100$

k	χ^2	leg
2	10,8	**21,3**
3	13,4	17,5
4	**15,6**	16,5
5	14,6	16,0
6	13,5	14,9
7	14,4	14,6
8	13,1	14,0
9	13,3	12,3
10	13,7	11,1

TABLEAU 6

Puissances empiriques pour $\rho = 0,5$ et $n = 250$

k	χ^2	leg
2	21,7	**50,3**
3	27,8	44,4
4	30,0	49,8
5	**33,8**	44,4
6	33,7	45,4
7	32,3	42,4
8	33,4	42,5
9	33,6	39,8
10	30,5	38,0

TABLEAU 7
Puissances empiriques pour $\rho = -0,5$ *et* $n = 100$

k	χ^2	leg
2	18,0	**34,1**
3	**21,0**	30,1
4	20,3	29,9
5	19,7	27,1
6	20,4	24,9
7	19,2	22,7
8	17,8	21,7
9	18,1	19,3
10	18,2	18,3

TABLEAU 8
Puissances empiriques pour $\rho = 0,7$ *et* $n = 100$

k	χ^2	leg
2	20,9	**47,9**
3	25,5	33,8
4	29,9	43,6
5	29,5	35,1
6	31,9	39,8
7	32,9	34,3
8	32,7	35.5
9	33,2	30,7
10	**33,3**	31,9

TABLEAU 9

Puissances empiriques pour $\rho = 0,7$ et $n = 250$

k	χ^2	leg
2	45,7	94,4
3	64,5	88,1
4	72,9	**94,6**
5	79,4	91,7
6	82,3	93,7
7	84,5	92,3
8	84,9	92,9
9	86,3	91,9
10	**87,4**	91,9

TABLEAU 10

Puissances empiriques pour $\rho = -0,7$ et $n = 100$

k	χ^2	leg
2	47,8	**72,6**
3	54,4	69,8
4	59,3	71,2
5	**60,9**	69,6
6	58,4	68,0
7	60,8	66,2
8	60,4	63,1
9	59,6	60,9
10	59,1	57,8

Remerciements

Nous remercions le referee qui, par ses observations, nous a permis d'améliorer la rédaction du manuscrit.

Références

[1] ANDERSON T.W. (1971), *The statistical analysis of time series*. J.Wiley, New-York.

[2] BLACHER R. (1985), *A new chi-squared independence test*. Rapport de recherché n° 512 PMAG (Grenoble).

[3] BOSQ D. (1980), *Sur une classe de tests qui contient le test du* χ^2. Publication de l'I.S.U.P. Volume 25 Fasc .1–2 p.1-16.

[4] BOSQ D. (1983), Lois limites et efficacité asymptotique des tests hilbertiens de dimension finie sous des hypothèses adjacentes. *Statistiques et Analyse de données*, Vol 8 n°1 pp.1-40

[5] BOSQ D. (1989), Tests du χ^2 Généralisés. Comparaison avec le test du χ^2 classique. *Rev. Statist. Appliquée*. Vol 27 n°1 p. 43–52.

[6] BOSQ D. (2000), *Linear processes in function spaces. Theory and applications*. Lecture notes in statistics n°149. Springer-Verlag, New-York.

[7] CARBON M. (1982), *Sur l'estimation asymptotique d'une classe de paramètres fonctionnels pour un processus stationnaire*. Thèse 3^{e} cycle, Lille I.

[8] DOOB J.L. (1967), *Stochastic processes*. Wiley, New-York

[9] GADIAGA D. (1982), *Sur une classe de tests qui contient le test du* χ^2 *le cas d'un processus stationnaire*. Thèse 3ème cycle Lille I.

[10] GADIAGA D. (1983), Tests hilbertiens et test du χ^2 pour un processus stationnaire et mélangeant. *CRAS* t.296 p.171–174.

[11] GADIAGA D. (2003), *Tests Fonctionnels d'Ajustement et de Non Influence pour des Variables Aléatoires Dépendantes*. Thèse d'État – Université de Ouagadougou – Novembre 2003.

[12] GADIAGA D. et IGNACCOLO R. (2002), *Tests de non influence*. Preprint L.S.T.A., Université Paris 6.

[13] GRENANDER U., ROSENBLATT R. (1957), *Statistical analysis of stationnary time series*. Wiley, New-York.

[14] IGNACCOLO R. (2002), *Tests d'Ajustement Fonctionnels pour des observations corrélées*. Thèse de Doctorat présentée à l'Université de Paris 6 et Université de Milan – Novembre 2002.

[15] ROSENBLATT R. (1956), *A central limit theorem and a strong mixing condition*. Proceeding, Nat. Acad. Sci. USA, Vol 42, p.42–47.

[16] WALTER P. (1969), The central limit problem for the mixing sequences of random variables. *Z. Wahrs.verw. Geb*. 12, p.155–171.