Available online at www.sciencedirect.com
SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 337 (2003) 801–804

COMPTES RENDUS MATHEMATIQUE

Statistique/Probabilités

La convergence faible des U-statistiques multivariées pour des processus non stationnaires

Michel Harel, Echarif Elharfaoui

Laboratoire de statistique et probabilités UMR C55830, Université Paul Sabatier Toulouse III, 31062 Toulouse, France

Reçu le 9 juillet 2003 ; accepté le 29 septembre 2003

Présenté par Paul Deheuvels

Résumé

Le but est d'étudier le comportement asymptotique de la U-statistique multivariée pour des processus non stationnaires, d'une fonctionnelle régulière $\theta(\mathbf{F})$ où $\mathbf{F}$ est la fonction de répartition (f.r.) d'une observation. On étudie pour commencer le comportement asymptotique des processus de vecteurs aléatoires indépendants, non stationnaires. Puis on examine le cas des processus de vecteurs aléatoires non stationnaires, dépendants avec un coefficient de mélange (taux d'absolue régularité $\beta(m)$ $(m \geqslant 1)$). On établit la convergence en loi, sous des hypothèses d'intégrabilité uniforme d'ordre q (>2) des fonctions à noyaux symétriques pour des observations indépendantes non stationnaires, et sous l'hypothèse importante : la convergence des fonctions de répartition non stationnaires pour une certaine norme. ***Pour citer cet article : M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 337 (2003).***

Abstract

The slow convergence of multivariate U-statistic for nonstationary processes. The object is to study the asymptotic behavior of the multivariate U-statistic for nonstationary independent processes of a regular functional $\theta(\mathbf{F})$ where $\mathbf{F}$ is the distribution function of an observation. First, the asymptotic behavior of nonstationary independent processes is studied. Then the use of nonstationary dependent processes with a coefficient of mixing (absolute regularity rate $\beta(m)$ $(m \geqslant 1)$) is addressed. The convergence in law is established, under assumptions of uniform integrability of order q (>2) of the kernel symmetric functions for nonstationary independent observations and under the important assumption: the convergence of the nonstationary distributions functions for some norm. ***To cite this article: M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 337 (2003).***

1. Introduction

Soit $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ une suite de vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^r$ $(r \geqslant 2)$, où $\mathbf{X}_\alpha$ a pour f.r. $\mathbf{F}_\alpha$ $(\alpha \in \mathbb{N}^*)$, soit $\Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)})$ une fonction à noyau symétrique en ses $m(\gamma)$ $(\leqslant n)$ vecteurs. Si la suite de f.r. $\mathbf{F}_\alpha$ converge pour une certaine norme vers une f.r. $\mathbf{F}$, on définit le réel $\theta^{(\gamma)}(\mathbf{F})$ par

Adresses e-mail : harel@unilim.fr (M. Harel), elharfa@cict.fr (E. Elharfaoui).

1631-073X/$ – see front matter © 2003 Académie des sciences. Publié par Elsevier SAS. Tous droits réservés.
doi:10.1016/j.crma.2003.09.034

$$\theta^{(\gamma)}(\mathbf{F}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)})\,\mathrm{d}\mathbf{F}(\mathbf{x}_1) \cdots \mathrm{d}\mathbf{F}(\mathbf{x}_{m(\gamma)})\,; \quad \gamma = 1, \ldots, g\ (g \geqslant 1), \tag{1}$$

$\theta^{(\gamma)}(\mathbf{F})$ est appelée la fonctionnelle régulière de $\mathbf{F}$.

On définit l'estimateur sans biais asymptotique de $\theta^{(\gamma)}(\mathbf{F})$ par :

$$U_n^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum_{1 \leqslant \alpha_1 < \cdots < \alpha_{m(\gamma)} \leqslant n} \Phi^{(\gamma)}(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_{m(\gamma)}})\,; \quad \gamma = 1, \ldots, g, \tag{2}$$

où $\sum$ désigne la somme sur les $\binom{n}{m(\gamma)}$ combinaisons de $m(\gamma)$ éléments distincts dans $\{1, \ldots, n\}$.

La statistique $U_n^{(\gamma)}$ est appelée U-statistique multivariée.

Dans cet article, on étudie le comportement asymptotique de la U-statistique multivariée pour des processus non stationnaires, d'une fonctionnelle régulière $\theta(\mathbf{F})$ où $\mathbf{F}$ est la f.r. d'une observation. Dans la Section 2, on généralise les résultats de Hoeffding [2] pour des U-statistiques multivariées quand les $\mathbf{X}_\alpha$ ont des distributions différentes, et on étudie la convergence en loi des processus non stationnaires indépendants. Dans la Section 3, on prolonge le résultat de Yoshihara [3] au cas des processus identiquement distribués, absolument réguliers et on généralise aussi le résultat de Harel et Puri [1] établi pour des processus de variables aléatoires non stationnaires au cas des processus de vecteurs aléatoires non stationnaires dépendants, nous passerons donc du cas indépendant étudié dans la Section 2 au cas dépendant.

Signalons enfin que les U-statistiques interviennent dans plusieurs domaines de la statistique mathématique. Nos résultats peuvent être appliqués, par exemple, à la distribution asymptotique d'une large classe de tests de rangs qui sont des outils fondamentaux dans de nombreuses applications.

2. Théorème limite d'une U-statistique pour des processus indépendants

Pour établir le théorème limite central de la U-statistique, nous commencerons par définir une fonction à noyau symétrique associée à des processus non stationnaires comme suit :

2.1. *Définitions*

Soient $i_1, i_2, \ldots, i_p$ des entiers arbitraires tels que $1 \leqslant i_1 < i_2 < \cdots < i_p \leqslant n$, pour tout p $(1 \leqslant p \leqslant m(\gamma) - 1)$, on pose : $I_{p,n}(i_1, \ldots, i_p) = \{(i_{p+1}, \ldots, i_{m(\gamma)});\ 1 \leqslant i_{p+1} < \cdots < i_{m(\gamma)} \leqslant n,\ i_l \notin \{i_1, \ldots, i_p\},\ p+1 \leqslant l \leqslant m(\gamma)\}$ et

$$\Phi_{p,n}^{(\gamma),(i_1,\ldots,i_p)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \sum_{(i_{p+1},\ldots,i_{m(\gamma)}) \in I_{p,n}(i_1,\ldots,i_p)} \lambda(\mathbf{x}_1, \ldots, \mathbf{x}_p), \tag{3}$$

où $\lambda(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \int_{(\mathbb{R}^r)^{m(\gamma)-p}} \Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)})\,\mathrm{d}\mathbf{F}_{i_{p+1}}(\mathbf{x}_{p+1}) \cdots \mathrm{d}\mathbf{F}_{i_{m(\gamma)}}(\mathbf{x}_{m(\gamma)})$. On pose aussi : $I_{0,n} = \{(i_1, \ldots, i_m)\,;\ 1 \leqslant i_1 < \cdots < i_m \leqslant n\}$ et $\Phi_{0,n} = \sum_{(i_1,\ldots,i_m) \in I_{0,n}} \int_{(\mathbb{R}^r)^m} \Phi(\mathbf{x}_1, \ldots, \mathbf{x}_m)\,\mathrm{d}\mathbf{F}_{i_1}(\mathbf{x}_1) \cdots \mathrm{d}\mathbf{F}_{i_m}(\mathbf{x}_m)$. Si $p = m$, on pose $\Phi_{m,n}^{(i_1,\ldots,i_m)} = \Phi$.

Par suite pour tout p $(1 \leqslant p \leqslant m - 1)$, on définit :

$$U_{p,n}^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant n} \int_{(\mathbb{R}^r)^p} \Phi_{p,n}^{(\gamma),(i_1,\ldots,i_p)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) \prod_{j=1}^{p} \mathrm{d}\big(\mathbb{1}_{[\mathbf{X}_{i_j} \leqslant \mathbf{x}_j]} - \mathbf{F}_{i_j}(\mathbf{x}_j)\big), \tag{4}$$

où $\mathbb{1}_{[\,]}$ désigne la fonction indicatrice de $[\,]$.

Par définition, on vérifie facilement que :

$$U_n^{(\gamma)} = \theta_n^{(\gamma)} + \sum_{p=1}^{m(\gamma)} \binom{m(\gamma)}{p} U_{p,n}^{(\gamma)}\,; \quad \gamma = 1, \ldots, g\ (g \geqslant 1) \tag{5}$$

qui est une généralisation de la décomposition de Hoeffding, où

$$\theta_n^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum_{1 \leqslant \alpha_1 < \cdots < \alpha_{m(\gamma)} \leqslant n} \theta^{(\gamma)}_{\alpha_1,\ldots,\alpha_{m(\gamma)}}, \tag{6}$$

avec

$$\theta^{(\gamma)}_{\alpha_1,\ldots,\alpha_{m(\gamma)}} = E\big[\Phi^{(\gamma)}(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_{m(\gamma)}})\big]\,; \quad \gamma = 1, \ldots, g\ (g \geqslant 1). \tag{7}$$

On peut écrire aussi que : $\theta_n^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \Phi_{0,n}$.

2.2. Théorème principal

En utilisant les définitions et les notations ci-dessus, on obtient le théorème suivant :

Théorème 2.1. *Supposons qu'il existe un nombre strictement positif δ tel que pour $q = 2 + \delta$:*

$$\sup_{n \in \mathbb{N}^*} \max_{1 \leqslant \alpha_1 < \cdots < \alpha_{m(\gamma)} \leqslant n} E\big|\Phi^{(\gamma)}(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_{m(\gamma)}})\big|^q < +\infty \tag{8}$$

et que

$$\lim_{n \to +\infty} E\big[\sqrt{n} U_{1,n}^{(\gamma)} \sqrt{n} U_{1,n}^{(\delta)}\big] := \sigma_{\gamma,\delta} \tag{9}$$

existe, lorsque $1 \leqslant \gamma \leqslant g$, $1 \leqslant \delta \leqslant g$ où $U_{1,n}^{(\gamma)}$ $(1 \leqslant \gamma \leqslant g)$ est définie comme dans (4).

Posons $\mathbf{0} = (0, \ldots, 0)^t \in \mathbb{R}^g$ et $\boldsymbol{\Sigma}' = (= (\sigma'_{\gamma,\delta})_{\gamma,\delta=1,\ldots,g})$, avec $\sigma'_{\gamma,\delta} = m(\gamma)m(\delta)\sigma_{\gamma,\delta}$.

Alors on a la convergence en loi de $(\sqrt{n}\{U_n^{(1)} - \theta_n^{(1)}\}, \ldots, \sqrt{n}\{U_n^{(g)} - \theta_n^{(g)}\})$ vers la loi normale $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}')$, quand n tend vers $+\infty$.

En utilisant la décomposition de Hoeffding, la preuve du Théorème 2.1 est établie grâce aux deux résultats suivants :

Lemme 2.1. *Sous les conditions du Théorème* 2.1, *le vecteur aléatoire $\big(\sqrt{n} U_{1,n}^{(1)}, \ldots, \sqrt{n} U_{1,n}^{(g)}\big)$ converge en loi vers la loi normale g-variée de moyenne $\mathbf{0}$ et de matrice covariance $\boldsymbol{\Sigma}$ $(= (\sigma_{\gamma,\delta})_{\gamma,\delta=1,\ldots,g})$, quand $n \to +\infty$.*

Lemme 2.2. *Sous les conditions du Théorème* 2.1, *pour tout $p \geqslant 2$, $\sqrt{n} U_{p,n}^{(\gamma)}$ $(\gamma = 1, \ldots, g)$ converge en probabilité vers* 0, *quand n tend vers $+\infty$.*

Définition 2.1. Soient P et Q deux mesures de probabilité définies sur l'espace probabilisable $(\mathbb{R}^r, \mathfrak{B}^r)$ où $\mathfrak{B}$ est la tribu borélienne, on rappelle que la distance entre P et Q en norme de variation totale est définie par :

$$\|P - Q\|_V = \sup_{A \in \mathfrak{B}^r} \big|P(A) - Q(A)\big|.$$

Par conséquent, si $\mathbf{F}$ et $\mathbf{G}$ sont les fonctions de répartition de deux vecteurs aléatoires réels $\mathbf{X}$ et $\mathbf{Y}$, on note par convention : $\|\mathbf{F} - \mathbf{G}\|_V = \|P_{\mathbf{X}} - P_{\mathbf{Y}}\|_V$, où $P_{\mathbf{X}}$ et $P_{\mathbf{Y}}$ sont les lois images de $\mathbf{X}$ et $\mathbf{Y}$.

On déduit du Théorème 2.1, le corollaire suivant :

Corollaire 2.1. *On considère la U-statistique définie dans* (2). *Si la condition* (8) *est satisfaite, et si l'on suppose de plus qu'il existe une fonction de répartition* $\mathbf{F}$ *telle que* :

$$\|\mathbf{F}_\alpha - \mathbf{F}\|_V = \mathrm{O}\big(\tau^\alpha\big)\,; \quad 0 < \tau < 1,\ \alpha \in \mathbb{N}^*, \tag{10}$$

alors : $\lim_{n\to+\infty} \theta_n^{(\gamma)} = \theta^{(\gamma)}$ $(\gamma = 1, \ldots, g)$, *où la fonctionnelle* $\theta^{(\gamma)}$ *est définie en* (1). *De plus, la conclusion du Théorème* 2.1 *reste vraie en remplaçant* $\theta_n^{(\gamma)}$ *par* $\theta^{(\gamma)}$.

3. Théorème limite d'une U-statistique sous la condition d'absolue régularité

On considère, maintenant $\{\mathbf{X}_n\}_{n\in\mathbb{N}^*}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^r$ et absolument réguliers de taux de mélange $\beta(m)$ $(m \geqslant 1)$. Supposons qu'il existe un réel strictement positif δ' tels que :

$$\beta(m) = \mathrm{O}\big(m^{-(2+\delta')/\delta'}\big)\,; \quad 0 < \delta' < \delta. \tag{11}$$

Pour une définition de l'absolue régularité dans le cas non stationnaire voir Harel et Puri [1].

On note $\mathbf{H}_{i,j}$ la f.r. conjointe du vecteur aléatoire $(\mathbf{X}_i, \mathbf{X}_j)$ et on suppose qu'il existe une suite de vecteurs aléatoires $\mathbf{X}_i^*$ strictement stationnaire, de f.r. $\mathbf{F}$, absolument régulière et de même taux que la suite $\mathbf{X}_i$ telle qu'en notant $\mathbf{H}_l^*$ la f.r. conjointe définie sur $\mathbb{R}^r \times \mathbb{R}^r$ de $(\mathbf{X}_i^*, \mathbf{X}_j^*)$ $(l = j - i)$ on ait

$$\|\mathbf{H}_{i,j} - \mathbf{H}_{j-i}^*\|_V = \mathrm{O}\big(\tau^i\big)\,; \quad 0 < \tau < 1,\ m \leqslant i < j. \tag{12}$$

On définit la suite $\sigma_{\gamma,n}^2$ par :

$$\sigma_{\gamma,n}^2 = \frac{1}{n}\sum_{i=1}^{n}\big\{E\big(\Phi_{1,n}^{(\gamma),(i)}(\mathbf{X}_i)\big)^2 - \big(\theta_n^{(\gamma)}\big)^2\big\} + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n-i}\big\{E\big(\Phi_{1,n}^{(\gamma),(i)}(\mathbf{X}_i)\Phi_{1,n}^{(\gamma),(j)}(\mathbf{X}_j)\big) - \big(\theta_n^{(\gamma)}\big)^2\big\} \tag{13}$$

et la constante σ_γ^2 strictement positive par :

$$\sigma_\gamma^2 = \big\{E\big(\Phi_1^{(\gamma)}(\mathbf{X}_1^*)\big)^2 - \big(\theta^{(\gamma)}\big)^2\big\} + 2\sum_{k=1}^{\infty}\big\{E\big(\Phi_1^{(\gamma)}(\mathbf{X}_1^*)\Phi_1^{(\gamma)}(\mathbf{X}_{k+1}^*)\big) - \big(\theta^{(\gamma)}\big)^2\big\}, \tag{14}$$

où $\Phi_1^{(\gamma)}(\mathbf{X}_i^*) = \int_{(\mathbb{R}^r)^{m(\gamma)-1}} \Phi^{(\gamma)}(\mathbf{X}_i^*, \mathbf{x}_2, \ldots, \mathbf{x}_{m(\gamma)})\,\mathrm{d}\mathbf{F}(\mathbf{x}_2)\cdots\mathrm{d}\mathbf{F}(\mathbf{x}_{m(\gamma)})$. Sous la condition (13), on définit :

$$\sigma_{\gamma,\delta} = \big\{E\big[\big(\Phi_1^{(\gamma)}(\mathbf{X}_1^*) - \theta^{(\gamma)}\big)\big(\Phi_1^{(\delta)}(\mathbf{X}_1^*) - \theta^{(\delta)}\big)\big]\big\} + 2\sum_{k=1}^{\infty}\big\{E\big(\Phi_1^{(\gamma)}(\mathbf{X}_1^*)\Phi_1^{(\delta)}(\mathbf{X}_{k+1}^*) - \theta^{(\gamma)}\theta^{(\delta)}\big)\big\}. \tag{15}$$

Théorème 3.1. *Si les conditions* (8), (11) *et* (13) *sont satisfaites, alors la suite* $\sigma_{\gamma,n}^2$ *définie en* (14) *converge vers la constante* σ_γ^2 *définie en* (15), *et le vecteur aléatoire* $(\sqrt{n}\{U_n^{(1)} - \theta_n^{(1)}\}, \ldots, \sqrt{n}\{U_n^{(g)} - \theta_n^{(g)}\})$ *converge en loi vers la loi normale g-variée de moyenne* $\mathbf{0}$ *et de matrice de covariance* $(m(\gamma)m(\delta)\sigma_{\gamma,\delta})_{\gamma,\delta=1,\ldots,g}$ *où* $\sigma_{\gamma,\delta}$ *est défini en* (16).

Références

[1] M. Harel, M.L. Puri, Limiting behavior of U-statistics, V-statistics and one-sample rank order statistics for nonstationary absolutely regular processes, J. Multivariate Anal. 30 (1989) 181–204.

[2] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Statist. 14 (3) (1948).

[3] K. Yoshihara, Limiting behavior of U-statistics for stationary, absolutely regular processes, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 35 (1976) 237–252.