Available online at www.sciencedirect.com
SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 340 (2005) 449–452

COMPTES RENDUS
MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Comportement asymptotique des fonctions de répartition perturbées pour des processus non stationnaires et absolument réguliers

Michel Harel [1], Echarif Elharfaoui

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

Reçu le 3 décembre 2004 ; accepté après révision le 18 janvier 2005

Présenté par Paul Deheuvels

Résumé

Le but est d'établir la normalité asymptotique des statistiques associées à la fonction de répartition empirique perturbée via la convergence en loi d'une U-statistique multivariée. On généralise les résultats de Sun (1993) du cas de variables aléatoires identiquement distribuées, absolument régulières au cas de vecteurs aléatoires non stationnaires, absolument réguliers. ***Pour citer cet article : M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Abstract

Asymptotic behavior of the perturbed empirical distribution functions for nonstationary absolutely regular processes. The object is to study the asymptotic normality of the statistics associated to the perturbed empirical distribution function via the slow convergence of multivariate U-statistic. We extend the results of Sun (1993) from the case of identically distributed absolutely regular random variables to the case of nonstationary absolutely regular random vectors. ***To cite this article: M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

1. Introduction

On considère $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^g$ $(g \geqslant 2)$, où $\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(g)})$ a pour fonction de répartition (f.r.) $\mathbf{F}_n$ $(n \in \mathbb{N}^*)$, absolument régulière de taux $\beta(m)$ $(m \geqslant 1)$. Supposons qu'il existe deux nombres réels δ et δ' $(0 < \delta' < \delta \leqslant \frac{2}{5})$ tels que :

Adresses e-mail : harel@unilim.fr (M. Harel), eelharfaoui@yahoo.fr (E. Elharfaoui).
[1] IUFM Limousin, 209, boulevard de Vanteaux, 87036 Limoges cedex, France.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.01.018

$$\beta(m) = \mathrm{O}(m^{-(2+\delta')/\delta'}). \tag{1}$$

Pour tout γ $(1 \leqslant \gamma \leqslant g)$ considérons un entier $m(\gamma)$ $(\leqslant n)$, soit $\Phi^{(\gamma)}$ une application de $(\mathbb{R}^g)^{m(\gamma)}$ dans $\mathbb{R}$ appelée noyau de degré $m(\gamma)$ c'est-à-dire $\Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)})$ est invariante pour toute permutation sur les $m(\gamma)$ indices des $\mathbf{x}_n$ $(\in \mathbb{R}^g)$, $n = 1, \ldots, m(\gamma)$. Si la suite de f.r. $\mathbf{F}_n$ converge pour une certaine norme vers une f.r. $\mathbf{F}$ de densité $\mathbf{f}$, on définit le paramètre fonctionnel par :

$$\boldsymbol{\xi} = (\xi^{(1)}, \ldots, \xi^{(g)})\,; \quad g \geqslant 2, \tag{2}$$

où pour tout $\gamma \in \{1, \ldots, g\}$,

$$\xi^{(\gamma)} = \int\limits_{(\mathbb{R}^g)^{m(\gamma)}} \Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)}) \prod_{j=1}^{m(\gamma)} \mathrm{d}\mathbf{F}(\mathbf{x}_j). \tag{3}$$

On considère la U-statistique $\boldsymbol{U}_n$ comme estimateur de $\boldsymbol{\xi}$, basée sur l'échantillon $\mathbf{X}_1, \ldots, \mathbf{X}_n$ définie par :

$$\boldsymbol{U}_n = (U_n^{(1)}, \ldots, U_n^{(g)}),$$

où

$$U_n^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_{m(\gamma)} \leqslant n} \Phi^{(\gamma)}(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_{m(\gamma)}})\,; \quad \gamma = 1, \ldots, g. \tag{4}$$

On considère la fonction de répartition perturbée $\widehat{\mathbf{F}}_n$ définie par :

$$\widehat{\mathbf{F}}_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{K}_n(\mathbf{x} - \mathbf{X}_j)\,; \quad \mathbf{x} = (x^{(1)}, \ldots, x^{(g)}) \in \mathbb{R}^g \tag{5}$$

où $\{\mathbf{K}_n\}_{n \geqslant 1}$ est une suite de fonctions de répartition continues qui converge faiblement vers la fonction de répartition $\mathbf{s}(\mathbf{x})$ définie par :

$$\mathbf{s}(\mathbf{x}) = \begin{cases} 1; & \mathbf{x} \geqslant \mathbf{0}, \\ 0; & \text{sinon}, \end{cases}$$

où $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^g$ et $\mathbf{x} = (x^{(1)}, \ldots, x^{(g)}) \in \mathbb{R}^g$. Soit $\mathbf{k}$ une densité de probabilité sur $\mathbb{R}^g$, et $\{a_n\}$ une suite de nombres positifs telle que : $a_n \to 0$, lorsque $n \to +\infty$. En utilisant l'estimateur à noyau $\hat{\mathbf{f}}_n$ de la densité $\mathbf{f}$ (voir Rosenblatt [3]) défini par :

$$\hat{\mathbf{f}}_n(\mathbf{x}) = \frac{1}{n a_n^g} \sum_{j=1}^{n} \mathbf{k}\left(\frac{\mathbf{x} - \mathbf{X}_j}{a_n}\right)\,; \quad \mathbf{x} = (x^{(1)}, \ldots, x^{(g)}) \in \mathbb{R}^g, \tag{6}$$

on obtient un estimateur de $\mathbf{F}_n$ défini par

$$\widehat{\mathbf{F}}_n(\mathbf{x}) = \int\limits_{-\infty}^{x^{(1)}} \cdots \int\limits_{-\infty}^{x^{(g)}} \hat{\mathbf{f}}_n(\mathbf{t})\,\mathrm{d}\mathbf{t}\,; \quad \mathbf{x} = (x^{(1)}, \ldots, x^{(g)}) \in \mathbb{R}^g. \tag{7}$$

On s'intéresse au comportement asymptotique de l'estimateur $\widehat{\mathbf{F}}_n(\boldsymbol{U}_n)$. Dans ce qui suit, nous allons proposer une généralisation du théorème limite central établi par Sun [4] dans le cas univarié pour des variables aléatoires réelles identiquement distribuées et absolument régulières au cas des statistiques multivariées pour des processus non stationnaires et absolument réguliers. Le principe consiste à étudier la normalité asymptotique des statistiques liées à la fonction de répartition empirique, en utilisant les propriétés asymptotiques de $\boldsymbol{U}_n$.

2. Préliminaire

Pour étudier la normalité asymptotique de $\widehat{\mathbf{F}}_n(\boldsymbol{U}_n)$, on utilise une généralisation de la décomposition de Hoeffding définie ci-dessous :

$$U_n^{(\gamma)} = \theta_n^{(\gamma)} + \sum_{p=1}^{m(\gamma)} \binom{m(\gamma)}{p} U_{p,n}^{(\gamma)}, \tag{8}$$

où

$$\theta_n^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_{m(\gamma)} \leqslant n} \int_{(\mathbb{R}^g)^{m(\gamma)}} \Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)}) \times \mathrm{d}\mathbf{F}_{i_1}(\mathbf{x}_1) \cdots \mathrm{d}\mathbf{F}_{i_{m(\gamma)}}(\mathbf{x}_{m(\gamma)}),$$

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

et

$$\Phi_{p,n}^{(\gamma),(i_1,\ldots,i_p)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \sum_{(i_{p+1},\ldots,i_{m(\gamma)}) \in I_{p,n}(i_1,\ldots,i_p)} \lambda^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_p), \tag{9}$$

avec

$$\lambda^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \int_{(\mathbb{R}^g)^{m(\gamma)-p}} \Phi^{(\gamma)}(\mathbf{x}_1, \ldots, \mathbf{x}_{m(\gamma)})\, \mathrm{d}\mathbf{F}_{i_{p+1}}(\mathbf{x}_{p+1}) \cdots \mathrm{d}\mathbf{F}_{i_{m(\gamma)}}(\mathbf{x}_{m(\gamma)})$$

et

$$I_{p,n}(i_1, \ldots, i_p) = \big\{(i_{p+1}, \ldots, i_{m(\gamma)})\,;\, 1 \leqslant i_{p+1} < \cdots < i_{m(\gamma)} \leqslant n, \quad i_l \notin \{i_1, \ldots, i_p\},\ p+1 \leqslant l \leqslant m(\gamma)\big\},$$

$\forall p\ :\ 1 \leqslant p \leqslant m(\gamma) - 1$ et pour tout $i_1, \ldots, i_p$ des entiers arbitraires tels que : $1 \leqslant i_1 < \cdots < i_p \leqslant n$. 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 $\mathbb{R}^g \times \mathbb{R}^g$ de $(\mathbf{X}_i^*, \mathbf{X}_j^*)$, $l = j - i$ on ait

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

où $\|\ \|_V$ est la norme de variation totale voir Harel et Elharfaoui [1].

On définit

$$\mu_{n,l} = E\big(\mathbf{K}_n(\boldsymbol{\xi} - \mathbf{X}_l)\big) = \int_{\mathbb{R}^g} \mathbf{K}_n(\boldsymbol{\xi} - \mathbf{x})\, \mathrm{d}\mathbf{F}_l(\mathbf{x}) \leqslant 1, \quad 1 \leqslant l \leqslant n, \tag{11}$$

aussi

$$\mu_n = \frac{1}{n} \sum_{l=1}^{n} \mu_{n,l} \tag{12}$$

et

$$A_l = \mathbf{s}(\boldsymbol{\xi} - \mathbf{X}_l^*) - \mathbf{F}(\boldsymbol{\xi}) + \sum_{\gamma=1}^{g} \big(m(\gamma)\big)^2 \big(\Phi_1^{(\gamma)}(\mathbf{X}_l^*) - \xi^{(\gamma)}\big) \frac{\partial \mathbf{F}}{\partial x^{(\gamma)}}(\boldsymbol{\xi}), \tag{13}$$

où

$$\Phi_1^{(\gamma)}(\mathbf{X}_l^*) = \int_{(\mathbb{R}^g)^{m(\gamma)-1}} \Phi^{(\gamma)}(\mathbf{X}_l^*, \mathbf{x}_2, \ldots, \mathbf{x}_{m(\gamma)})\,\mathrm{d}\mathbf{F}(\mathbf{x}_2)\cdots\mathrm{d}\mathbf{F}(\mathbf{x}_{m(\gamma)}).$$

D'après la condition (10), μ_n converge vers $\mathbf{F}(\boldsymbol{\xi})$ quand n tend vers $+\infty$. On définit aussi

$$\sigma^2 = E(A_1^2) + 2\sum_{k=1}^{\infty} E(A_1 A_{k+1}). \tag{14}$$

On suppose qu'il existe un nombre réel δ'' $(0 < \delta'' < \delta)$ tel que :

$$\sum\nolimits_{m\geqslant 1} \big(\beta(m)\big)^{\delta''/(2+\delta'')} < +\infty. \tag{15}$$

Pour une définition de l'absolue régularité dans le cas non stationnaire voir Harel et Puri [2].

3. Convergence faible de la fonction de répartition empirique perturbée

En supposant que les conditions introduites dans le paragraphe 2 sont satisfaites, on établit que $\widehat{\mathbf{F}}_n(\boldsymbol{U}_n)$ est un estimateur convergeant vers $\mathbf{F}(\boldsymbol{\xi})$ et la normalité asymptotique nous permettra d'établir des intervalles de confiance pour $\mathbf{F}(\boldsymbol{\xi})$. Enfin, en utilisant les notations introduites dans le paragraphe 2, on peut donc écrire le résultat suivant :

Théorème 3.1. *On suppose que*

(i) *Il existe un nombre réel* δ $(0 < \delta'' < \delta' < \delta \leqslant \frac{2}{5})$ *et il existe* $M_0 > 0$ *tels que* :

$$\sup_{n\in\mathbb{N}^*} \sup_{1\leqslant i_1<\cdots<i_{m(\gamma)}\leqslant n} E\left|\Phi^{(\gamma)}(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_{m(\gamma)}})\right|^{2+\delta} \leqslant M_0 < +\infty.$$

(ii) *Le taux d'absolue régularité vérifie* : (1) *et* (15).

(iii) *La condition* (10) *est vérifiée.*

(iv) $$\int_{\mathbb{R}^g} \|\mathbf{t}\|\mathbf{k}(\mathbf{t})\,\mathrm{d}\mathbf{t} < +\infty, \quad \|\mathbf{t}\| = \sup_{1\leqslant\gamma\leqslant g} |t^{(\gamma)}|.$$

(v) $\mathbf{F}_n$ *et* $\mathbf{F}$ *sont de classe* C^2 *à dérivées partielles premières et secondes bornées.*

Alors $\sqrt{n}\{\widehat{\mathbf{F}}_n(\boldsymbol{U}_n) - \mathbf{F}(\boldsymbol{\xi})\}$ *converge en loi vers une loi normale* $\mathcal{N}(0, \sigma^2)$, *quand* n *tend vers* $+\infty$ *où* σ^2 *est définie en* (14).

Références

[1] M. Harel, E. Elharfaoui, La convergence faible des U-statistiques multivariées pour des processus non stationnaires, C. R. Acad. Sci. Paris, Ser. I 337 (2003) 801–804.

[2] 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.

[3] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist. 27 (1956) 832–837.

[4] S. Sun, Asymptotic behavior of the perturbed empirical distribution functions evaluated at a random point for absolutely regular sequences, J. Multivariate Anal. 47 (1993) 230–249.