Statistics and Probability Letters 106 (2015) 157–164

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Stability of expected L-statistics against weak dependence of observations

A. Okolewski *, M. Kaluszka

Institute of Mathematics, Lodz University of Technology, 90-924 Lodz, Poland

ARTICLE INFO

Article history:
Received 14 April 2015
Received in revised form 15 July 2015
Accepted 16 July 2015
Available online 23 July 2015

Keywords:
Order statistics
L-statistics
Dependent observations
Mixings

ABSTRACT

We study the stability of L-statistics with respect to dependence structures belonging to nonparametric neighbourhoods of independence of Kolmogorov and χ^2 types. The resulting bounds are expressed in terms of numerical characteristics of independent sample and parameters controlling neighbourhood's size.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Denote by $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ the order statistics from the sample $X_1, \ldots, X_n$. Linear combinations of order statistics, called L-statistics, form an important class of estimators (see, e.g. Serfling, 1980; David and Nagaraja, 2003). In the financial context, they accommodate numerous indices of economic inequality as well as actuarial risk measures (see Kremer, 1998; Dowd et al., 2008; Greselin et al., 2009). Bounds on expected L-statistics from possibly dependent samples were studied by Arnold and Groeneveld (1979), Rychlik (1993a,b), Papadatos (2001) and Kaluszka et al. (2005), among others. A comprehensive survey on this topic can be found in Rychlik (1998, 2001a). The sensitivity of L-statistics under arbitrary violations of independence assumption has been examined for the first time by Rychlik (1993b, 2001b, 2007). Kaluszka and Okolewski (2011, 2014) presented some (usually not attainable) evaluations of stability of L-statistics with respect to dependence structures belonging to several Kolmogorov-type neighbourhoods of independence, and applied them to derive additional multiple life premium loading related to the dependence of lifetimes.

In this paper we study the effect of dependence on expected L-estimate in the case when the dependence structure of observations is not fully known, but it belongs to some Kolmogorov-type or χ^2-type neighbourhood of independence, motivated by ψ-mixing (for notions of ψ-mixing see, e.g., Dedecker et al., 2007). We give sharp bounds on the difference of expected L-estimates calculated from independent and dependent samples with the same univariate marginal distributions. They are expressed in terms of numerical characteristics of the independent sample and coefficients describing the neighbourhood size. To the best of our knowledge, the presented bounds are the first ones for expected L-statistics from weakly dependent observations which are attainable.

* Corresponding author.
E-mail addresses: oko@p.lodz.pl (A. Okolewski), kaluszka@lodz.pl (M. Kaluszka).

http://dx.doi.org/10.1016/j.spl.2015.07.019
0167-7152/

2. Stability of expected L-estimates

Let $X_1, X_2, \dots, X_n$, $n \geq 2$, be non-negative random variables with distribution functions $F_1, F_2, \dots, F_n$ and let $X'_1, \dots, X'_n$ be independent random variables such that $X'_i =^d X_i$ for $i = 1, \dots, n$, where $=^d$ means the equality in distribution. Denote by $X'_{1:n} \leq \cdots \leq X'_{n:n}$ the order statistics from the sample $X'_1, \dots, X'_n$ and by $X'_{1:I}$ the minimum $\min\{X'_i : i \in I\}$, $\emptyset \neq I \subset \{1, 2, \dots, n\}$. We will denote by C the Farlie–Gumbel–Morgenstern (FGM) n-copula defined as

$$C(u_1, \dots, u_n) = \left(1 + \sum_{t=2}^{n} \sum_{I \in \mathcal{C}_t} \rho_I \prod_{j \in I} (1 - u_j)\right) \prod_{i=1}^{n} u_i, \tag{1}$$

where $\mathcal{C}_t$ stands for the family of all t-element subsets of $\{1, 2, \dots, n\}$ and the parameters $\rho_I \in [-1, 1]$, $I \in \mathcal{I}_2 = \bigcup_{t=2}^{n} C_t$, are such that

$$1 + \sum_{t=2}^{n} \sum_{I \in \mathcal{C}_t} \rho_I \prod_{i \in I} \psi_i \geq 0 \quad \text{for all } \psi_1, \dots, \psi_n \in \{-1, 1\} \tag{2}$$

(see Kotz et al., 2000). The admissible parameter set Θ constitutes a closed convex polyhedron in $\mathbb{R}^{2^n - n - 1}$, which contains the ball with centre at zero and radius 1 in l_1 norm. Some extension of the FGM n-copula has been proposed by Cambanis (1977) (cf. Hashorva and Hüsler, 1999; Hashorva, 2001).

Let λ_k, $k = 1, \dots, n$, be real numbers and let $L = \sum_{k=1}^{n} \lambda_k X_{k:n}$, $L' = \sum_{k=1}^{n} \lambda_k X'_{k:n}$ and

$$\tilde{\lambda}_t = \sum_{k=1}^{t} \lambda_{n-k+1} (-1)^{t-k} \binom{t-1}{k-1}, \quad t = 1, 2, 3, \dots, n. \tag{3}$$

We focus our attention on deriving upper bounds. Their lower counterparts can be immediately obtained as $\inf \mathsf{E} \sum_{k=1}^{n} \lambda_k X_{k:n} = -\sup \mathsf{E} \sum_{k=1}^{n} (-\lambda_k) X_{k:n}$. We shall assume that the integrals appearing in the propositions exist and are finite. Moreover, we adopt the convention that $0/0 = 0$, denote by $|A|$ and $int(A)$ the cardinality and the interior of a set A, respectively, and write $\mathcal{C}_{k,I} = \{Z \subset I : |Z| = k\}$ and $\mathrm{sgn}(a) = -\mathbf{1}(a < 0) + \mathbf{1}(a > 0)$, where $\mathbf{1}(z) = 1$ if z is true and $\mathbf{1}(z) = 0$ otherwise.

We first provide evaluations for Kolmogorov-type neighbourhoods of independence.

Proposition 1. *Suppose that for any $t = 2, 3, \dots, n$ and any $I \in \mathcal{C}_t$ there exists $\varepsilon_I \geq 0$ such that*

$$\sup_{x > 0} \frac{\left| \mathsf{P}\left(\bigcap_{i \in I} \{X_i > x\}\right) - \prod_{i \in I} \mathsf{P}(X_i > x) \right|}{\prod_{i \in I} \mathsf{P}(X_i > x)} \leq \varepsilon_I, \tag{4}$$

i.e. the X_i's satisfy some ψ-mixing type dependence condition. Then

$$\mathsf{E}(L - L') \leq \sum_{t=2}^{n} |\tilde{\lambda}_t| \sum_{I \in \mathcal{C}_t} \varepsilon_I \mathsf{E} X'_{1:I}. \tag{5}$$

If the ε_I's are sufficiently small, i.e. $(\tilde{\rho}_I)_{I \in \mathcal{I}_2} \in int(\Theta)$, where $\tilde{\rho}_I = \sum_{k=2}^{t} (-1)^k \mathrm{sgn}(\tilde{\lambda}_k) \sum_{J \in \mathcal{C}_{k,I}} \varepsilon_J$, then the bound (5) is attained for the X_i's which have the FGM n-copula (1) with $\rho_I = \tilde{\rho}_I \prod_{i \in I} p_i^{-1}$, and the marginal distribution functions $F_i = p_i \mathbf{1}_{[0, c_i)} + \mathbf{1}_{[c_i, \infty)}$, $i = 1, 2, \dots, n$, in which c'_i s are arbitrary positive real numbers and $p_1, p_2, \dots, p_n \in (0, 1)$ are so chosen that $(\rho_I)_{I \in \mathcal{I}_2} \in int(\Theta)$.

Proof. We will use the following well-known identity (cf. Blom et al., 1994, p. 32):

$$\mathsf{P}\left(\sum_{i=1}^{n} \mathbf{I}_{A_i} \geq k\right) = \sum_{t=k}^{n} (-1)^{t-k} \binom{t-1}{k-1} \sum_{1 \leq i_1 < \cdots < i_t \leq n} \mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_t}), \tag{6}$$

where $A_1, \dots, A_n$ are arbitrary events and $\mathbf{I}_A(\omega) = \mathbf{1}(\omega \in A)$, $\omega \in \Omega$. Of course

$$\mathsf{P}(X_{k:n} > x) = \mathsf{P}\left(\sum_{i=1}^{n} \mathbf{1}(X_i > x) \geq n - k + 1\right).$$

From (6) we have

$$\mathsf{P}(X_{k:n} > x) = \sum_{t=n-k+1}^{n} (-1)^{t-(n-k+1)} \binom{t-1}{n-k} S_t(x),$$