Statistics and Probability Letters 119 (2016) 259–263

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Asymptotic normality of numbers of observations near order statistics from stationary processes

Krzysztof Jasiński

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, ul. Chopina 12/18, 87-100 Torun, Poland

ARTICLE INFO

Article history:
Received 2 July 2016
Received in revised form 19 August 2016
Accepted 23 August 2016
Available online 30 August 2016

Keywords:
Central limit theorems
Extreme, intermediate, and central order statistics
Near order statistic observations
Stationary processes
α-mixing condition

ABSTRACT

This paper is concerned with asymptotic normality of numbers of observations near order statistics. For some type of discontinuous marginal distributions, we extend knowing results to strictly stationary and ergodic observations satisfying an extra condition which guarantees some local independence.

1. Introduction

Let $(X_n, n \geq 1)$ be a sequence of random variables (rv's) with a common cumulative distribution function (cdf) F. Denote by $X_{1:n} \leq \cdots \leq X_{n:n}$ the order statistics based on the random sample $(X_1, \ldots, X_n)$. If $(k_n, n \geq 1)$ is a sequence of positive integers such that $k_n \leq n$ for all n and $k_n/n \to \lambda \in [0, 1]$ as $n \to \infty$ then $X_{k_n:n}$, $n \geq 1$, are referred to as one of three natural cases: the case of central order statistics (corresponding to $\lambda \in (0, 1)$), that of extreme order statistics (when k_n or $n - k_n$ is fixed) and that of intermediate order statistics (when $\lambda \in \{0, 1\}$ and both k_n and $n - k_n$ approach infinity). In this paper, we will study the asymptotic behavior of the following counting rv:

$$K_{k_n:n}(A) = \#\{j \in \{1, \ldots, n\} : X_{k_n:n} - X_j \in A\} \tag{1.1}$$

as $n \to \infty$, where A is a Borel subset of real numbers and $1 \leq k_n \leq n$. The rv in (1.1) provides information on how many observations fall into a random region determined by the set A and the order statistic $X_{k_n:n}$.

Exact and asymptotic properties of the rv $K_{k_n:n}(A)$ have been studied in the literature due to their applications to different practical problems. In particular, if $k_n = n$ and $A = \{0\}$, the rv in (1.1) counts elements in the sample of size n that are tied with the sample maximum. Moreover, if the cdf F is concentrated on nonnegative integers then $K_{n:n}(\{0\})$ can be interpreted as the numbers of winners in a game with n players whose scores are $X_1, \ldots, X_n$. Under the assumption that $(X_n, n \geq 1)$ is a sequence of independent and identically distributed (i.i.d.) rv's, numerous authors considered the existence of the limit of probability of no ties ($\lim_{n\to\infty} \Pr(K_{n:n}(\{0\})) = 1$), and related problems. For more details see, for example, Eisenberg et al. (1993), Brands et al. (1994), Qi (1997), Bruss and Grübel (2003), Berred and Stepanov (2005), Eisenberg (2009) and Gouet et al. (2009).

E-mail address: krzys@mat.umk.pl.

http://dx.doi.org/10.1016/j.spl.2016.08.014
0167-7152/

If F is continuous then $K_{k_n:n}(\{0\}) = 1$ almost surely. So we do not obtain any information. Therefore, in that case Pakes and Steutel (1997) introduced the rv

$$K_{n:n}((0, a)) = \#\{j \in \{1, \ldots, n\} : X_j \in (X_{n:n} - a, X_{n:n})\}, \quad a > 0,$$

which counts the number of random points near the maximum. It is worth pointing out that this type of the rv can be used in actuarial mathematics to count insurance claims whose size is at prescribed distance from the existing largest claim size (see Li and Pakes, 2001 and Hashorva, 2003). Properties of the rv $K_{n:n}((0, a))$ have been quite intensively studied. It was also generalized to the numbers of near r-extremes, $K_{n-r:n}((0, a))$ and $K_{n-r:n}((-a, 0))$, $a > 0$, $0 \leq r \leq n - 1$. The extension was discussed in Pakes and Li (1998), Pakes (2000), Hu and Su (2003), Balakrishnan and Stepanov (2005), Dembińska and Balakrishnan (2008, 2010). Next, this concept was extended and the asymptotic properties of the numbers of near order statistic observations were obtained, $K_{k_n:n}((0, a))$ and $K_{k_n:n}((-a, 0))$, $a > 0$, $k_n/n \to \lambda \in [0, 1]$ (see Dembińska et al., 2007; Pakes, 2009; Iliopoulos et al., 2012 for more details). Finally, these results were generalized to any Borel set A, $K_{k_n:n}(A)$ by Dembińska (2012a) and Dembińska and Iliopoulos (2012). Some generalizations to the case of not necessarily independent sequences of observations are given by Hashorva (2003), Hashorva and Hüsler (2004) and Balakrishnan et al. (2009).

In the literature one problem, that was extensively investigated, is devoted to the question whether the rv $K_{k_n:n}(A)$ can be centered and normed to yield a normal limit law. The case of extreme order statistics and continuous F was discussed by Pakes and Steutel (1997), Hashorva and Hüsler (2004) and Pakes (2009) under the assumption that the left endpoint of the support of F is finite. For central order statistics and continuous F, these properties were studied by Pakes (2009), Iliopoulos et al. (2012) and Dembińska (2012a). Moreover, two situations: one where the cdf F is discontinuous and the order statistics are extreme, central, intermediate and the other where it is continuous and the order statistics are extreme, were considered by Dembińska (2012b). However, all these results were obtained under i.i.d. assumption of the sequence $(X_n, n \geq 1)$. The aim of this paper is to extend the results for discontinuous case given by Dembińska (2014b). We will replace the i.i.d. assumption with a weaker one that $(X_n, n \geq 1)$ is a strictly stationary and ergodic sequence. This extension is important because dependence is often observed in practical applications. We obtain our result while the cdf F is discontinuous, under some standard conditions which guarantee some local independence and hold for all three cases discussed in the literature: central, extreme and intermediate order statistics. Discontinuity here means that the corresponding λth quantile of the cdf F is not an accumulation point of its support.

Throughout the paper we make use of the following notation. The rv's X_n, $n \geq 1$, exist in a probability space $(\Omega, \mathcal{F}, P)$. $\mathbb{R}$ and $\mathbb{Z}$ represent the sets of real numbers and integers, respectively. $\mathcal{B}(\mathbb{R})$ stands for the Borel σ-field of subsets of $\mathbb{R}$. The support of the distribution F is denoted by $\mathrm{supp}(F)$ and we set $\gamma_0 := \inf \mathrm{supp}(F) = \inf\{x \in \mathbb{R} : F(x) > 0\}$, $\gamma_1 := \sup \mathrm{supp}(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}$. By γ_λ we denote the unique λth quantile of F where $\lambda \in (0, 1)$. We write $I(\cdot)$ for the indicator function, that is $I(x \in A) = 1$ if $x \in A$ and $I(x \in A) = 0$ otherwise; and we use I_A as an abbreviation for $I(x \in A)$. Moreover, by $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{a.s.}}$ we denote convergence in distribution and almost sure convergence, respectively.

2. Main result

Let $(X_n, n \geq 1)$ be a sequence of not necessarily independent but identically distributed rv's with a common discontinuous cdf F. More precisely, we relax the standard assumption that X_n, $n \geq 1$ are independent. We replace it by the weaker one that $(X_n, n \geq 1)$ is a strictly stationary and ergodic sequence. However, an extra condition is required to ensure that central, extreme, intermediate order statistics from such stationary and ergodic processes have similar asymptotic behavior as the respective ones from i.i.d. observations. This condition is called strong mixing (α-mixing) condition and guarantees a specific type of some local asymptotic independence. Before we formulate it, we introduce some notation.

For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, we define the measure of dependence

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A},\, B \in \mathcal{B}} |\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)|.$$

If $\mathbb{X} = (X_k, k \in \mathbb{Z})$ is a two-sided random sequence, then the coefficients of the α-mixing condition are given by

$$\alpha_{\mathbb{X}}(n) := \sup_{j \in \mathbb{Z}} \alpha\left(\sigma(\ldots, X_{j-1}, X_j), \sigma(X_{j+n}, X_{j+n+1}, \ldots)\right), \quad n \geq 1,$$

where $\sigma(X_k, k \in J)$, $J \subset \mathbb{Z}$, is the σ-field generated by the rv's X_k, $k \in J$. For a one-sided random sequence $\mathbb{X} = (X_n, n \geq 1)$, $\alpha_{\mathbb{X}}(n)$ is by definition the same as for two-sided sequence of the form $(\ldots, 0, 0, 0, X_1, X_2, X_3, \ldots)$. Under these notation, we say that $\mathbb{X}$ is strongly mixing (α-mixing), if $\alpha_{\mathbb{X}}(n) \to 0$ as $n \to \infty$. For discussion of α-mixing condition see Bradley (2007).

Before we present the main theorem that asserts the asymptotic normality of a centered and normed version of the rv $K_{k_n:n}(A)$ when the cdf F is discontinuous, we begin with the following result which describes the asymptotic behavior of the order statistics from stationary and ergodic sequences. It extends Lemma 1 of Dembińska (2012b), because the independence assumption is relaxed.