Statistics and Probability Letters 132 (2018) 7–16

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On limiting distribution of U-statistics based on associated random variables

Mansi Garg *, Isha Dewan

Indian Statistical Institute, New Delhi-110016, India

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 17 August 2017
Accepted 30 August 2017
Available online 18 September 2017

Keywords:
Associated random variables
U-statistics
Central limit theorem
Skewness
Kurtosis

ABSTRACT

Let $\{X_n, n \geq 1\}$ be a sequence of stationary associated random variables. We discuss another set of conditions under which a central limit theorem for U-statistics based on $\{X_n, n \geq 1\}$ holds. We look at U-statistics based on differentiable kernels of degree 2 and above. As applications, we discuss consistent estimators of second, third and fourth central moments, and estimators of skewness and kurtosis based on them.

1. Introduction

In this paper, we obtain another set of assumptions under which a central limit theorem for U-statistics based on stationary associated random variables holds. We look at non-degenerate U-statistics based on differentiable (component-wise monotonic or non-monotonic) kernels of any finite degree $k \geq 2$. The proof requires relatively non-restrictive assumptions. We also illustrate some applications of our results using examples. Apropos our discussion, we give the following.

Definition 1.1 (*Esary et al., 1967*)**.** A finite collection of random variables $\{X_j, 1 \leq j \leq n\}$ is said to be associated, if for any choice of component-wise nondecreasing functions $h, g : \mathbb{R}^n \to \mathbb{R}$, we have

$$Cov(h(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \geq 0,$$

whenever it exists. An infinite collection of random variables $\{X_j, j \geq 1\}$ is associated if every finite sub-collection is associated.

A detailed presentation of the asymptotic results and examples relating to associated sequences can be found in Bulinski and Shashkin (2007), Oliveira (2012) and Prakasa Rao (2012).

For the rest of the paper, assume that $\{X_n, n \geq 1\}$ is a sequence of stationary associated random variables with F as the marginal distribution function of X_1. We next briefly discuss existing results on central limit theorem of U-statistics based on $\{X_n, n \geq 1\}$.

Dewan and Prakasa Rao, (2001) gave a central limit theorem for degenerate and non-degenerate U-statistics using an orthogonal expansion of the underlying kernel. Dewan and Prakasa Rao (2002) and its corrigendum (Dewan and Prakasa Rao, 2015) obtained a central limit theorem for U-statistics with differentiable kernels of degree 2, using Hoeffding's

* Corresponding author.
E-mail addresses: mansibirla@gmail.com (M. Garg), isha@isid.ac.in (I. Dewan).

http://dx.doi.org/10.1016/j.spl.2017.08.016

decomposition. The limiting distribution of U-statistics based on differentiable kernels can also be obtained using the results of Beutner and Zähle (2012, 2014) and Garg and Dewan (2015). We have discussed the difference in our assumptions with an example in Section 4.

The paper is organized as follows. In Section 2, we state a few results and definitions which will be required to prove our main results. Limiting distribution of non-degenerate U-statistics based on $\{X_n, n \geq 1\}$ is given in Section 3. In Section 4, we apply our results to discuss the asymptotic distribution of estimators of second, third and the fourth central moments. We also discuss estimators of skewness and kurtosis, when the underlying sample is from $\{X_n, n \geq 1\}$.

2. Preliminaries

In this section, we give results and definitions which will be needed to prove our main results given in Section 3.

Definition 2.1. Hoeffding's decomposition for U-statistics based on a symmetric measurable function $\rho : \mathbb{R}^2 \to \mathbb{R}$. Define the U-statistic, U_n, by

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \rho(X_i, X_j). \tag{2.1}$$

Let $\theta = \int_{\mathbb{R}^2} \rho(x_1, x_2)\, dF(x_1) dF(x_2)$ and $\rho_1(x_1) = \int_{\mathbb{R}} \rho(x_1, x_2)\, dF(x_2)$. Further, let

$$h^{(1)}(x_1) = \rho_1(x_1) - \theta \text{ and } h^{(2)}(x_1, x_2) = \rho(x_1, x_2) - \rho_1(x_1) - \rho_1(x_2) + \theta. \tag{2.2}$$

Then, Hoeffding's decomposition for U_n is

$$U_n = \theta + 2H_n^{(1)} + H_n^{(2)}, \tag{2.3}$$

where $H_n^{(j)}$ is the U-statistic of degree j based on the kernel $h^{(j)}, j = 1, 2$. When X_j, $1 \leq j \leq n$ are *i.i.d.*, $E(U_n) = \theta$.

Remark 2.2. An extension of Hoeffding's decomposition for U-statistics of a finite degree $k > 2$ can be found in Lee (1990).

Lemma 2.3 (Newman, 1980). *Let X and Y be two associated random variables with $E(X^2) < \infty$ and $E(Y^2) < \infty$. Let f and g be differentiable functions with $\sup_x |f'(x)| < \infty$ and $\sup_y |g'(y)| < \infty$. Then,*

$$\begin{aligned} Cov(f(X), g(Y)) &= \int_{\mathbb{R}^2} f'(x) g'(y) [P(X \leq x, Y \leq y) - P(X \leq x) P(Y \leq y)] dxdy \\ &\leq \sup_x |f'(x)| \sup_y |g'(y)|\, Cov(X, Y). \end{aligned}$$

Lemma 2.4 (Lebowitz, 1972). *Define, for A and B, subsets of $\{1, 2, \ldots, n\}$ and real x_j's,*

$$H_{A,B}(x_j, j \in A \cup B) = P[X_j > x_j, j \in A \cup B] - P[X_k > x_k, k \in A] P[X_l > x_l, l \in B].$$

If the random variables $X_1, X_2, \ldots, X_n$ are associated, then

$$0 \leq H_{A,B} \leq \sum_{i \in A} \sum_{j \in B} H_{\{i\},\{j\}}. \tag{2.4}$$

Definition 2.5 (Newman, 1984). Let g and $\tilde{g}$ be two real-valued functions on $\mathbb{R}^m$, for some $m \in \mathbb{N}$. $g \ll \tilde{g}$ iff $\tilde{g} + g$ and $\tilde{g} - g$ are both coordinate-wise nondecreasing. If $g \ll \tilde{g}$, then $\tilde{g}$ will be a coordinate-wise nondecreasing function.

Lemma 2.6 (Newman, 1984). *For each j, $j \geq 1$, let $Y_j = f(X_j)$ and $\tilde{Y}_j = \tilde{f}(X_j)$. Suppose that $f \ll \tilde{f}$. Define $\sigma^2 = Var(Y_1) + 2\sum_{j=2}^{\infty} Cov(Y_1, Y_j)$. Let $\sigma^2 > 0$ and $\sum_{j=1}^{\infty} Cov(\tilde{Y}_1, \tilde{Y}_j) < \infty$. Then,*

$$\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} (Y_j - E(Y_j)) \xrightarrow{\mathcal{L}} N(0, 1) \text{ as } n \to \infty. \tag{2.5}$$

In the following, $g \ll_A \tilde{g}$ if $g \ll \tilde{g}$ and both g and $\tilde{g}$ depend only on x_j's with $j \in A$. A is a finite subset of $\{k, k \geq 1\}$.

Lemma 2.7 (Newman, 1984). *Let $g_1 \ll_A \tilde{g}_1$ and $g_2 \ll_A \tilde{g}_2$. Then,*

$$|Cov(g_1(X_1, X_2, \ldots), g_2(X_1, X_2, \ldots))| \leq Cov(\tilde{g}_1(X_1, X_2, \ldots), \tilde{g}_2(X_1, X_2, \ldots)). \tag{2.6}$$