Journal of Multivariate Analysis 121 (2013) 176–192

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Asymptotics of randomly weighted u- and v-statistics: Application to bootstrap☆

Miklós Csörgő, Masoud M. Nasari *

School of Mathematics and Statistics, Carleton University, 1125 Colonel By Drive, Ottawa, Canada

ARTICLE INFO

Article history:
Received 22 October 2012
Available online 27 July 2013

AMS subject classifications:
62G20
62G09

Keywords:
Conditional central limit theorems
Laws of large numbers
Multinomial distribution
u- and v-statistics
Randomly weighted u- and v-statistics
Weighted arrays of random variables
Weighted bootstrap

ABSTRACT

This paper is mainly concerned with asymptotic studies of weighted bootstrap for u- and v-statistics. We derive the consistency of the weighted bootstrap u- and v-statistics, based on i.i.d. and non i.i.d. observations from some more general results which we first establish for sums of randomly weighted arrays of random variables. Some of the results in this paper significantly extend some well-known results on consistency of u-statistics and also consistency of sums of arrays of random variables. We also employ a new approach to conditioning to derive a conditional central limit theorem (CLT) for weighted bootstrap u- and v-statistics, assuming the same conditions as the classical CLT for regular u- and v-statistics.

1. Introduction

The main purpose of this study is to investigate the validity of bootstrap u- and v-statistics resulting from the so-called m-out-of-n scheme of bootstrap. The use of this scheme yields randomly weighted variants of the original u-, v-statistics that exhibit an explicit connection between the original u-, v-statistics and their bootstrap versions. Moreover, it clearly identifies the two types of stochastic variations involved in a bootstrap problem for statistics of the form of partial sums in general, and for u-, v-statistics in particular. In this exposition, our investigation of bootstrap u-, v-statistics results in establishing their consistency and convergence in distribution. While investigating the convergence in distribution for the bootstrap is done only for u-, v-statistics that are based on i.i.d. observations when the m-out-of-n scheme is used, their consistency is studied in a more general context. In fact, our results on consistency are restricted to neither u-, v-statistics, nor to i.i.d. observations or the bootstrap. Rather, the results on consistency are established for randomly weighted arrays of random variables *per se*. These results in turn are used to derive consistency for randomly weighted u-statistics, as well as for bootstrap u-, v-statistics based on i.i.d. observations and observations with absolute regularity property (cf. Corollaries 3.1 and 3.2, respectively). Some of the results in this work extend well-known results on the strong law of large numbers for arrays of random variables and also for u-statistics (cf. Theorem 2.3 and Remark 2.4), and some of them shed light on the consistency of bootstrap when the original sample size can stay finite (cf. Theorems 2.2 and 3.2).

The conditions assumed for the results in this paper are the same as the ones required for the classical CLTs and classical strong and weak laws of large numbers in the non-weighted case. In other words, there are no further restrictions imposed on the observations.

☆ Research supported by NSERC Canada Discovery Grants of M. Csörgő at Carleton University.
* Corresponding author.
E-mail addresses: mmnasari@connect.carleton.ca, masoudnasari@yahoo.com (M.M. Nasari).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.07.008

The material in this paper is organized as follows. In Section 2 laws of large numbers are provided for randomly weighted arrays of random variables. The presentation of these results is so that they can be used in establishing the consistency of bootstrap u-, v-statistics in Section 3. Section 4 is devoted to establishing conditional (given the weights) CLTs for bootstrapped u-, v-statistics. In Section 6, remarks are made on the validity of results for higher dimensional arrays of random variables and u-, v-statistics of order greater than 2. Simulation results, illustrating the validity of the CLTs in Section 4 are presented in Section 5 for the bootstrap sample variance and also the bootstrap Gini's mean difference. Both are of great interest in practice. The proofs are given in Section 7.

2. Laws of large numbers for randomly weighted arrays of random variables

Consistency of bootstrap means as randomly weighted sums was pioneered by Athreya [4] and followed by S. Csörgő [9] and Arenal-Gutiérrez et al. [3]. Since then the problem has received a great deal of attention from researchers. Among the contributions in the field, we specifically mention two results, respectively due to Arenal-Gutiérrez et al. [3], and Rosalsky and Sreehari [18], as the former influenced Theorem 3.2 and the latter motivated Theorem 2.1 of this exposition.

In this paper we consider two sequences of possibly double triangular (with respect to m and n) arrays of random variables, $\{X_{ij}^{(n,m)};\ 1 \le i,j \le n\}$ and $\{\varepsilon_{ij}^{(n,m)};\ 1 \le i,j \le n\}$, $n,m \ge 1$, which are defined on the same probability space $(\Omega_{X,\varepsilon}, \mathfrak{F}_{X,\varepsilon}, P_{X,\varepsilon})$. Also, by $(\Omega_X, \mathfrak{F}_X, P_X)$ and $(\Omega_\varepsilon, \mathfrak{F}_\varepsilon, P_\epsilon)$, we denote the marginal probability spaces of the $X^{(n,m)}$s and $\varepsilon^{(n,m)}$s, respectively. We shall, often, refer to the $X^{(n,m)}$s as the observations (data) and $\varepsilon^{(n,m)}$s are referred to as the weights. We shall investigate the large sample behavior of the randomly weighted sums $\sum_{1\le i,j\le n} \varepsilon_{ij}^{(n,m)} X_{ij}^{(n,m)}$, $n,m \ge 1$. The observations and the weights are so that they can be employed in studying the m-out-of-n scheme of bootstrap for u-statistics which is discussed in Sections 3 and 4. We note in passing that when the observations and the weights are independent, then their joint probability space can of course be defined as the direct product probability space $(\Omega_X \times \Omega_\varepsilon, \mathfrak{F}_X \otimes \mathfrak{F}_\varepsilon, P_X \times P_\varepsilon)$ of their marginals.

In this section we present some strong and weak consistency results for sums of randomly weighted arrays of random variables.

Except for Theorem 2.1 and its application to bootstrap u-statistics in Theorem 3.1 of the next section, the method of conditioning plays an important role in the establishment of the results in this exposition. More precisely, employing hierarchical arguments, we derive our results via conditioning on the weights $\varepsilon^{(n,m)}$ in some stochastic way with respect to P_ε. The latter results, in turn, can be extended to unconditional ones in terms of the joint probability measure $P_{X,\varepsilon}$. We let $P_{\cdot|\varepsilon}(\cdot)$ and $E_{\cdot|\varepsilon}(\cdot)$, respectively stand for the conditional probability and conditional expected value given the weights $\varepsilon^{(n,m)}$.

Theorem 2.1 is a strong law of large numbers for sums of randomly weighted arrays of random variables. It was motivated by, and it generalizes, Theorem 1 of Rosalsky and Sreehari [18] to arrays of random variables so that it can be used in studying the validity of the so called m-out-of-n method of bootstrap u-statistics in Section 3.

For the sake of convenient comparison of the just mentioned two results, we first present Theorem 1 of Rosalsky and Sreehari [18]: *let $\{Y_{i,n},\ 1 \le i \le n, n \ge 1\}$ and $\{X_{i,n}, 1 \le i \le n, n \ge 1\}$ be triangular arrays of random variables and assume that $\{Y_{i,n}, 1 \le i \le n, n \ge 1\}$ consists of integrable random variables and satisfies the condition that there exists a triangular array of positive constants $\{a_{i,n}, 1 \le i \le n, n \ge 1\}$ such that for all $\delta > 0$*

$$P_Y\big(\cup_{1\le i\le n}\{|Y_{i,n} - E_Y(Y_{i,n})| > \delta a_{i,n}\}\ i.o.(n)\big) = 0.$$

If

$$\sum_{1\le i\le n} E_Y(Y_{i,n})X_{i,n} \quad \text{and} \quad \sum_{1\le i\le n} a_{i,n}|X_{i,n}| \quad \text{converge a.s.-}P_X,$$

then $\sum_{1\le i\le n} Y_{i,n}X_{i,n}$ converges a.s.-$P_{X,Y}$ to the a.s.-P_X limit of $\sum_{1\le i\le n} E_Y(Y_{i,n})X_{i,n}$.

Our strong law of large numbers for sums of randomly weighted arrays of random variables reads as follows.

Theorem 2.1. *Consider the two possibly (double) triangular arrays $\{X_{ij}^{(n,m)};\ 1 \le i,j \le n\}$ and $\{\varepsilon_{ij}^{(n,m)};\ 1 \le i,j \le n\}$, $n,m \ge 1$, of random variables which are defined on the same probability space. Let the sequence of positive integers m be such that $m = m(n) \to +\infty$, as $n \to +\infty$. Also let $\{c_{ij}^{(n,m)};\ 1 \le i,j \le n\}$ be a possibly (double) triangular array of real numbers and $\{a_{ij}^{(n,m)};\ 1 \le i,j \le n\}$ be a possibly (double) triangular array of positive real numbers such that, for $\delta > 0$,*

$$\text{(a) } P_\varepsilon(\cup_{1\le i,j\le n} |\varepsilon_{ij}^{(n,m)} - c_{ij}^{(n,m)}| > \delta\, a_{ij}^{(n,m)},\ i.o.(n)\,) = 0, \tag{1}$$

and, as $m(n), n \to +\infty$,

$$\text{(b) } \sum_{1\le i,j\le n} a_{ij}^{(n,m)} |X_{ij}^{(n,m)}| < +\infty \quad a.s.\text{-}P_X,$$

$$\text{(c) } \sum_{1\le i,j\le n} c_{ij}^{(n,m)} X_{ij}^{(n,m)} < +\infty \quad a.s.\text{-}P_X.$$

Then, (a)–(c)*, as $n \to \infty$, imply that*

$\sum_{1\le i,j\le n} \varepsilon_{ij}^{(n,m)} X_{ij}^{(n,m)}$ converges a.s.-$P_{X,\varepsilon}$ to the same a.s.-P_X limit as that of $\sum_{1\le i,j\le n} c_{ij}^{(n,m)} X_{ij}^{(n,m)}$.