Journal of Multivariate Analysis 117 (2013) 257–280

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Dependent wild bootstrap for degenerate U- and V-statistics

Anne Leucht [a], Michael H. Neumann [b,*]

[a] Universität Mannheim, Abteilung Volkswirtschaftslehre, L7, 3-5, D-68131 Mannheim, Germany

[b] Friedrich-Schiller-Universität Jena, Institut für Stochastik, Ernst-Abbe-Platz 2, D-07743 Jena, Germany

ARTICLE INFO

Article history:
Received 24 April 2012
Available online 18 March 2013

AMS 2010 subject classifications:
primary 62M07
62G09
secondary 62E20

Keywords:
Bootstrap
Weak dependence
U-statistic
V-statistic
Cramér–von Mises test
Two-sample test

ABSTRACT

Degenerate U- and V-statistics play an important role in the field of hypothesis testing since numerous test statistics can be formulated in terms of these quantities. Therefore, consistent bootstrap methods for U- and V-statistics can be applied in order to determine critical values for these tests. We prove a new asymptotic result for degenerate U- and V-statistics of weakly dependent random variables. As our main contribution, we propose a new model-free bootstrap method for U- and V-statistics of dependent random variables. Our method is a modification of the dependent wild bootstrap recently proposed by Shao [X. Shao, The dependent wild bootstrap, J. Amer. Statist. Assoc. 105 (2010) 218–235], where we do not directly bootstrap the underlying random variables but the summands of the U- and V-statistics. Asymptotic theory for the original and bootstrap statistics is derived under simple and easily verifiable conditions. We discuss applications to a Cramér–von Mises-type test and a two sample test for the marginal distribution of a time series in detail. The finite sample behavior of the Cramér–von Mises test is explored in a small simulation study. While the empirical size was reasonably close to the nominal one, we obtained nontrivial empirical power in all cases considered.

1. Introduction

U- and related von Mises- (V-)statistics play an important role in mathematical statistics. In the case of hypothesis testing, major interest is on degenerate statistics of this type since they approximate important test quantities under the null hypothesis. Well-known examples are the Cramér–von Mises and the χ^2-statistics. Especially in the case of dependent random variables, the distribution of such a statistic has quite an involved form and depends on characteristics of the underlying process in a complicated manner. For the determination of critical values, we propose new versions of a model-free bootstrap method that can be viewed as variants of the dependent wild bootstrap recently proposed by Shao [39] for smooth functions of the mean.

In Section 2 we derive the limit distributions of degenerate U- and V-statistics under a condition of weak dependence introduced by Dedecker and Prieur [10]. The classical approach is based on a spectral decomposition of the kernel function and was first taken by Gregory [22] in the case of i.i.d. random variables. Later it has been used for mixing random variables by Eagleson [17], Carlstein [9], and Borisov and Volodko [8] as well as for associated random variables by Dewan and Prakasa Rao [14] and Huang and Zhang [23]. This method works actually perfectly well in the case of independent random variables, however, it has to be taken with care in the dependent case. The authors mentioned above imposed additional conditions on the corresponding eigenvalues and eigenfunctions that can hardly be checked in practice. Making use of the observation that typical test statistics of L_2-type can be approximated by V-statistics with positive semidefinite kernel functions, Leucht and Neumann [28] could simplify technical issues and derived the asymptotics for U- and V-statistics for ergodic processes.

* Corresponding author.
E-mail addresses: leucht@uni-mannheim.de (A. Leucht), michael.neumann@uni-jena.de (M.H. Neumann).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.03.003

They imposed a stricter form of degeneracy that is satisfied by V-statistics resulting from model-specification tests for the conditional mean function or from goodness-of-fit tests for the conditional distribution of time series data. However, it is violated by the classical Cramér–von Mises test statistic; see also their Remark 2. Here we derive the limit distribution under usual degeneracy and under easily verifiable conditions imposed directly on the kernel function.

For bootstrapping degenerate U- or V-statistics of dependent random variables, there are so far only consistency results tailor-made for model-based bootstrap methods; see [26,28]. However, goodness-of-fit tests based on the empirical distribution are most appropriate if no particular model class is available. To the best of our knowledge, the literature does not provide consistency results on model-free bootstrap methods for degenerate U- and V-statistics under dependence. In Section 3 we introduce new variants of the dependent wild bootstrap that are suitable for degenerate U- and V-statistics. For dependent random variables $X_1, \ldots, X_n$ satisfying certain conditions, and a smooth function H, Shao [39] proposed to approximate the distribution of $H(\bar{X}_n) - H(EX_1)$ by that of $H(n^{-1}\sum_{t=1}^n X_t^*) - H(\bar{X}_n)$, where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$, $X_t^* = \bar{X}_n + (X_t - \bar{X}_n)W_{t,n}^*$ $(t = 1, \ldots, n)$ and $(W_{t,n}^*)_{t=1}^n$ are weakly dependent random variables independent of $X_1, \ldots, X_n$. Because of the obvious similarity to Wu's [43] wild bootstrap for independent random variables, Shao called this method the dependent wild bootstrap. The role played by the triangular scheme $(W_{t,n}^*)_{t=1}^n$, $n \in \mathbb{N}$, is easily explained: While a nondegenerate distribution of $W_{t,n}^*$ with $E^*W_{t,n}^* = 0$ introduces the necessary randomness, the condition of $\operatorname{cov}^*(W_{s,n}^*, W_{t,n}^*) \longrightarrow_{n\to\infty} 1$ takes care that the dependence structure of the original process $X_1, \ldots, X_n$ is asymptotically captured. Knowledge of the mechanism producing the limit distribution of a V-statistic $V_n = n^{-1}\sum_{s,t=1}^n h(X_s, X_t)$ helps us to devise a variant of the dependent wild bootstrap appropriate for U- and V-statistics. Under the conditions imposed below, the V-statistic can be rewritten as $V_n = \sum_k \lambda_k \left(n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)\right)^2$, where $(\lambda_k)_k$ are the nonzero eigenvalues and $(\Phi_k)_k$ the corresponding eigenfunctions of a certain integral equation. The random variables $(\Phi_k(X_t))_{t=1}^n$ have zero mean and inherit the property of weak dependence from the original process. The limit distribution of V_n results from joint asymptotic normality of $n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)$, $k \in \mathbb{N}$. Therefore, it is quite a natural attempt to approximate the distribution of V_n by that of $V_n^* = \sum_k \lambda_k \left(n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)\, W_{t,n}^*\right)^2 = n^{-1}\sum_{s,t=1}^n h(X_s, X_t)W_{s,n}^*W_{t,n}^*$. A similar method, based on independent auxiliary variables $(W_t^*)_{t=1,\ldots,n}$, was proposed in [13] for degenerate U-statistics of independent random variables. We prove that the distribution of V_n^* consistently approximates that of V_n. In particular, we require only a very weak condition on the tuning parameter of this method that is obviously a necessary one. An analogous result holds also for U-statistics. Besides the fact that our resampling method is very easy to implement, it is also computationally less intensive than ordinary block bootstrap algorithms which cannot be applied naively to U-statistics but require a modification of the kernel on the bootstrap side; see also Remark 4 in Section 3 for details.

In Section 4 we apply our general results to two particular test problems. We consider a goodness-of-fit test of Cramér–von Mises type and a test of equality of the marginal distributions of two matched samples. Even in the former case where we restrict our attention to tests of simple hypotheses, we cannot simply use simulations to determine an appropriate critical value since the dependence structure is left unspecified. We can also not use tabulated critical values from the independent case since the effect of dependence is not negligible here. It was already pointed out by Gleser and Moore [21] that the null is rejected too often if the quantiles of the i.i.d. setting are used as an approximation for the corresponding quantities under positive dependence. On the other hand, both test statistics can be rewritten as V-statistics that are degenerate under the null hypothesis and it can be seen that the conditions imposed in the Sections 2 and 3 are fulfilled. Hence, our version of the dependent wild bootstrap method provides asymptotically correct critical values.

Section 5 contains a numerical analysis of the proposed bootstrap method for the classical Cramér–von Mises test in the time series context. It can be seen from Fig. 1 that the bootstrap distributions of the test statistic are much more accurate approximations than the distribution from the independent case. Moreover, it turns out that the actual size of the test is not too far from the prescribed one. Some impression on the power is also given by a few examples. We additionally included a real-data application of the two-sample test introduced in Section 4.2.

The proofs of the theorems and some auxiliary results are deferred to a final Section 6. Besides many approximations, a key tool for the asymptotic theory is a multivariate central limit theorem (CLT) for weakly dependent random variables. In case of the original process, a univariate CLT in conjunction with the Cramér–Wold device would obviously do the job. However, on the bootstrap side we have to deal with a triangular scheme and all conditions required for a CLT are only fulfilled in probability. This fact makes a direct application of the Cramér–Wold device much more cumbersome. Therefore, we establish as a by-product a multivariate generalization of the CLT of Neumann [32] that implies then a multivariate bootstrap CLT as a direct consequence. We think that these technical tools are of interest beyond this work.

2. Asymptotic distributions of *U*- and *V*-statistics

Suppose that observations $X_1, \ldots, X_n$ from a strictly stationary process are available. In view of Kolmogorov's consistency theorem there exists a two-sided process with the corresponding finite-dimensional distributions. To simplify the presentation, we assume in the sequel that our observations stem from such a two-sided process $(X_t)_{t\in\mathbb{Z}}$.

In this section, we derive the limit distributions of

$$U_n = \frac{1}{n}\sum_{\substack{s,t=1\\ s\neq t}}^n h(X_s, X_t) \quad \text{and} \quad V_n = \frac{1}{n}\sum_{s,t=1}^n h(X_s, X_t),$$