Statistical Methodology 11 (2013) 36–52

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

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

Strong approximations for weighted bootstrap of empirical and quantile processes with applications

Sergio Alvarez-Andrade, Salim Bouzebda *

Laboratoire de Mathématiques Appliquées, Université de Technologie de Compiègne, B.P. 529, 60205 Compiègne cedex, France

ARTICLE INFO

Article history:
Received 12 April 2012
Received in revised form
27 July 2012
Accepted 9 September 2012

Keywords:
General bootstrap
Brownian bridge
Best approximation
Empirical process
Quantile process
Kernel density estimator

ABSTRACT

The main purpose of this paper is to investigate the strong approximation of the weighed bootstrap of empirical and quantile processes. The bootstrap idea is to reweight the original empirical distribution by stochastic weights. Our results are applied in two concrete statistical problems: the Q–Q processes as well as the kernel-type density estimator. Finally, a general notion of bootstrapped empirical quantile processes, from randomly censored data, constructed by exchangeably weighting samples is presented.

1. Introduction

The present paper is mainly concerned with the strong approximations of the weighted bootstrap empirical and quantile processes. Recall that the bootstrap technique, which is a form of resampling procedures for statistical inference, was introduced in [23]'s seminal paper. The bootstrap may be described briefly as follows. Let $T(F)$ be a functional of an unknown distribution function (df) $F(\cdot)$, $X_1, \ldots, X_n$ a sample from $F(\cdot)$, and $X_1^*, \ldots, X_n^*$ an independent and identically distributed (i.i.d.) sample with common distribution given by the empirical distribution $F_n(\cdot)$ of the original sample. The distribution of $\{T(F_n) - T(F)\}$ is then approximated by that of $\{T(F_n^*) - T(F_n)\}$ conditionally upon $X_1, \ldots, X_n$, with $F_n^*(\cdot)$ being the empirical distribution of $X_1^*, \ldots, X_n^*$. The key idea behind the bootstrap is that if a sample is representative of the underlying population, then one can make inferences about the population characteristics by resampling from the current sample. Roughly speaking, it is known that the bootstrap works in the i.i.d. case if and only if the central limit theorem holds for the random variable under consideration. For further discussion we refer the reader to the

* Corresponding author.
E-mail addresses: sergio.alvarez@utc.fr (S. Alvarez-Andrade), bouzebda@gmail.com, salim.bouzebda@utc.fr (S. Bouzebda).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2012.09.001

landmark paper by Giné and Zinn [28]. The following notation is needed for the statement of our results. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables (rv's) with common df $F(t) = P(X_1 \le t)$. For each $n \ge 1$, the empirical distribution function of $X_1, \ldots, X_n$, is given by

$$F_n(t) = n^{-1}\#\{X_i \le t : 1 \le i \le n\}, \quad \text{for } -\infty < t < \infty,$$

where # stands for cardinality. The quantile function (qf) pertaining to $F(\cdot)$, is defined, for $u \in (0, 1)$, by $Q(u) = \inf\{x : F(x) \ge u\}$. The empirical quantile function is given, for each $n \ge 1$ and $u \in (0, 1)$, by $Q_n(u) = \inf\{x : F_n(x) \ge u\}$. Given the sample $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_m^*$ be conditionally independent rv's with common distribution function $F_n(\cdot)$. Let

$$F_{m,n}(t) = m^{-1}\#\{X_i^* \le t : 1 \le i \le m\}, \quad \text{for } -\infty < t < \infty,$$

denote the *classical* Efron (or multinomial) bootstrap (see, e.g. [23,25] for more details). Consider also the bootstrapped empirical quantile function, belonging to $F_{m,n}(\cdot)$,

$$Q_{m,n}(u) = \inf\{x : F_{m,n}(x) \ge u\}, \quad \text{for } 0 < u < 1.$$

Define the *bootstrapped empirical* and *quantile processes*, respectively, by

$$\xi_{m,n}(t) := m^{1/2}(F_{m,n}(t) - F_n(t)), \quad \text{for } -\infty < t < \infty, \tag{1.1}$$

and

$$\zeta_{m,n}(t) := m^{1/2}(Q_{m,n}(t) - Q_n(t)), \quad \text{for } 0 < t < 1. \tag{1.2}$$

Among many other things, [8] established weak convergence of the processes in (1.1) and (1.2), which enabled them to deduce the asymptotic validity of the bootstrap method in forming confidence bounds for $F(\cdot)$. Shorack [47] gave a simple proof of weak convergence of the process in (1.1) (see also [48, Section 23.1]). The Bickel and Freedman result for $\xi_{m,n}(\cdot)$ has been subsequently generalized for empirical processes based on observations in $\mathbb{R}^d$, $d > 1$, as well as in very general sample spaces and for various set and function-indexed random objects (see, for example [6,27]). This line of research found its "*final results*" in the works of [28,18]. There is a huge literature on the application of the bootstrap methodology to nonparametric kernel density and regression estimation, among other statistical procedures, and it is not the purpose of this paper to survey this extensive literature. This being said, one of the possible drawbacks of [23]'s original bootstrap formulation is that some observations may be used more than once while others are not sampled at all. To overcome this difficulty, a more general formulation of the bootstrap has been introduced: the *weighted* (or *smooth*) bootstrap, which has also been shown to be computationally more efficient in several applications. For a survey of further results and a deeper discussion on weighted bootstrap consult [4,46]. The performance of different kinds of bootstrap procedures is reviewed by Bickel and Freedman [9] in terms of expansions. Exactly as for Efron's bootstrap, the question of rates of convergence is an important one (both in probability and in statistics) and has occupied a great number of authors (see [35,19,16,30] and the references therein).

In this paper, the strong approximations for the weighted bootstrap empirical and quantile processes by a sequence of Brownian bridges will be investigated. The reason for employing the strong approximation theory instead of the weak convergence theory is motivated by their usefulness is probability theory as well as in statistical applications. Precisely, many well-known probability theorems can be considered as consequences of results about strong approximation of sequences of sums by corresponding Gaussian sequences. We shall mention that the rates of convergence for the distribution of *smooth* functionals of the empirical and quantile processes can also be deduced from the strong approximation results. We refer to [35,16], [15, Chapter 3], [19, Chapters 4–5] and [48, Chapter 12] for expositions and references about this problem. To the best of our knowledge, the results presented here respond to a problem that has not been studied systematically until the present, and it gives the main motivation to this paper.

The remainder of the present paper is organized as follows. Section 2 introduces the notation and definitions needed to state the strong approximations of the weighed bootstrap of uniform empirical and quantile processes, which are given in Theorems 3 and 4. Section 3 is devoted to