Journal of Multivariate Analysis 116 (2013) 52–62

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

On general bootstrap of empirical estimator of a semi-Markov kernel with applications

Salim Bouzebda *, Nikolaos Limnios

Laboratoire de mathématiques appliquées-L.M.A.C., Université de technologie de Compiègne, B.P. 529, 60205 Compiègne cedex, France

ARTICLE INFO

Article history:
Received 4 April 2012
Available online 8 December 2012

AMS subject classifications:
60G45
60J28
60K15
60G09
62F40

Keywords:
Semi-Markov processes
Semi-Markov kernel
Empirical estimator
Invariance principle
Bootstrap
Exchangeable bootstrap
Confidence intervals
Change point problem

ABSTRACT

The aim of this paper is to introduce a general bootstrap by exchangeable weight random variables for empirical estimators of the semi-Markov kernels and of the conditional transition probabilities for semi-Markov processes with countable state space. Asymptotic properties of these generalized bootstrapped empirical distributions are obtained by a martingale approach. We show how to apply our results to the construction of confidence intervals and change point problem where the limiting distribution of the proposed statistic is derived under the null hypothesis.

1. Introduction and notations

Semi-Markov processes are an extension of jump Markov processes and renewal processes. More specifically, they allow the use of any distribution for the sojourn times instead of the exponential (geometric) distributions in the Markov processes (chains) case. This feature has led to successful applications in survival analysis [1], reliability [21], queueing theory, finance and insurance [19]. The theory of the semi-Markov processes was given by Pyke [27,28]. For recent references in this area along with statistical applications see, e.g., [22,3]. Nonparametric estimation of semi-Markov kernel and conditional transition probability has been the subject of intense investigation for many years leading to the development of a large variety of methods, see, e.g., [21,3] and the references therein. Note that the limiting distributions of these estimators, or their functionals, are rather complicated, which does not permit explicit computation in practice. To overcome that difficulty, we shall propose a general bootstrap of empirical semi-Markov kernels and of the conditional transition probabilities and study some of its asymptotic properties by means of martingale techniques. The interest in considering a general bootstrap instead of particular cases lies in the fact that we need, in general, a more flexible modeling to handle the problems in practice. In a variety of statistical problems, the bootstrap provides a simple method for circumventing technical difficulties due to intractable distribution theory and has become a powerful tool for setting confidence intervals and critical values of tests for

* Corresponding author.
E-mail addresses: salim.bouzebda@utc.fr, bouzebda@gmail.com (S. Bouzebda), nikolaos.limnios@utc.fr (N. Limnios).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.11.008

composite hypotheses. A substantial body of literature, reviewed in [5], gives conditions for the bootstrap to be satisfied in order to provide desirable distributional approximations. In [8], the performance of different kinds of bootstrap procedures is investigated through asymptotic results and small sample simulation studies. Note that the bootstrap, according to Efron's original formulation (see [17]), presents some drawbacks. Namely, some observations may be used more than once while others are not sampled at all. To overcome that problem, 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 on weighted bootstrap the reader is referred to Barbe and Bertail [2]. Another resampling scheme was proposed in [29] and was extensively studied by Bickel and Freedman [7], who suggested the name "weighted bootstrap", e.g., Bayesian Bootstrap when the vector of weights

$$(W_{n1}, \ldots, W_{nn}) = (D_{n1}, \ldots, D_{nn}),$$

is equal in distribution to the vector of n spacings of $n-1$ ordered uniform $(0, 1)$ random variables, that is,

$$(D_{n1}, \ldots, D_{nn}) \sim \text{Dirichlet}(n; 1, \ldots, 1).$$

The interested reader may refer to Lo [23]. The case

$$(D_{n1}, \ldots, D_{nn}) \sim \text{Dirichlet}(n; 4, \ldots, 4),$$

was considered in [31, Remark 2.3] and [32, Remark 5]. These resampling plans lead to the interest of a unified approach, generically designated as general weighted resampling, was first proposed by Mason and Newton [24] and amongst others extended by Præstgaard and Wellner [26]. We may refer also to Pauly [25], where the weighted bootstrap versions of the linear statistic of arrays of martingale differences were analyzed. To our best knowledge, general weighted resampling in the semi-Markov setting was open, giving the main motivation to our paper. The results obtained in the present paper are useful in many statistical problems, in the semi-Markov framework, as it will be illustrated in the construction of confidence intervals, the change point problem and the computation of the p-value of the statistical test.

We start by giving some notation and definitions that are needed for the forthcoming sections. All random processes are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This is a technical requirement which allows for the construction of the Gaussian processes in our theorems, and is not restrictive since one can expand the probability space to make it rich enough (see e.g., Appendix 2 in [13,16] and [6, Lemma A1]). To define semi-Markov processes or equivalently Markov renewal processes, it is natural, first, to define semi-Markov kernels (see, for example, [22] for a description in details). Consider an infinite countable set, say E, and an E-valued càdlàg time-homogeneous semi-Markov process Z_t, for $t \in \mathbb{R}_+$, with embedded Markov renewal process (J_k, S_k), for $k \in \mathbb{N}$, where (J_k) is the E-valued embedded Markov chain (EMC) of the successive visited states, and

$$0 = S_0 \leq S_1 \leq \cdots \leq S_k \leq S_{k+1} \leq \cdots$$

are the jump times of the semi-Markov process (Z_t). Define also $X_k := S_k - S_{k-1}$, for $k \geq 1$, the sojourn times, and the process $N(t)$, for $t \in \mathbb{R}_+$, which counts the number of jumps of (Z_t), in the time interval $(0, t]$, by

$$N(t) := \sup\{k \geq 0 : S_k \leq t\}.$$

Let us also define $N_i(t)$ as the number of visits of (Z_t) to state $i \in E$ up to time t, and $N_{ij}(t)$ the number of direct jumps of (Z_t) from state i to state j up to time t, (see, e.g., [22]). More precisely,

$$N_i(t) := \sum_{k=1}^{N(t)} \mathbb{1}_{\{J_{k-1}=i\}} \quad \text{and} \quad N_{ij}(t) := \sum_{k=1}^{N(t)} \mathbb{1}_{\{J_k=i, J_{k-1}=j\}},$$

where $\mathbb{1}_A$ stands for the indicator function of the event A. Particularly, in the case where we consider the renewal process $(S_k^i)_{k\geq 0}$ (eventually delayed, see, e.g., [22]) of successive times of visits to state i, then $N_i(t)$ is the counting process of renewals. Let μ_{ii} be the mean recurrence times of (S_n^i), i.e.,

$$\mu_{ii} = \mathbb{E}[S_2^i - S_1^i],$$

and let us denote by $\nu = (\nu_i)_{i\in E}$ the stationary distribution of the embedded Markov chain (J_k). Let $Q(t) = (Q_{ij}(t), i, j \in E)$, for $t \geq 0$, be the semi-Markov kernel, which is defined by

$$\begin{aligned} Q_{ij}(t) &:= \mathbb{P}(J_{k+1} = j, X_{k+1} \leq t | J_k = i) \\ &= P(i,j) F_{ij}(t), \quad t \geq 0,\ i, j \in E, \end{aligned} \tag{1.1}$$

where

$$P(i,j) := \mathbb{P}(J_{k+1} = j | J_k = i),$$

is the transition kernel of the EMC (J_k), and

$$F_{ij}(t) := \mathbb{P}(X_{k+1} \leq t | J_k = i, J_{k+1} = j),$$