Note on the regeneration-based bootstrap for atomic Markov chains

Bertail, Patrice
Laboratoire de Statistiques, CREST

Clémençon, Stéphan
MODAL'X, Université Paris X Nanterre
Laboratoire de Probabilités et Modèles Aléatoires,
UMR CNRS 7599 - Universités Paris VI et VII

Abstract

In this paper, we show how the original Bootstrap method introduced by Datta & McCormick (1993), namely the regeneration-based Bootstrap, for approximating the sampling distribution of sample mean statistics in the atomic Markovian setting may be modified, so as to be second order correct. We prove that the drawback of the original construction mainly relies on a wrong estimation of the skewness of the sampling distribution and that it is possible to correct it by suitable standardization of the regeneration-based bootstrap statistic and recentering of the bootstrap distribution. An asymptotic result establishing the second order accuracy of this bootstrap estimate up to $O(n^{-1}\log(n))$ (close to the rate obtained in an i.i.d. setting) is also stated under weak moment assumptions.

1 Introduction

Among the numerous methods that have been suggested to adapt Efron's Bootstrap to weakly dependent settings, the view underlying the construction proposed in Datta & McCormick (1993) (see also Athreya & Fuh (1989)) for bootstrapping sample mean statistics in the atomic Markovian framework is one of the most interesting ones. Curiously, the beautiful ideas introduced in this paper, based on the renewal properties of Markov chains with an atom, do not seem to be widely known and used in the statistical and econometric Bootstrap literature. This may be partly explained by the fact that they only consider the restrictive case of Markov chains possessing a known atom under rather strong assumptions regarding to ergodicity properties. Moreover, because of some inappropriate standardization, the method proposed in Datta & McCormick (1993) is not second order correct and performs poorly in the applications. The purpose of this paper is to explain why the original regeneration-based Bootstrap procedure fails to be second order accurate on the one hand and to show how it is possible to correct it by some specific standardization and recentering on the other hand. It is noteworthy that the regeneration-based bootstrap, modified in this way, allows to get an accuracy, in the case when the chain is stationary, very close to the one obtained in the i.i.d. setting, which is not the case for other Bootstrap methods introduced to deal with the dependent case (see Götze & Künsch (1996) for instance). For the sake of simplicity, we only focus here on the case of the sample mean statistic renormalized by its true asymptotic variance. In section 2, the atomic Markovian framework we consider is set out and some notations are given for later use. In section 3 a preliminary result is established, which provides an explicit expression for the asymptotic skewness coefficient of the sample mean statistic in our setting. Our proposal, based on this preliminary result, for correcting the original regeneration-based bootstrap is described in section 4. An asymptotic result proving the second order accuracy with a remainder of order $O_P(n^{-1}\log(n))$ of this bootstrap procedure is also stated. The proof is given in section 5.

2 Assumptions and notation

Here and throughout, $X = (X_n)_{n\in\mathbb{N}}$ is a time-homogeneous positive recurrent Markov chain valued in a countably generated state space $(E,\ \mathcal{E})$ with transition probability $\Pi(x, dy)$ and stationary probability distribution $\mu(dy)$ (see Revuz (1984) for an exhaustive treaty of the basic concepts of the Markov chain theory). For any probability distribution ν on $(E,\ \mathcal{E})$ (respectively, for any $x \in E$), let P_ν (resp., P_x) denote the probability on the underlying space such that $X_0 \sim \nu$ (resp., $X_0 = x$) and let $E_\nu(.)$ (resp., $E_x(.)$) denote the P_ν-expectation (resp., the P_x-expectation). In what follows, we will suppose that the underlying probability space is the canonical space of the Markov chain, that is by no means restrictive regarding to our results. Let us assume further that the chain possesses a known accessible atom A, i.e. a measurable set $A \in \mathcal{E}$ such that $\mu(A) > 0$ and for all x, y in A, $\Pi(x, .) = \Pi(y, .)$. We will denote by P_A (respectively, by $E_A(.)$) the probability measure on the underlying space such that $X_0 \in A$ (resp., the P_A-expectation). We denote the consecutive return times to the atom A by

$$\tau_A = \tau_A(1) = \inf\{n \geqslant 1,\ X_n \in A\},$$
$$\tau_A(k+1) = \inf\{n > \tau_A(k),\ X_n \in A\},\ \text{for } k \geqslant 1.$$

Throughout this paper, $I(\mathcal{A})$ denote the indicator function of the event $\mathcal{A}$. In this setting, the stationary distribution μ may be represented as an occupation's measure (see Theorem 17.1.7 in Meyn & Tweedie (1996) for instance): for any $B \in \mathcal{E}$,

$$\mu(B) = E_A(\tau_A)^{-1} E_A(\sum_{i=1}^{\tau_A} I\{X_i \in B\}).$$

Moreover the study of the asymptotic properties of such a Markov chain is made much easier by applying the so-called regenerative method. This consists in dividing its trajectories into "blocks" corresponding to pieces of the sample path between successive visits to the atom A, $\mathcal{B}_k = (X_{\tau_A(k)+1}, ..., X_{\tau_A(k+1)})$, $k \geqslant 1$, and in exploiting the fact that, by virtue of the strong Markov property, the $\mathcal{B}_k$'s are i.i.d. random variables valued in the torus $T = \cup_{n=1}^{\infty} E^n$. In the sequel, we shall denote by $l(\mathcal{B}_k) = \tau_A(k+1) - \tau_A(k)$ the length of the block $\mathcal{B}_k$, $k \geqslant 1$. And for any measurable function $f : E \to \Re$, we will set

$$S_A(f) = \sum_{i=1}^{\tau_A} f(X_i),$$

$$f(\mathcal{B}_j) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i), \text{ for } j \geqslant 1.$$

We point out that the atomic setting includes the whole class of Harris recurrent Markov chains with a countable state space (for which, any recurrent state is an accessible atom), as well as many other specific Markovian models, widely used for modeling storage and queuing systems for instance (refer to Asmussen (1987) for an overview).

3 Preliminary result

Let f be a real valued function defined on the state space $(E, \mathcal{E})$ and set $S_n(f) = \sum_{i=1}^{n} f(X_i)$. Under the assumption that the expectation $E_A(S_A(|f|))$ is finite, the function f is clearly μ-integrable (note that $\mu(f) = E_\mu(f(X_1)) = \mu(A)E_A(S_A(f))$ by the representation of μ using the atom A) and with the additional assumption that the initial probability distribution ν is such that $P_\nu(\tau_A < \infty) = 1$, the regenerative method mentioned above allows to show straightforwardly that $\mu_n(f) = S_n(f)/n$ is a strongly consistent estimator of the parameter $\mu(f)$ under P_ν: $S_n(f)/n \to \mu(f)$ P_ν a.s., as $n \to \infty$. Moreover, under the further assumptions that the expectations $E_A(\tau_A^2)$, $E_\nu(\tau_A)$, $E_A(S_A(|f|^2))$ and $E_\nu(S_A(|f|))$ are finite, the CLT holds too under P_ν:

$$n^{1/2}(S_n(f)/n - \mu(f)) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2), \text{ as } n \to \infty,$$

with a limiting variance $\sigma_f^2 = \mu(A)E_A(S_A(f - \mu(f))^2)$ (see Theorems 17.2.1 and 17.2.2 in Meyn & Tweedie (1996) for instance).

Even if it entails to replace f by $f - \mu(f)$, we assume that $\mu(f) = 0$ in the remainder of this section. The following theorem gives two different forms for the asymptotic skewness of $n^{1/2}(S_n(f)/n)$, which determines the main term in its Edgeworth expansion (see Datta & McCormick (1993b)).

Theorem 1 If the series $\sum_{i\geqslant 1}\{E_\mu(f^2(X_1)f(X_{i+1}))+E_\mu(f(X_1)f^2(X_{i+1}))\}$ and $\sum_{i,\ j\geqslant 1}E_\mu(f(X_1)f(X_{i+1})f(X_{i+j+1}))$ converge absolutely, then we have:

$$\lim_{n\to\infty} n^{-1}E_\mu((S_n(f))^3) = E_\mu(f(X_1)^3) + 3\sum_{i=1}^{\infty}\{E_\mu(f^2(X_1)f(X_{i+1})) + E_\mu(f(X_1)f^2(X_{i+1}))\} + 6\sum_{i,\ j=1}^{\infty} E_\mu(f(X_1)f(X_{i+1})f(X_{i+j+1})). \quad (1)$$

Moreover, if the expectations $E_A(\tau_A^4)$ and $E_A\left(S_A(|f|)^4\right)$ are finite, $\sigma_f^2 > 0$ and $\overline{\lim}_{|t|\to\infty}|E_A(\exp(itS_A(f)))| < 1$, then we have also:

$$\lim_{n\to\infty} n^{-1}E_\mu((S_n(f))^3) = E_A(\tau_A)^{-1}\{E_A(S_A(f)^3) - 3\sigma_f^2 E_A(\tau_A S_A(f))\}. \quad (2)$$

Proof. For all $n \geqslant 1$ we have by stationarity

$$\begin{aligned} n^{-1}E_\mu(S_n(f)^3) &= E_\mu(f(X_1)^3) \\ &+ 3n^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E_\mu(f(X_1)^2 f(X_{j-i+1})) + E_\mu(f(X_1)f(X_{j-i+1})^2) \\ &+ 6n^{-1}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} E_\mu(f(X_1)f(X_{j-i+1})f(X_{k-j+1})) \\ &= E_\mu(f(X_1)^3) + \\ &3\frac{n-1}{n}\sum_{l=1}^{n} E_\mu(f(X_1)^2 f(X_{l+1})) + E_\mu(f(X_1)f(X_{l+1})^2) + \\ &6\frac{n-2}{n}\sum_{l=1}^{n-1}\sum_{m=1}^{n} E_\mu(f(X_1)f(X_{l+1})f(X_{m+l+1})) \end{aligned}$$

and thus one clearly gets (1) from the convergence of the right hand side as $n \to \infty$. Besides, under the assumption that the "block" moment conditions

$$E_A(\tau_A^4) < \infty,\ E_A\left(S_A(|f|)^4\right) < \infty$$

are fulfilled, as well as the Cramer condition $\overline{\lim}_{|t|\to\infty}|E_A(\exp(itS_A(f)))| < 1$, (2) straightforwardly results from Theorem 3 in Malinovskii (1987) (see also Theorem 3 in Malinovskii (1985)). As a matter of fact, according to this result, for any initial probability ν such that $E_\nu(\tau_A^2) < \infty$ and $E_\nu(S_A(|f|)^2) < \infty$, we have as $n \to \infty$

$$n^{-1}E_\nu(S_n(f))^3) = \alpha^{-1}\{E_A(S_A(f)^3) - 3\sigma_f^2(2\beta(f) - \alpha\eta_\nu(f)\} + O(n^{-1/2}),$$

with $\alpha = E_A(\tau_A)$, $\eta_\nu(f) = E_\nu(S_A(f)) + \alpha^{-1}E_A(\sum_{i=1}^{\tau_A}(\tau_A - i)f(X_i))$ and $\beta(f) = E_A(\tau_A S_A(f))$. The assumptions $E_A(\tau_A^4) < \infty$ and $E_A\left(S_A(|f|)^4\right) <$

∞ ensure thus that $E_\mu(\tau_A^2) < \infty$ and $E_\mu(S_A(|f|)^2) < \infty$. As a matter of fact, by the representation of the stationary probability measure using the atom A (i.e. $\mu(B) = \alpha^{-1}E_A(\sum_{i=1}^{\tau_A} I\{X_i \in B\})$, for all $B \in \mathcal{E}$, and $E_\mu(H(X)) = \int \mu(dx)E_x(H(X))$ for any measurable function H defined on the canonical space), the following stronger bounds hold. We have

$$E_\mu(\tau_A^3) = \alpha^{-1}E_A(\sum_{i=1}^{\tau_A} E_{X_i}(\tau_A)^3).$$

Therefore, on the event $\{i \leqslant \tau_A\}$, we have $\theta^i \circ \tau_A = \tau_A - i$, denoting by θ the shift operator on the canonical space. Hence we deduce from the Markov property that

$$E_\mu(\tau_A^3) = \alpha^{-1}E_A(\sum_{i=1}^{\tau_A} E((\tau_A - i)^3 \mid \mathcal{F}_i^X)),$$

where $\mathcal{F}_i^X$ denotes the σ-field generated by the X_j's, for $j \leqslant i$. Finally, we have

$$E_\mu(\tau_A^3) < \alpha^{-1}E_A(\sum_{i=1}^{\tau_A}(\tau_A - i)^3) < \alpha^{-1}E_A(\tau_A^4) < \infty.$$

In a similar fashion, we have

$$\begin{aligned} E_\mu(S_A(|f|)^3) &= \alpha^{-1}E_A(\sum_{i=1}^{\tau_A}(\sum_{j=i}^{\tau_A}|f(X_i)|)^3) \leqslant \alpha^{-1}E_A(\tau_A S_A(f)^3) \\ &\leqslant \alpha^{-1}(E_A(\tau_A^4))^{1/4}E_A(S_A(f)^4)^{3/4} < \infty \end{aligned}$$

by using Hölder inequality. Moreover, we have $\alpha\eta_\mu(f) = \beta(f)$. By the representation of the stationary distribution μ using the atom A again, we have

$$\begin{aligned} \alpha E_\mu(\sum_{i=1}^{\tau_A} f(X_i)) &= E_A(\sum_{i=1}^{\tau_A} E_{X_i}(\sum_{k=1}^{\tau_A} f(X_k))) = E_A(\sum_{1\leqslant i<k\leqslant\tau_A} f(X_k)) \\ &= E_A(\sum_{k=1}^{\tau_A}(k-1)f(X_k)) = E_A(\sum_{k=1}^{\tau_A} kf(X_k)), \end{aligned}$$

since we assumed $\mu(f) = \alpha^{-1}E_A(\sum_{k=1}^{\tau_A} f(X_k)) = 0$. This allows to deduce straightforwardly that $\alpha\eta_\mu = \beta$ and so to prove (2). ■

We point out that it is possible to take advantage of the identity (2) for constructing an empirical estimate of the asymptotic skewness based on the regenerative blocks (refer to Bertail & Clémençon (2003) and the proof of the main theorem below). Besides, it is noteworthy that the asymptotic skewness, namely $k_{3,f} = \sigma_f^{-3}\lim_{n\to\infty} n^{-1}E_\mu(S_n(f)^3)$, generally differs from $\alpha^{-1}\sigma_f^{-3}E_A((S_A(f)^3)$. Under the assumptions of Theorem 1, whereas $\sum_{k\geqslant1}\{f^2(X_k) - E_\mu(f^2(X_k))\}$ converges absolutely under P_μ, we have that $\sum_{k=1}^{\infty} f^2(X_k) = \infty$, P_μ a.s. (note that $f^2(X_k)$ is not centered under P_μ). Thus, this crucial observation shows that exchanging the expectation and the summation in $\sum_{i=1}^{\infty} E_\mu(f(X_1)f^2(X_{i+1}))$, as done in Datta & McCormick (1993a), is not possible. Observe further that such an illicit operation would allow to derive the false identity claiming that the sum $E_\mu(f(X_1)^3)+3\sum_{i=1}^{\infty}\{E_\mu(f^2(X_1)f(X_{i+1}))+E_\mu(f(X_1)f^2(X_{i+1}))\}+6\sum_{i,\,j=1}^{\infty} E_\mu(f(X_1)f(X_{i+1})f(X_{i+j+1}))$ equals to the term $\alpha^{-1}E_A((S_A(f)^3)$, (or equivalently that $k_{3,f}$ equals to $\alpha^{-1}\sigma_f^{-3}E_A((S_A(f)^3))$, on which the argument of Datta & McCormick (1993a), for studying the second order properties of the regeneration-based bootstrap methodology they introduced, is based. As will be shown precisely in the next section, this particularly entails that the regeneration-based bootstrap estimate of the sampling distribution of the sample mean statistic $\mu_n(f) = S_n(f)/n$, originally proposed by Datta & McCormick (1993a) has an Edgeworth expansion, that does not match with the expansion of $\mu_n(f)$ (which is due to the skewness term), and consequently is not second order accurate.

4 The stationary regenerative block-bootstrap

The preliminary result of section 3 clearly advocates for the following modification of the regeneration-based bootstrap procedure introduced by Datta & McCormick (1993a) to deal with atomic Markov chains, which we call the stationary regenerative block-bootstrap (SRBB). For estimating the sampling distribution $H_\mu(x) = P_\mu(n^{1/2}\sigma_f^{-1}(\mu_n(f) - \mu(f)) \leqslant x)$ of the studentized sample mean statistic (see below the definition of the asymptotic variance estimator $\sigma_n^2(f)$) computed from observations $X_1, ..., X_n$ drawn from a stationary version of the chain X, it is performed in four steps as follows.

1. Count the number of visits $l_n = \sum_{i=1}^{n} I\{X_i \in A\}$ to the atom A up to time n. And divide the observed sample path $X^{(n)} = (X_1, ..., X_n)$ into $l_n + 1$ blocks, valued in the torus $T = \cup_{n=1}^{\infty} E^n$, corresponding to the pieces of the sample path between consecutive visits to the atom A:

$$\mathcal{B}_0 = (X_1, ..., X_{\tau_A(1)}),\ \mathcal{B}_1 = (X_{\tau_A(1)+1}, ..., X_{\tau_A(2)}), ...,$$
$$\mathcal{B}_{l_n-1} = (X_{\tau_A(l_n-1)+1}, ..., X_{\tau_A(l_n)}),\ \mathcal{B}_{l_n}^{(n)} = (X_{\tau_A(l_n)+1}, ..., X_n).$$

2. Draw an array of $l_n - 1$ bootstrap data blocks $(\mathcal{B}_{1,n}^*, ..., \mathcal{B}_{l_n-1,n}^*)$ independently from the empirical distribution $F_n = (l_n - 1)^{-1}\sum_{i=1}^{l_n-1} \delta_{\mathcal{B}_i}$ of the blocks $\mathcal{B}_1, ..., \mathcal{B}_{l_n-1}$, conditioned on $X^{(n)}$. Practically the bootstrap blocks are taken with replacement from the primary blocks.

3. From the bootstrap data blocks generated at step 2, reconstruct a pseudo-trajectory by binding the blocks together, getting the reconstructed SRBB sample path

$$X^{*(n)} = (\mathcal{B}_{0,n}^*,\ \mathcal{B}_{1,n}^*, ..., \mathcal{B}_{l_n-1,n}^*,\ \mathcal{B}_{l_n,n}^*).$$

with

$$\mathcal{B}^*_{0,n} = \mathcal{B}_0 \text{ and } \mathcal{B}^*_{l_n,n} = \ \mathcal{B}^{(n)}_{l_n}.$$

Whereas the number of blocks $l_n - 1$ is fixed (conditionally to the data sample), the length of the reconstructed segment $(\mathcal{B}^*_{1,n}, ..., \mathcal{B}^*_{l_n-1,n})$ of the pseudo-trajectory is random. We denote by $n^* = \sum_{i=1}^{l_n-1} l(\mathcal{B}^*_{1,n})$ the length of this segment.

4. Compute the SRBB statistic, with the usual convention regarding to empty summation,

$$S^*_{n^*}(f) = \sum_{j=0}^{l_n} f(\mathcal{B}^*_{j,n}),$$

and the following estimate of σ_f^2,

$$\sigma_n^2(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{j=1}^{l_n-1} (f(\mathcal{B}_j) - \widetilde{\mu}_n(f) l(\mathcal{B}_j))^2,$$

with $\widetilde{\mu}_n(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{j=1}^{l_n-1} f(\mathcal{B}_j)$. Then, the recentered distribution of

$$t^*_n = n^{*-1/2} \frac{S^*_{n^*}(f) - S_n(f)}{\sigma_n(f)},$$

conditioned on $X^{(n)}$, is the SRBB distribution

$$H_{SRBB}(x) = P^*(t^*_n - E^*(t^*_n | X^{(n)}) \leq x \mid X^{(n)})$$

where $P^*(. \mid X^{(n)})$ (respectively, $E^*(. \mid X^{(n)})$) denotes the conditional probability (resp., the conditional expectation) given $X^{(n)}$.

• We point out that the bootstrap estimator $H_{SRBB}(x)$ of $H_\mu(x)$ differs from the bootstrap estimator originally proposed by Datta & McCormick (1993a) in two ways. First, the standardization of the bootstrap statistic depends on the random length n^* of the reconstructed bootstrap data segment, whereas the standardization $n^{-1/2}(S^*_{n^*}(f) - S_n(f))/\sigma_n(f)$ is used in Datta & McCormick (1993a). Secondly, the bootstrap distribution is recentered so as to be unbiased. As will be shown below, this random standardization actually allows to recover the correct skewness coefficient $k_{3,f}$ at the price of an additional bias, that may be rectified by recentering suitably the statistic t^*_n of interest (observe that, because of the random standardization by $n^{*-1/2}$, recentering the distribution does not amount to recenter the SRBB statistic $S^*_{n^*}(f)$).

• The construction of the estimator $\sigma_n^2(f)$ naturally relies on the expression $\sigma_f^2 = \mu(A) E_A((S_A(f) - \mu(f)\tau_A)^2)$ for the asymptotic variance, its properties are studied in Bertail & Clémençon (2003). Besides, we have not used the first and last (non regenerative) data blocks $\mathcal{B}_0$ and $\mathcal{B}^{(n)}_{l_n}$ in the computation of our estimate $\sigma_n^2(f)$, because this would make its study much more tricky, while being all the same from the estimation point of view in the stationary framework considered here.

• We also emphasize that one may naturally compute a Monte-Carlo approximation to $H_{SRBB}(x)$ by the following scheme: repeat independently the procedure above Q times, so as to generate $t^*_{n,1}$, ..., $t^*_{n,Q}$, and compute

$$H^{(Q)}_{SRBB}(x) = Q^{-1} \sum_{q=1}^{Q} I\{t^*_{n,q} - Q^{-1} \sum_{p=1}^{Q} t^*_{n,p} \leqslant x\}.$$

The following theorem establish the second order validity of the SRBB estimator up to order $O_P(n^{-1}\log(n))$, which is close to the rate $O_P(n^{-1})$ that can be obtained in the i.i.d case.

Theorem 2 Assume that the chain X fulfills the following conditions.
(i) Conditional Cramer condition:

$$\overline{lim}_{t\to\infty} E_A|E_A(\exp(itS_A(f))|\tau_A)| < 1.$$

(ii) Non degenerate asymptotic variance: $\sigma^2_f > 0$.
(iii) "Block" moment conditions:

$$E_A(\tau_A^4) < \infty,\ E_A(S_A(|f|)^6) < \infty.$$

(iv) Non trivial regeneration set: $E_A(\tau_A) > 1$.
Then, the following Edgeworth expansion is valid uniformly over R,

$$\Delta_n = \sup_{x\in R} |H_\mu(x) - E_n^{(2)}(x)| = O(n^{-1}), \text{ as } n \to \infty, \tag{3}$$

with

$$\begin{aligned} E_n^{(2)}(x) &= \Phi(x) - n^{-1/2}\frac{k_{3,f}}{6}(x^2-1)\phi(x), \\ k_{3,f} &= E_A(\tau_A)^{-1}\{E_A(S_A(f)^3) - 3\sigma_f^2 E_A(\tau_A S_A(f))\}/\sigma_f^3. \end{aligned}$$

And the SRBB estimator is second order accurate in the following sense

$$\Delta_n^S = \sup_{x\in R} |H_{SRBB}(x) - H_\mu(x)| = O_{P_\mu}(n^{-1}\log(n)) \tag{4}$$

uniformly over R, as $n \to \infty$.

• The proof essentially relies on the Edgeworth expansion (E.E. in abbreviated form) obtained in Malinovskii (1987). And dealing with the Bootstrap part mainly reduces to study the E.E. of a bootstrapped V-statistic of degree 2 based on i.i.d. r.v.'s (the bootstrap blocks). The validity of E.E. for V-statistics has been proved in Götze (1979), Bickel, Götze & van Zwet (1986) for instance. The accuracy of the Bootstrap for U-statistics of degree 2 is easy to obtain up to $o_P(n^{-1/2})$. But further conditional Cramer conditions are generally assumed to check the validity up to $O_P(n^{-1})$. Here we use the results of Lai & Wang (1993), proving the validity of the Bootstrap of U-V statistics up to $O_P(n^{-1})$, under conditions which reduce to the conditional Cramer condition (i) in our case. The validity of the SRBB under weaker Cramer conditions will be investigated elsewhere.

• When f is bounded, (iii) reduces to the condition $E_A(\tau_A^6) < \infty$, which typically holds as soon as the strong mixing coefficients sequence decreases at a polynomial rate $n^{-\rho}$ for some $\rho > 5$ (see Bolthausen (1982)).

5 Proof of the main result

The proof of the E.E. (3) for the non studentized sample mean may be found in Malinovskii (1987) (see Theorem 1, refer also to Bertail & Clémençon (2003)). Notice that the conditional Cramer condition implies the usual Cramer condition $\overline{\lim}_{|t|\to\infty}|E_A(\exp(itS_A(f)))| < 1$ and that the bias vanishes in the stationary case. Consider the recentered variables for $j \geqslant 1$,

$$F(\mathcal{B}_j) = f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j),$$
$$F(\mathcal{B}_{j,n}^*) = f(\mathcal{B}_{j,n}^*) - \widetilde{\mu}_n(f)l(\mathcal{B}_{j,n}^*).$$

Notice that the mean length of the bootstrap data blocks $\mathcal{B}_{j,n}^*$, $j \geqslant 1$, for given $X^{(n)}$ is

$$\overline{l_B} \underset{def}{=} E^*(l(\mathcal{B}_{j,n}^*) \mid X^{(n)}) = (l_n - 1)^{-1} \sum_{k=1}^{l_n - 1} l(\mathcal{B}_k),$$

and observe further that $E^*(F(\mathcal{B}_{j,n}^*) \mid X^{(n)}) = 0$ and

$$V^*(F(\mathcal{B}_{j,n}^*) \mid X^{(n)}) = \frac{1}{l_n - 1} \sum_{k=1}^{l_n - 1} F(\mathcal{B}_k)^2 = \overline{l_B}\sigma_n^2(f) =_{def} \widehat{\sigma}_F^2,$$

denoting by $V^*(. \mid X^{(n)})$ the conditional variance for given $X^{(n)}$. Note that the empirical estimator $\sigma_n^2(f)$ of the asymptotic variance is essentially a bootstrap estimator of the variance of the recentered blocks, rescaled by an estimator of $E_A(\tau_A)$, namely $\overline{l_B}$. The following technical results will be useful in the proof. Lemma 3 is a standard result due to Chibisov (1972).

Lemma 3 Assume that W_n admits an E.E. on the normal distribution up to $O(n^{-1}\log(n)^\delta)$, $\delta > 0$, as $n \to \infty$. Assume further that R_n is such that, for some $\eta > 0$, $P(n|R_n| > \eta\log(n)^\delta) = O(n^{-1}\log(n)^\delta)$ or $O(n^{-1})$ as $n \to \infty$, then $W_n + R_n$ has the same E.E. as W_n up to $O(n^{-1}\log(n)^\delta)$.

Lemma 4 Under the hypotheses of Theorem 2, we have for some constant $\eta > 0$,

$$P_\mu(n^{1/2}\left|nl_n^{-1} - \alpha\right| \geqslant \eta(\log(n))^{1/2}) = O\left(n^{-1}\right), \text{ as } n \to \infty. \qquad (5)$$

Proof. Following the argument given in Clémençon (2001) based on the Fuk & Nagaev's inequality for sums of independent unbounded r.v.'s (see also Theorem 6.1 in Rio (2000) for a proof based on block mixing techniques), there exists constants c_0 and c_1 such that the following probability inequality holds for all n,

$$P_\mu(\left|l_n/n-\alpha^{-1}\right|\geqslant x)\leqslant c_0\{\exp(-\frac{nx^2}{c_1+xy})+ nP_A(\tau_A>y)+P_A(\tau_A>n/2)+P_\mu(\tau_A>n/2)\}.$$

On the one hand, choosing $y=n^{1/2}$ and bounding the last three terms at the right hand side by Chebyshev's inequality (given that the expectations $E_A(\tau_A^2)$ and $E_\mu(\tau_A)$ are finite), one gets that, for a constant $\zeta>0$ large enough

$$P_\mu(\left|l_n/n-\alpha^{-1}\right|\geqslant\zeta)=O(n^{-1}),\text{ as } n\to\infty, \tag{6}$$

and on the other hand with the choice $x=\eta(\log n/n)^{1/2}$ and $y=(n/\log n)^{1/2}$ and using Chebyshev's inequality again (given that the expectations $E_A(\tau_A^4)$ and $E_\mu(\tau_A)$ are finite), one obtains that

$$P_\mu(n^{1/2}\left|l_n/n-\alpha^{-1}\right|\geqslant\eta(\log n)^{1/2})=O(n^{-1}),\text{ as } n\to\infty. \tag{7}$$

Now, by combining bounds (6) and (7) , the proof is finished by straightforward calculations. ■

Notice first that, because of the recentering of $S^*_{n^*}(f)$ by the original statistic $S_n(f)$, the data of the first and last (non regenerative) blocks $\mathcal{B}_0$ and $\mathcal{B}^{(n)}_{l_n}$ disappear in the numerator. Hence, we may rewrite the bootstrap version of the studentized sample mean the following way

$$t_n^*=\frac{\sum_{j=1}^{l_n-1}F(\mathcal{B}^*_{j,n})}{(\sum_{j=1}^{l_n-1}l(\mathcal{B}^*_{j,n}))^{1/2}\sigma_n(f)} =\frac{\sum_{j=1}^{l_n-1}\{f(\mathcal{B}^*_{j,n})-\widetilde{\mu}_n(f)l(\mathcal{B}^*_{j,n})\}}{(l_n-1)^{1/2}\,(1+L_n^*)^{1/2}\,\widehat{\sigma}_F} \tag{8}$$

with

$$L_n^*=\overline{l_B}^{-1}\{(l_n-1)^{-1}\sum_{j=1}^{l_n-1}l(\mathcal{B}^*_{j,n})-\overline{l_B}\}$$

Using standard bootstrap results in the i.i.d. case (see Singh (1981) for the lattice case), we have for a constant $\eta>0$ large enough,

$$P^*((l_n-1)L_n^{*2}>\eta\log(l_n)\mid X^{(n)})=O(l_n^{-1}),\text{ as } n\to\infty.$$

It follows from lemma (3) with $\delta=1$ that up to $O(l_n^{-1}\log(l_n))$, we can linearize (8) and the problem reduces to find the E.E. of

$$\widetilde{t}_n^*=(l_n-1)^{-1/2}\widehat{\sigma}_F^{-1}\left\{\sum_{j=1}^{l_n-1}F(\mathcal{B}^*_{j,n})\{1-\frac{1}{2}(l_n-1)^{-1}\sum_{k=1}^{l_n-1}(l(\mathcal{B}^*_{k,n})-\overline{l_B})/\overline{l_B})\}\right\}$$

This may be seen as a bootstrapped V-statistic of degree 2 based on the i.i.d. blocks $\mathcal{B}^*_{j,n}$, $j \geqslant 1$ or . The main (linear) part of the corresponding U-statistic is $F(\mathcal{B}^*_{j,n})/\widehat{\sigma}_F^{-1}$, the (degenerate) quadratic term is given by

$$\beta_n(\mathcal{B}^*_{j,n},\mathcal{B}^*_{k,n}) = \frac{1}{2}\widehat{\sigma}_F^{-1}\left\{F(\mathcal{B}^*_{j,n})\frac{l(\mathcal{B}^*_{k,n})-\overline{l_B}}{\overline{l_B}}) + F(\mathcal{B}^*_{k,n})\frac{l(\mathcal{B}^*_{j,n})-\overline{l_B}}{\overline{l_B}}\right\}.$$

The validity of the Bootstrap for general U or V statistics is proved in Lai & Wang (1993), up to $O_P(n^{-1})$ under assumptions on the second gradient of the U-statistics, which are easier to check than the usual conditional Cramer conditions or conditions on the eigenvalues of the second order gradient of the U-statistic (see also Bickel, Götze & van Zwet (1986)). The conditional Cramer condition used here implies their Cramer type condition (see p 521 of their paper, as well as related results in Bai & Rao (1991) for the validity of E.E. under conditional Cramer type conditions). Using the arguments in Lai & Wang (1993), one may thus check that, conditionally to $X^{(n)}$, $\widetilde{t}^*_n$ admits up to $O(l_n^{-1}\log(l_n))$ an E.E. of the form (see also Barbe & Bertail (1995) for the form of the E.E. up to $o_P(n^{-1/2})$),

$$\begin{aligned} P^*\left(t^*_n \le x \mid X^{(n)}\right) = \Phi(x) &- \frac{\Phi^{(3)}(x)}{6\sqrt{l_n-1}}\{\frac{1}{l_n-1}\sum_{j=1}^{l_n-1}\frac{\{f(\mathcal{B}_j)-\widetilde{\mu}_n(f)l(\mathcal{B}_j)\}^3}{\widehat{\sigma}_F^3}\} \\ &- \frac{x\Phi^{(2)}(x)}{2\sqrt{l_n-1}}\{\frac{1}{l_n-1}\sum_{j=1}^{l_n-1}\{f(\mathcal{B}_j)-\widetilde{\mu}_n(f)l(\mathcal{B}_j)\}(\frac{l(\mathcal{B}_j)-\overline{l_B}}{\widehat{\sigma}_F\,\overline{l_B}})\} \\ &+ O(l_n^{-1}\log(l_n)). \end{aligned} \tag{9}$$

Now from lemma 4, we obtain (unconditionally) as $n \to \infty$,

$$\frac{1}{(l_n-1)^{1/2}} = \frac{E_A(\tau_A)^{1/2}}{n^{1/2}} + O_{P_\mu}(n^{-1}\log(n)^{1/2}), \tag{10}$$

and similarly

$$l_n^{-1}\log(l_n) = O_{P_\mu}(n^{-1}\log(n)). \tag{11}$$

Now under assumption (iii), by the SLLN and the CLT for the i.i.d. blocks we have as $n \to \infty$ (see also Bertail & Clémençon (2003))

$$\frac{1}{l_n-1}\sum_{j=1}^{l_n-1}\frac{\{f(\mathcal{B}_j)-\widetilde{\mu}_n(f)l(\mathcal{B}_j)\}^3}{\widehat{\sigma}_F^3} = \frac{E_A(S_A(f)^3)}{E_A(\tau_A)^{3/2}{\sigma_f}^3} + O_{P_\mu}(n^{-1/2}) \tag{12}$$

and

$$\frac{1}{l_n-1}\sum_{j=1}^{l_n-1}(f(\mathcal{B}_j)-\widetilde{\mu}_n(f)l(\mathcal{B}_j))(\frac{l(\mathcal{B}_j)-\overline{l_B}}{\widehat{\sigma}_F\,\overline{l_B}}) = E\tau_A^{-3/2}{\sigma_f}^{-1}\beta + O_{P_\mu}(n^{-1/2}), \tag{13}$$

as $n \to \infty$, provided that the denominator is defined, which is the case as soon as $l_n > 1$. Therefore we have $P_\mu(l_n \le 1) = O(n^{-1})$ as $n \to \infty$ (see lemma 4 for instance) and combining the conditional E.E. (9) with the approximations (10), (11), (12), (13), it follows that the Bootstrap distribution has in P_μ probability an E.E. of the form

$$\begin{aligned} P^*\left(t_n^* \leq x \mid X^{(n)}\right) &= \Phi(x) - n^{-1/2}k_{3,f}(x^2-1)\phi(x) \\ &\quad + n^{-1/2}E_A(\tau_A)^{-1}\sigma_f{}^{-1}\beta/2\,\phi(x) + O_{P_\mu}(n^{-1}\log(n)). \end{aligned}$$

Notice the bias $n^{-1/2}E_A(\tau_A)^{-1}\sigma(f)^{-1}\beta/2$ which appears because of the random standardization. Recentering by the conditional expectation of t_n^* given $X^{(n)}$ immediately leads to the asymptotic result (4) of Theorem 2.

References

[1] Asmussen, S. (1987). Applied Probability and Queues. John Wiley & Sons, NY.

[2] Athreya, K.B., Fuh, C.D. (1989). Bootstrapping Markov chain: countable case. Tech Rep. 89-7, Institute of Statistical Science, Academia Sinica, Taiwan.

[3] Bai, Z.D., Rao, C.R. (1991). Edgeworth expansion of a function of sample means. Ann. Statist., **19**, 1295-1315.

[4] Barbe, Ph., Bertail, P. (1995). The Weighted Bootstrap. Lecture Notes in Statistics, **98**, Springer Verlag, New-York.

[5] Bertail, P., Clémençon, S. (2003). Edgeworth expansions for suitably normalized sample mean statistics for atomic Markov chains. Submitted for publication.

[6] Bickel, P.J., Götze, F. , van Zwet , W. R. (1986). The Edgeworth expansion for U-statistics of degree 2. Ann. Statist. , **14**, 1463-1484.

[7] Bolthausen, E. (1982). The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains. Z. Wahrsch. Verw. Gebiete, **60**, 283-289.

[8] Chibisov, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. Theory Probab. Appl., **17**, 620-630.

[9] Clémençon, S. (2001). Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. Stat. Prob. Letters, **55**, 227-238.

[10] Datta, S., McCormick W.P., (1993a). Regeneration-based bootstrap for Markov chains. The Canadian Journal of Statistics, **21**, No.2, 181-193.

[11] Datta, S., McCormick W.P., (1993b). On the first-Order Edgeworth Expansion for a Markov Chain. J. Mult. Anal. , **44**, 345-359.

[12] Götze, F. (1979). Asymptotic expansions for von-Mises functionals. Zeit. Warsh. Verw. Gebiete, **50**, 333-355.

[13] Götze, F., Künsch, H.R. (1996), Second order correctness of the blockwise bootstrap for stationary observations, Ann. Statist., **24**, 1914-1933.

[14] Malinovskii, V. K. (1985). On some asymptotic relations and identities for Harris recurrent Markov Chains. Statistics and Control of Stochastic Processes, 317-336.

[15] Malinovskii, V. K. (1987). Limit theorems for Harris Markov chains I. Theory Prob. Appl., **31**, 269-285.

[16] Meyn, S.P., Tweedie, R.L., (1996). Markov chains and stochastic stability. Springer, Berlin.

[17] Lai, T. L. , Wang, J. Q. (1993). Edgeworth expansions for symetric statistics with applications to bootstrap methods. Statistica Sinica, **3**, 517-542.

[18] Revuz, D (1984). Markov chains. North-Holland, 2nd edition.

[19] Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendantes. Springer Verlag.

[20] Singh, K. (1981). On the asymptotic accuracy of Efron's Bootstrap. Ann. Statist. , **9**, 1187-1195.