Bernoulli **12**(4), 2006, 689–712

Regenerative block bootstrap for Markov chains

PATRICE BERTAIL[1] and STÉPHAN CLÉMENÇON[2,3]

[1]*CREST, Timbre J340, 3 ave. Pierre Larousse, 94240 Malakoff Cedex, France. E-mail: patrice.bertail@ensae.fr*
[2]*MODAL'X, Université Paris X, 200 avenue de la République, 92000 Nanterre, France*
[3]*Laboratoire de Probabilités et Modèles Aléatoires, UMR CNRS 7599, Université Paris VI, 4 place Jussieu, 75252 Paris Cedex 05, France. E-mail: clemençon@ccr.jussieu.fr*

A specific bootstrap method is introduced for positive recurrent Markov chains, based on the regenerative method and the Nummelin splitting technique. This construction involves generating a sequence of approximate pseudo-renewal times for a Harris chain X from data $X_1, \ldots, X_n$ and the parameters of a minorization condition satisfied by its transition probability kernel and then applying a variant of the methodology proposed by Datta and McCormick for bootstrapping additive functionals of type $n^{-1}\sum_{i=1}^{n} f(X_i)$ when the chain possesses an atom. This novel methodology mainly consists in dividing the sample path of the chain into data blocks corresponding to the successive visits to the atom and resampling the blocks until the (random) length of the reconstructed trajectory is at least n, so as to mimic the renewal structure of the chain. In the atomic case we prove that our method inherits the accuracy of the bootstrap in the independent and identically distributed case up to $O_{\mathbb{P}}(n^{-1})$ under weak conditions. In the general (not necessarily stationary) case asymptotic validity for this resampling procedure is established, provided that a consistent estimator of the transition kernel may be computed. The second-order validity is obtained in the stationary case (up to a rate close to $O_{\mathbb{P}}(n^{-1})$ for regular stationary chains). A data-driven method for choosing the parameters of the minorization condition is proposed and applications to specific Markovian models are discussed.

Keywords: bootstrap; Edgeworth expansion; Markov chain; Nummelin splitting technique; regenerative process

1. Introduction

In the statistical literature there has been substantial interest in transposing the naive bootstrap method (Efron, 1979) introduced in the independent and identically distributed (i.i.d.) setting to dependent settings. The now well-known idea of the *moving block bootstrap* (MBB) is to resample (overlapping or disjoint) blocks of observations to capture the dependence structure of the observations (see Lahiri 2003, for a recent survey and exhaustive references). However, as noticed by many authors, the results obtained by using this method are not completely satisfactory for the following reasons. First, the MBB approach usually requires stationarity for the observations and generally fails in a general non-stationary framework. Secondly, the asymptotic behaviour of the MBB distribution crucially depends on the estimation of the bias and of the asymptotic variance of the

statistic of interest, which makes it difficult to apply in practice (see Götze and Künsch 1996; Lahiri 2003). From a theoretical viewpoint, the rate of convergence of the MBB distribution is much slower than that of the bootstrap in the i.i.d. case: at best it is of order $O_{\mathbb{P}}(n^{-3/4})$ under restrictive conditions, stipulating the finiteness of moments at any order and an exponential rate for the decay of the strong mixing coefficients, while the bootstrap achieves $O_{\mathbb{P}}(n^{-1})$ in the i.i.d. setting. Finally, the choice of the size of the blocks is a key point in achieving an accurate estimation: this practical problem still remains open in the general case.

Recently, various authors have been interested in bootstrapping some particular types of Markov chain (see Lahiri 2003; Franke *et al.* 2002; and the references therein). However, second-order results in this framework are scarcely available, except in Horowitz (2003) for discrete chains. Unfortunately, these results are weakened by the unrealistic technical assumptions (m-dependence) made on the Markovian models considered. Most bootstrap methods proposed in the literature are asymptotically equivalent at first order. Obtaining their exact rate of convergence is thus of prime importance in helping practitioners to choose a particular bootstrap technique. Our goal is to propose a specific bootstrap method based on the renewal properties of Markov chains, which almost achieves the same rate as that in the i.i.d. case in a general (possibly non-stationary) framework.

This method originates from Athreya and Fuh (1989) and Datta and McCormick (1993), and exploits the regeneration properties of Markov chains when a (recurrent) state is visited infinitely often. The main idea underlying the method is to resample a deterministic number of data blocks corresponding to regeneration cycles. However, because of some inadequate standardization, the *regeneration-based bootstrap* method proposed in Datta and McCormick (1993) is not second-order correct (its rate is $O_{\mathbb{P}}(n^{-1/2})$ only). Bertail and Clémençon (2005) have proposed a modification of the procedure introduced by Datta and McCormick (1993), which is second-order correct up to $O_{\mathbb{P}}(n^{-1}\log(n))$ in the unstudentized case (i.e. when the variance is known) when the chain is stationary. However, this method fails to be second-order correct in the non-stationary case, as a careful look at the Edgeworth expansion (EE) of the statistic of interest shows (see Bertail and Clémençon 2004, 2005). As a matter of fact, the first cycle and the randomness of the number of cycles in a finite-length trajectory play a crucial role in this asymptotic expansion. To avoid the problems caused by the first (non-regenerative) block, it is preferable to construct estimates using the data collected from the first regeneration time (i.e. the first visit to A) only, so as to get rid of a first-order bias term that cannot be estimated or recovered by any resampling method with a single realization of the chain. Our proposal (see Section 2) consists, then, in imitating the renewal structure of the chain by sampling regeneration data blocks, until the length of the reconstructed bootstrap series is larger than the length n of the original data series. In this way, we approximate the distribution of the (random) number of regeneration blocks in a series of length n and remove significant bias terms.

This resampling method, which we call the *regenerative block bootstrap* (RBB), has a uniform rate of convergence of order $O_{\mathbb{P}}(n^{-1})$, the optimal rate in the i.i.d. case. Unlike the MBB, there is no need in the RBB procedure to choose the size of the blocks, which are entirely determined by the data. Besides, the second-order accuracy of the RBB holds under weak conditions (stipulating a polynomial rate for the decay of the strong mixing

coefficients only). In Section 3 we show how these results may be extended to the much broader class of Harris Markov chains. Our proposal is based on a practical use of the *splitting technique* introduced in Nummelin (1978) and an empirical method to build approximatively a realization drawn from an extension of the chain with a regeneration set. We establish the asymptotic validity of this procedure, even in a non-stationary framework, which is clearly more suitable for many applications. Its second-order validity is only shown in the unstudentized stationary case, up to a rate close to that in the i.i.d. setting. The technical study of the second-order properties of this method and of the optimal rate that may be attained in the studentized case will be carried out at length in a forthcoming paper. We give an entirely data-based procedure for choosing an 'optimal' regeneration set that maximizes an estimation of the expected number of data blocks conditionally on the data. Here we mainly focus on the case of the sample mean in the positive recurrent case, but the ideas set out in this paper may be straightforwardly extended to much more general functionals and even to the null recurrent case, when specific models are considered. Technical proofs are postponed to the final Section 4.

2. Bootstrapping Markov chains with an atom

2.1. Notation and definitions

Here we introduce some notation and recall some key concepts of Markov chain theory (see Meyn and Tweedie 1996 for further details) that are needed throughout the paper. Let $X = (X_n)_{n\in\mathbb{N}}$ be a Markov chain on a countably generated state space $(E, \mathcal{E})$, with transition probability Π and initial probability distribution ν (the assumption that $\mathcal{E}$ is countably generated plays a standard role in the analysis of communicating sets for the chain and is really not restrictive in practice; see the discussion in Orey 1971). Thus for any $B \in \mathcal{E}$ and $n \in \mathbb{N}$, we have

$$X_0 \sim \nu \text{ and } \mathbb{P}(X_{n+1} \in B | X_0, \ldots, X_n) = \Pi(X_n, B) \qquad \text{almost surely.}$$

Recall the following notions. The first formalizes the idea of a communicating structure between specific subsets, while the second considers the set of time points at which such communication may occur.

- The chain is *irreducible* if there exists a σ-finite measure ψ such that for all sets $B \in \mathcal{E}$, when $\psi(B) > 0$, the chain visits B with strictly positive probability, no matter what the starting point.
- Assuming ψ-irreducibility, there exist $d' \in \mathbb{N}^*$ and disjoints sets $D_1, \ldots, D_{d'} (D_{d'+1} = D_1)$ weighted by ψ such that $\psi(E \setminus \cup_{1\leqslant i\leqslant d'} D_i) = 0$ and, for all $x \in D_i$, $\Pi(x, D_{i+1}) = 1$. The *period* of the chain is the greatest common divisor d of such integers; it is *aperiodic* if $d = 1$.

In what follows, $\mathbb{P}_\nu$ (or $\mathbb{P}_x$ for x in E) denotes the probability measure on the underlying space such that $X_0 \sim \nu$ (or $X_0 = x$), $\mathbb{E}_\nu(\cdot)$ is the $\mathbb{P}_\nu$-expectation (or $\mathbb{E}_x(\cdot)$ the $\mathbb{P}_x$-expectation), and $\mathbb{I}\{\mathcal{A}\}$ is the indicator function of the event $\mathcal{A}$.

Assume that X is aperiodic, ψ-irreducible and possesses an accessible atom, that is to say, a set $A \in \mathcal{E}$ such that for all x, y in A, $\Pi(x, \cdot) = \Pi(y, \cdot)$ and $\psi(A) > 0$. Denote by $\tau_A = \tau_A(1) = \inf\{n \geqslant 1,\ X_n \in A\}$ the hitting time on A, by $\tau_A(j) = \inf\{n > \tau_A(j-1),\ X_n \in A\}$ for $j \geqslant 2$, the successive return times to A, and by $\mathbb{E}_A(\cdot)$ the expectation conditionally on $X_0 \in A$. Suppose further that X is Harris recurrent, hence the probability of returning infinitely often to the atom A is equal to one, no matter what the starting point: for all $x \in E$, $\mathbb{P}_x(\tau_A < \infty) = 1$. Then it follows from the strong Markov property that, for any initial law ν, the sample paths may be divided into i.i.d. blocks of random length corresponding to consecutive visits to A,

$$\mathcal{B}_1 = (X_{\tau_A(1)+1}, \ldots, X_{\tau_A(2)}), \ldots, \mathcal{B}_j = (X_{\tau_A(j)+1}, \ldots, X_{\tau_A(j+1)}), \ldots,$$

taking values in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$. The sequence $(\tau_A(j))_{j \geqslant 1}$ defines successive times at which the chain forgets its past, called *regeneration times*.

When an accessible atom exists, the *stochastic stability* properties of X amount to properties concerning the speed of return time to the atom only. For instance, X is positive recurrent if and only if $\mathbb{E}_A(\tau_A) < \infty$ (see Theorem 10.2.2 in Meyn and Tweedie 1996, known as Kac's theorem). Then the unique invariant probability law μ is the *occupation measure* given, for all $B \in \mathcal{E}$, by

$$\mu(B) = \frac{1}{\mathbb{E}_A(\tau_A)}\, \mathbb{E}_A\left(\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\}\right).$$

For such chains, limit theorems can be derived from the application of the corresponding results to the i.i.d. blocks $(\mathcal{B}_n)_{n \geqslant 1}$ (see Smith 1955 for an introduction to the *regenerative method*). Refer to Meyn and Tweedie (1996) for the law of large numbers, the central limit theorem (CLT) and the law of the iterated logarithm, Bolthausen (1982) for the Berry–Esseen theorem, and Malinovskii (1987, 1989) for other refinements of the CLT. The same technique can also be applied to establish moment and probability inequalities, which are not asymptotic results (see Clémençon 2001).

2.2. Preliminary remarks

Let $X^{(n)} = (X_1, \ldots, X_n)$ be observations drawn from X with an a priori known accessible atom A, which we suppose positive recurrent. This covers the case of countable chains, for which any recurrent state is an atom, as well as many Markovian models with regeneration times, widely used in operational research for modelling queuing/storage systems with the empty state $A = \{0\}$ as an atom (see, for instance, Asmussen 1987; Meyn and Tweedie 1996). In the following we denote by $l_n = \sum_{i=1}^{n} \mathbb{I}\{X_i \in A\}$ the number of successive visits to the atom, giving rise to $l_n + 1$ data blocks $\mathcal{B}_0 = (X_1, \ldots, X_{\tau_A(1)})$, $\mathcal{B}_1, \ldots, \mathcal{B}_{l_n - 1}$, $\mathcal{B}_{l_n}^{(n)} = (X_{\tau_A(l_n)+1}, \ldots, X_n)$, with the convention that $\mathcal{B}_{l_n}^{(n)} = \varnothing$ when $\tau_A(l_n) = n$. We denote by $l(\mathcal{B}_j) = \tau_A(j+1) - \tau_A(j)$, $j \geqslant 1$, the lengths of the regeneration blocks (note that $\mathbb{E}(l(\mathcal{B}_j)) = \mathbb{E}_A(\tau_A) = \mu(A)^{-1}$).

Let $f : E \to \mathbb{R}$ be a μ-integrable function. Consider first the estimator $\bar{\mu}_n(f) = n^{-1} \sum_{i=1}^{n} f(X_i)$ of the unknown asymptotic mean $\mu(f) = \mathbb{E}_\mu(f(X_1))$ computed

from the whole data segment $X^{(n)}$. In Bertail and Clémençon (2004) (see Proposition 3.1) it is shown that in the non-stationary case (i.e. when the initial law ν differs from μ), the first data block $\mathcal{B}_0$ induces a significant bias, of order $O(n^{-1})$, which cannot be estimated from a single realization $X^{(n)}$ of the chain starting from ν. This fact is well known in the Bayesian literature, when the matter is to control the convergence of Markov chain Monte Carlo algorithms. It is closely related to the important problem of *burn-in* (time) that is, the time that one should wait before the marginal of a (simulated) chain is close enough to the limit distribution (see, for instance, Hobert and Robert 2004). It is naturally impossible to approximate the second-order properties of such a statistic in the non-stationary case by using a resampling method. Practitioners are thus recommended not to use this estimator. The last (non-regenerative) block $\mathcal{B}_{l_n}^{(n)}$ induces a first-order term in the bias too. And though it seems possible to estimate its sampling distribution accurately, we shall consider statistics based on the observations $(\mathcal{B}_1, \ldots, \mathcal{B}_{l_n-1})$ collected between the first and last visits to the atom only (using $\mathcal{B}_{l_n}^{(n)}$ would make the resampling method below more complex, but the results would be similar). When estimating the limit mean $\mu(f)$, this leads to consideration of the sample mean based on the segment $\tilde{X}^{(n)} = (X_{\tau_A+1}, \ldots, X_{\tau_A(l_n)})$, $\mu_n(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{i=1+\tau_A}^{\tau_A(l_n)} f(X_i)$ with $\mu_n(f) = 0$ if $l_n < 2$ by convention. An adequate standardization $S_n = S_n(\tilde{X}^{(n)})$ is displayed in Bertail and Clémençon (2004) (see Section 2.4.1 below).

2.3. The regenerative block bootstrap algorithm

Although in this paper our asymptotic results are stated in the case of the sample mean only, we present here a valid algorithm, applicable to general statistics T_n for which there exists an adequate standardization S_n: this covers the case of non-degenerate U-statistics, as well as the case of differentiable functionals. For the reasons mentioned above, both the statistic T_n and the estimate of its asymptotic variance we consider are constructed from the regenerative data blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{l_n-1}$ only. We are thus interested in estimating accurately its sampling distribution under $\mathbb{P}_\nu$, say $H_{\mathbb{P}_\nu}^{(n)}(x) = \mathbb{P}_\nu(S_n^{-1}(T_n - \theta) \leqslant x)$. The RBB procedure is carried out in four steps as follows.

1. Count the number of visits l_n to the atom A up to time n. Divide the observed sample path $X^{(n)} = (X_1, \ldots, X_n)$ into $l_n + 1$ blocks, $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_{l_n-1}, \mathcal{B}_{l_n}^{(n)}$ taking values in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$, corresponding to the pieces of the sample path between consecutive visits to the atom A. Drop the first and last (non-regenerative) blocks.
2. Draw sequentially bootstrap data blocks $\mathcal{B}_{1,n}^*, \ldots, \mathcal{B}_{k,n}^*$ independently of the empirical distribution $F_n = (l_n - 1)^{-1} \sum_{j=1}^{l_n-1} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leqslant j \leqslant l_n-1}$ conditioned on $X^{(n)}$, until the length $l^*(k) = \sum_{j=1}^{k} 1(\mathcal{B}_{j,n}^*)$ of the bootstrap data series is larger than n. Let $l_n^* = \inf\{k \geqslant 1,\ l^*(k) > n\}$.
3. From the data blocks generated at step 2, reconstruct a pseudo-trajectory of size $l^*(l_n^* - 1)$by binding the blocks together, $X^{*(n)} = (\mathcal{B}_{1,n}^*, \ldots, \mathcal{B}_{l_n^*-1,n}^*)$. Compute the *RBB statistic* $T_n^* = T_n(X^{*(n)})$.
4. If $S_n = S(\mathcal{B}_1, \ldots, \mathcal{B}_{l_n-1})$ is an appropriate standardization of the original statistic T_n, compute $S_n^* = S(\mathcal{B}_{1,n}^*, \ldots, \mathcal{B}_{l_n^*-1,n}^*)$.

The *RBB distribution* is then given by

$$H_{\mathrm{RBB}}(x) = \mathbb{P}^*(S_n^{*-1}(T_n^* - T_n) \leqslant x | X^{(n)}),$$

where $\mathbb{P}^*(\cdot|X^{(n)})$ denotes the conditional probability given $X^{(n)}$. One may naturally compute a Monte Carlo approximation to $H_{\mathrm{RBB}}(x)$ by independently repeating the procedure above B times.

Remark 2.1. We point out that the RBB differs from the regeneration-based bootstrap proposed by Datta and McCormick (1993), which is not second-order correct up to $O_{\mathbb{P}}(n^{-1/2})$ and in which the number of resampled blocks is held fixed at $l_n - 1$, conditionally on the sample; it also differs from the modified version due to Bertail and Clémençon (2005). Note also that the principles underlying the RBB may be applied to any (possibly continuous-time and not necessarily Markovian) regenerative process (see Thorisson 2000).

2.4. Second-order accuracy of the RBB

Here we study the asymptotic validity of the RBB for the mean standardized by an adequate estimator of the asymptotic variance. This is the useful version for confidence intervals but also for practical use of the bootstrap (see Hall 1992). The accuracy achieved by the RBB is similar to the optimal rate of the i.i.d. bootstrap, contrary to the MBB (see Götze and Künsch 1996).

2.4.1. Further notation and preliminary remarks

We set $n_A = \tau_A(l_n) - \tau_A(1) = \sum_{j=1}^{l_n-1} l(\mathcal{B}_j)$ and $f(\mathcal{B}_j) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$, for any $j \geqslant 1$. We may thus write

$$\mu_n(f) - \mu(f) = n_A^{-1} \sum_{j=1}^{l_n-1} \{f(\mathcal{B}_j) - l(\mathcal{B}_j)\mu(f)\}.$$

By virtue of the strong Markov property, $\{f(\mathcal{B}_j) - l(\mathcal{B}_j)\mu(f)\}_{j\geqslant 1}$ are i.i.d. random variables with mean 0 and variance $\sigma_F^2 = \mathbb{E}(\{f(\mathcal{B}_j) - l(\mathcal{B}_j)\,\mu(f)\}^2)$. In the following, we also set $\alpha = \mathbb{E}_A(\tau_A)$ and $\beta = \mathrm{cov}(l(\mathcal{B}_j), f(\mathcal{B}_j) - l(\mathcal{B}_j)\mu(f)\})$. Assuming the expectations $\mathbb{E}_A(\sum_{i=1}^{\tau_A}|f(X_i)|^2)$, $\mathbb{E}_A(\tau_A^2)$, $\mathbb{E}_\nu(\sum_{i=1}^{\tau_A}|f(X_i)|)$ and $\mathbb{E}_\nu(\tau_A)$ are finite, the CLT holds (see Theorem 17.2.2 in Meyn and Tweedie 1996). We have, as $n \to \infty$, $n^{-1/2}(\mu_n(f) - \mu(f)) \to \mathcal{N}(0, \sigma_f^2)$ in distribution under $\mathbb{P}_\nu$, with the asymptotic variance $\sigma_f^2 = \alpha^{-1}\sigma_F^2$.

The following estimate of the asymptotic variance σ_f^2 may be naturally constructed using the regeneration times:

$$\sigma_n^2(f) = n_A^{-1} \sum_{j=1}^{l_n-1} \{f(\mathcal{B}_j) - \mu_n(f)l(\mathcal{B}_j)\}^2.$$

First-order properties of this estimator are studied in Bertail and Clémençon (2004). A straightforward application of the strong law of large numbers for positive recurrent Markov

chains shows it is strongly consistent. Under further regularity conditions, Bertail and Clémençon (2004) have also shown that its bias is of order $O(n^{-1})$ and it is asymptotically normal.

As shown below, this standardization does not worsen the performance of the RBB, while the standardization of the MBB distribution in the strong mixing case is the main barrier to achieving good rates (see Götze and Künsch 1996). Moreover, contrary to the MBB, the bootstrap counterparts in the studentized case are straightforwardly defined in our regenerative setting. With $n_A^* = \sum_{j=1}^{l_n^*-1} l(\mathcal{B}_j^*)$, define

$$\mu_n^*(f) = n_A^{*-1} \sum_{j=1}^{l_n^*-1} f(\mathcal{B}_j^*) \quad \text{and} \quad \sigma_n^{*2}(f) = n_A^{*-1} \sum_{j=1}^{l_n^*-1} \{f(\mathcal{B}_j^*) - \mu_n^*(f) l(\mathcal{B}_j^*)\}^2.$$

2.4.2. Main asymptotic result

We now state the asymptotic validity of the RBB in the atomic case.

Theorem 2.1. *Assume that the chain X fulfils the following conditions:*

(i) *(Cramér condition)* $\overline{\lim}_{|t|\to\infty} |\mathbb{E}_A(\exp(\mathrm{i}t(\sum_{i=1}^{\tau_A}\{f(X_i) - \mu(f)\})))| < 1$.

(ii) *(Non-degeneracy of the asymptotic variance)* $\sigma_f^2 > 0$.

(iii) *(Block moment conditions) For some* $s \geqslant 2$,

$$\mathbb{E}_A(\tau_A^s) < \infty, \ \mathbb{E}_A\left(\sum_{i=1}^{\tau_A} |f(X_i)|\right)^s < \infty.$$

(iv) *(Block moment conditions for the initial law* ν*)*

$$\mathbb{E}_\nu(\tau_A^2) < \infty, \ \mathbb{E}_\nu\left(\sum_{i=1}^{\tau_A} |f(X_i)|\right)^2 < \infty.$$

(v) *(Boundedness of the N-fold convolution of the joint density) There exists* $N \in \mathbb{N}^*$ *such that the N-fold convoluted density* G^{*N} *is bounded, denoting by G the density of the* $(f(\mathcal{B}_j) - l(\mathcal{B}_j)\mu(f))^2$.

Then, under assumptions (i)–(iv) *with* $s > 6$, *the RBB distribution estimate for the unstandardized sample mean is second-order accurate:*

$$\Delta_n^{\mathrm{U}} = \sup_{x\in\mathbb{R}} |H_{\mathrm{RBB}}^{\mathrm{U}}(x) - H_\nu^{\mathrm{U}}(x)| = O_{\mathbb{P}_\nu}(n^{-1}), \qquad \text{as } n \to \infty,$$

where

$$H_{\mathrm{RBB}}^{\mathrm{U}}(x) = \mathbb{P}^*(n_A^{1/2}\sigma_f^{-1}\{\mu_n^*(f) - \mu_n(f)\} \leqslant x | X^{(n)}),$$

$$H_\nu^{\mathrm{U}}(x) = \mathbb{P}_\nu(n_A^{1/2}\sigma_f^{-1}\{\mu_n(f) - \mu(f)\} \leqslant x).$$

Under assumptions (i)–(v) *with* $s > 8$*, the RBB distribution estimate for the studentized sample mean is also second-order correct:*

$$\Delta_n^{\mathrm{S}} = \sup_{x\in\mathbb{R}} |H_{\mathrm{RBB}}^{\mathrm{S}}(x) - H_{\nu}^{\mathrm{S}}(x)| = O_{\mathbb{P}_\nu}(n^{-1}), \qquad \text{as } n \to \infty,$$

where

$$H_{\mathrm{RBB}}^{\mathrm{S}}(x) = \mathbb{P}^*(n_A^{*1/2}\sigma_n^{*-1}(f)\{\mu_n^*(f) - \mu_n(f)\} \leqslant x | X^{(n)}),$$

$$H_{\nu}^{\mathrm{S}}(x) = \mathbb{P}_\nu(n_A^{1/2}\sigma_n^{-1}(f)\{\mu_n(f) - \mu(f)\} \leqslant x).$$

This result ensures that the RBB has the optimality of the i.i.d. bootstrap. This is noteworthy, since the RBB method applies to countable chains (for which any recurrent state is an atom) but also to many specific Markov chains widely used in practice for modelling queuing/storage systems (for a detailed account of such models, see Meyn and Tweedie 1996: Section 2.4; Asmussen 1987).

We point out that the relationship between the 'block moment' condition (iii) and the rate of decay of mixing coefficients was investigated in Bolthausen (1982): for instance, condition (iii) is typically fulfilled when f is bounded, if the strong mixing coefficients sequence decreases at an arithmetic rate $n^{-\rho}$, for some $\rho > s - 1$. Condition (iv) is needed to control the EE up to $O(n^{-1})$(see the partitioning arguments used in Malinovskii 1987; Bertail and Clémençon 2004: equation (19)). Intuitively, even if we use the truncated sum, too large a variance of the size of the block (or of the functional on the first block) may yield a large perturbation on the tail of the number of blocks which plays an important role in the second-order terms. Condition (v) is a technical assumption used in Bertail and Clémençon (2004) to obtain an EE in the studentized case. As may be shown by a straightforward calculation, if the density of the $\sum_{i=1+\tau_A(j)}^{\tau_A(j+1)}(f(X_i) - \mu(f))$ is bounded then the condition holds for $N = 2$.

Remark 2.2. The same results holds a.s. up to $O_{\mathbb{P}_\nu}(n^{-1}\log\log(n)^{1/2})$, just like in the i.i.d. case under the same moment conditions. This straightforwardly results from the law of the iterated logarithm applied to the empirical moments of the blocks appearing in the EE of the RBB distribution.

3. Approximate regenerative block bootstrap

3.1. General Harris chains and the splitting technique

For the sake of clarity, we briefly recall the *splitting technique* introduced in Nummelin (1978), which allows us to extend in some sense the probabilistic structure of any chain in order to artificially construct a regeneration set in the general Harris recurrent case. Here X is a general aperiodic, ψ-irreducible chain with transition kernel Π, taking values in a countably generated state space $(E, \mathcal{E})$. First, consider the following notion:

Definition 3.1. *A set $S \in \mathcal{E}$ is* small *if there exist $m \in \mathbb{N}^*$, a probability measure Φ supported by S, and $\delta > 0$ such that for all $x \in S$, for all $A \in \mathcal{E}$, $\Pi^m(x, A) \geqslant \delta\Phi(A)$, denoting by Π^m the mth iterate of Π. When this holds, we say that the chain satisfies the* minorization condition $\mathcal{M}(m, S, \delta, \Phi)$.

Recall that accessible small sets do exist for irreducible chains. Any accessible set actually contains a small set (see Jain and Jamison 1967). And in practice, finding such a set consists in most cases in exhibiting an accessible set, for which the probability that the chain returns to it in m steps is uniformly bounded from below (see Section 3.6). Suppose that the chain X satisfies $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ for some set S such that $\psi(S) > 0$. Even if it entails replacing the chain $(X_n)_{n\in\mathbb{N}}$ by the chain $((X_{nm}, \ldots, X_{n(m+1)-1}))_{n\in\mathbb{N}}$, we suppose that $m = 1$. The sample space is expanded so as to define a sequence $(Y_n)_{n\in\mathbb{N}}$ of independent random variables with parameter δ by defining the joint distribution $\mathbb{P}_{\nu,\mathcal{M}}$ whose construction relies on the following randomization of the transition probability Π each time the chain X hits the set S (note that it happens a.s. since the chain is Harris recurrent and $\psi(S) > 0$). If $X_n \in S$ and

- if $Y_n = 1$ (which happens with probability $\delta \in]0, 1[$), then X_{n+1} is distributed according to the probability measure Φ;
- if $Y_n = 0$ (which happens with probability $1 - \delta$), then X_{n+1} is distributed according to the probability measure $(1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta\Phi(\cdot))$.

This bivariate Markov chain $X^{\mathcal{M}} = ((X_n, Y_n))_{n\in\mathbb{N}}$ is called the *split chain*. It takes values in $E \times \{0, 1\}$ and possesses an atom, namely $S \times \{1\}$. The whole point of this construction is that $X^{\mathcal{M}}$ inherits all the communication and stochastic stability properties from X (irreducibility, Harris recurrence). In particular the blocks constructed for the split chain are i.i.d. (for the case $m = 1$, else they are 1-dependent). Hence the splitting method enables us to extend the regenerative method and to establish all the results known for atomic chains to general Harris chains. For simplicity's sake, we omit the subscript $\mathcal{M}$ in what follows and, in an abuse of notation, denote by $\mathbb{P}_\nu$ the extensions of the underlying probability we consider.

3.2. Approximating the regenerative blocks

In the following, we suppose for the sake of simplicity that condition $\mathcal{M}$ is fulfilled with $m = 1$. We assume, further, that the family of conditional distributions $\{\Pi(x, \mathrm{d}y)\}_{x\in E}$ and the initial distribution ν are dominated by a σ-finite reference measure λ, so that $\nu(\mathrm{d}y) = f(y)\lambda(\mathrm{d}y)$ and $\Pi(x, \mathrm{d}y) = p(x, y)\lambda(\mathrm{d}y)$, for all $x \in E$. Note that the minorization condition entails that Φ is absolutely continuous with respect to λ too, and that $p(x, y) \geqslant \delta\phi(y)$, $\lambda(\mathrm{d}y)$-a.s. for any $x \in S$, with $\Phi(\mathrm{d}y) = \phi(y)\mathrm{d}y$. Let Y be the binary random sequence constructed via the Nummelin technique from the parameters of condition $\mathcal{M}$. Our proposal for approximating the Nummelin construction is based on the crucial observation that the distribution of $Y^{(n)} = (Y_1, \ldots, Y_n)$ conditionally on $X^{(n+1)} =$

$(x_1, \ldots, x_{n+1})$ is the tensor product of Bernoulli distributions given, for all $\beta^{(n)} = (\beta_1, \ldots, \beta_n) \in \{0, 1\}^n$, $x^{(n+1)} = (x_1, \ldots, x_{n+1}) \in E^{n+1}$, by

$$\mathbb{P}_\nu\left(Y^{(n)} = \beta^{(n)} | X^{(n+1)} = x^{(n+1)}\right) = \prod_{i=1}^{n} \mathbb{P}_\nu(Y_i = \beta_i | X_i = x_i, X_{i+1} = x_{i+1}),$$

with, for $1 \leqslant i \leqslant n$,

- if $x_i \notin S$, $\mathbb{P}_\nu(Y_i = 1 | X_i = x_i, X_{i+1} = x_{i+1}) = \delta$,
- if $x_i \in S$, $\mathbb{P}_\nu(Y_i = 1 | X_i = x_i, X_{i+1} = x_{i+1}) = \dfrac{\delta\phi(x_{i+1})}{p(x_i, x_{i+1})}$.

Roughly speaking, conditioned on $X^{(n+1)}$, from $i = 1$ to n, Y_i is drawn from the Bernoulli distribution with parameter δ, unless X has hit the small set S at time i: in this case Y_i is drawn from the Bernoulli distribution with parameter $\delta\phi(X_{i+1})/p(X_i, X_{i+1})$. We denote this probability distribution by $\mathcal{L}^{(n)}(p, S, \delta, \phi, x^{(n+1)})$. If we were able to generate $Y_1, \ldots, Y_n$, so that $X^{\mathcal{M}(n)} = ((X_1, Y_1), \ldots, (X_n, Y_n))$ is a realization of the split chain $X^{\mathcal{M}}$ described in Section 3.1, then we could apply the RBB procedure to the sample path $X^{\mathcal{M}(n)}$ leading to asymptotically i.i.d. blocks. Unfortunately, as shown above, knowledge of the transition density $p(x, y)$ is required to make drawing $Y_1, \ldots, Y_n$ practical this way.

Our proposal for approximating this construction consists in computing an estimator $p_n(x, y)$ of the transition density $p(x, y)$ using data $X_1, \ldots, X_{n+1}$ (note that we may choose the estimate $p_n(x, y)$ of the transition density such that $p_n(x, y) \geqslant \delta\phi(y)$, $\lambda(\mathrm{d}y)$-a.s., and $p_n(X_i, X_{i+1}) > 0$, $1 \leqslant i \leqslant n$), and then drawing a binary random vector $(\hat{Y}_1, \ldots, \hat{Y}_n)$ conditionally on $X^{(n+1)} = (X_1, \ldots, X_{n+1})$, from the distribution $\mathcal{L}^{(n)}(p_n, S, \delta, \phi, X^{(n+1)})$, approximating in some sense the conditional distribution $\mathcal{L}^{(n)}(p, S, \delta, \phi, X^{(n+1)})$ of $(Y_1, \ldots, Y_n)$ for given $X^{(n+1)}$. Our method for bootstrapping Harris chains, which we call the *approximate regenerative block bootstrap* (ARBB), simply amounts, then, to applying the RBB procedure to the data $((X_1, \hat{Y}_1), \ldots, (X_n, \hat{Y}_n))$ as if they were drawn from the atomic chain $X^{\mathcal{M}}$.

3.3. A coupling result for $(X_i, \hat{Y}_i)_{1 \leqslant i \leqslant n}$ and $(X_i, Y_i)_{1 \leqslant i \leqslant n}$

We now prove that the distribution of $(X_i, \hat{Y}_i)_{1 \leqslant i \leqslant n}$ gets closer and closer to the distribution of $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ in the sense of the Mallows distance (also known as the Kantorovich or Wasserstein metric in the probability literature; see Bickel and Freedmann 1981) as $n \to \infty$. Hence, we express here the distance between the distributions $\mathbb{P}^Z$ and $\mathbb{P}^{Z'}$ of two random sequences $Z = (Z_n)_{n \in \mathbb{N}}$ and $Z' = (Z'_n)_{n \in \mathbb{N}}$, taking values in $\mathbb{R}^k$ (see Rachev and Rüschendorf 1998: 76), by

$$l_r(Z, Z') = l_r(\mathbb{P}^Z, \mathbb{P}^{Z'}) = \min\{L_r(W, W'); W \sim \mathbb{P}^Z, W' \sim \mathbb{P}^{Z'}\},$$

with $r \geqslant 1$ and $L_r(W, W') = (\mathbb{E}[D^r(W, W')])^{1/r}$, where D denotes the metric on the space $\chi(\mathbb{R}^k) = (\mathbb{R}^k)^\infty$ defined by $D(w, w') = \sum_{k=0}^{\infty} 2^{-k} \|w_k - w'_k\|_{\mathbb{R}^k}$, for any w, w' in $\chi(\mathbb{R}^k)$ ($\|\cdot\|_{\mathbb{R}^k}$ denoting the usual Euclidian norm of $\mathbb{R}^k$). Thus, viewing the sequences

$Z^{(n)} = (X_k, Y_k)_{1 \leqslant k \leqslant n}$ and $\hat{Z}^{(n)} = (X_k, \hat{Y}_k)_{1 \leqslant i \leqslant n}$ as the beginning segments of infinite series, we evaluate the deviation between the distribution $P^{(n)}$ of $Z^{(n)}$ and the distribution $\hat{P}^{(n)}$ of $\hat{Z}^{(n)}$ using $l_1(P^{(n)}, \hat{P}^{(n)})$.

Theorem 3.1. *Assume that:* (i) *S is chosen so that* $\inf_{x \in S} \phi(x) > 0$*; and* (ii) *$p$ is estimated by p_n at the rate α_n for the mean squared error (MSE) when error is measured by the L^∞ loss over S^2. Then* $l_1(P^{(n)}, \hat{P}^{(n)}) \leqslant (\delta \inf_{x \in S} \phi(x))^{-1} \alpha_n^{1/2}$.

This theorem is established by exhibiting a specific coupling of $(X_i, \hat{Y}_i)_{1 \leqslant i \leqslant n}$ and $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ (see Section 4.2)). It is a crucial tool for deriving the results stated in the next section. It also clearly shows that the closeness between the two distributions is tightly connected to the rate of convergence of the estimator $p_n(x, y)$ but also to the minorization condition parameters (see Section 3.6).

3.4. The ARBB algorithm

It is now easy to see how we can execute an ARBB algorithm:

1. From the data $X^{(n+1)} = (X_1, \ldots, X_{n+1})$, compute an estimate $p_n(x, y)$ of $p(x, y)$ such that $p_n(x, y) \geqslant \delta \phi(y)$, $\lambda(\mathrm{d}y)$-a.s., and $p_n(X_i, X_{i+1}) > 0$, $1 \leqslant i \leqslant n$.
2. Conditionally on $X^{(n+1)}$, draw a binary data vector $(\hat{Y}_1, \ldots, \hat{Y}_n)$ from the distribution estimate $\mathcal{L}^{(n)}(p_n, S, \delta, \phi, X^{(n+1)})$. From a practical point of view, it naturally suffices to draw the binary $\hat{Y}_i$s at times i when the chain visits the set S (i.e. when $X_i \in S$), which are the only time points at which the split chain may regenerate: at such a time point i, draw $\hat{Y}_i$ according to the Bernoulli distribution with parameter $\delta \phi(X_{i+1}) / p_n(X_i, X_{i+1})$).
3. Count the number of visits $\hat{l}_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S, \hat{Y}_i = 1)$ to the set $A_{\mathcal{M}} = S \times \{1\}$ up to time n. Define the successive hitting times of $A_{\mathcal{M}}$, $\hat{\tau}_{A_{\mathcal{M}}}(i)$, $i = 1, \ldots, \hat{l}_n$, and divide the sample path $X^{(n)}$ into $\hat{l}_n + 1$ blocks, $\hat{\mathcal{B}}_i = (X_{\hat{\tau}_{A_{\mathcal{M}}}(i)+1}, \ldots, X_{\hat{\tau}_{A_{\mathcal{M}}}(i+1)})$, corresponding to the pieces of the trajectory between consecutive visits to $A_{\mathcal{M}}$.
4. Apply steps 2–4 of the RBB algorithm to the collection of pseudo-blocks $\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_{\hat{l}_n - 1}$, producing the *ARBB distribution*

$$H_{\mathrm{ARBB}}(x) = \mathbb{P}^*(S_n^{*-1}(T_n^* - T_n) \leqslant x | X^{(n+1)}),$$

where $T_n^* = T(X^{*(n)})$ and $S_n^* = S(X^{*(n)})$ are respectively the *ARBB statistic* and the *ARBB standardization* obtained from the reconstructed path $X^{*(n)} = (\mathcal{B}_1^*, \ldots, \mathcal{B}_{l_n^* - 1}^*)$.

3.5. Asymptotic validity of the ARBB

The asymptotic variance is $\sigma_f^2 = \mathbb{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} \mathbb{E}_{A_{\mathcal{M}}}((\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} \{f(X_i) - \mu(f)\})^2)$, where $\tau_{A_{\mathcal{M}}} = \inf\{n \geqslant 1, (X_n, Y_n) \in S \times \{1\}\}$ and $\mathbb{E}_{A_{\mathcal{M}}}(\cdot)$ denotes the expectation conditionally on $(X_0, Y_0) \in A_{\mathcal{M}} = S \times \{1\}$. However, in the studentized case, one cannot use the standardization defined in Section 2.4.1 in the atomic setting for the split chain, since the

true regeneration times are unknown. We thus consider the following estimators based on the *pseudo-regeneration times* (i.e. times i when $(X_i, \hat{Y}_i) \in S \times \{1\}$) generated by the procedure detailed in Section 3.4,

$$\hat{\mu}_n(f) = \hat{n}_{A_M}^{-1} \sum_{j=1}^{\hat{l}_n - 1} f(\hat{\mathcal{B}}_j) \quad \text{and} \quad \hat{\sigma}_n^2(f) = \hat{n}_{A_M}^{-1} \sum_{j=1}^{\hat{l}_n - 1} \left\{ f(\hat{\mathcal{B}}_j) - \hat{\mu}_n(f) l(\hat{\mathcal{B}}_j) \right\}^2$$

of $\mu(f)$ and σ_f^2 respectively, with $\hat{n}_{A_M} = \hat{\tau}_{A_M}(\hat{l}_n) - \hat{\tau}_{A_M}(1) = \sum_{j=1}^{\hat{l}_n - 1} l(\hat{\mathcal{B}}_j)$ and $f(\hat{\mathcal{B}}_j) = \sum_{i=1+\hat{\tau}_A(j)}^{\hat{\tau}_A(j+1)} f(X_i)$. By convention, $\hat{\mu}_n(f)$ ($\hat{\sigma}_n^2(f)$, $\hat{n}_{A_M}$) equals 0, when $\hat{l}_n \leqslant 1$. Note that, analogously to the way we proceeded in the atomic case to avoid unrecoverable and large bias terms that cannot be approximated by using any resampling method (see Section 2.2), data collected before the first (or after the last) pseudo-regeneration time are not used to construct these estimators, yielding estimates that are not contaminated too much by the starting distribution.

Define the unstandardized distribution by

$$H_\nu^{\mathrm{U}}(x) = \mathbb{P}_\nu(\hat{n}_{A_M}^{1/2} \sigma_f^{-1} (\hat{\mu}_n(f) - \mu(f)) \leqslant x).$$

We also define the *pseudo-regeneration based studentized sample mean* $\hat{t}_{A_M,n} = \hat{n}_{A_M}^{1/2} (\hat{\mu}_n(f) - \mu(f))/\hat{\sigma}_n(f)$, with sampling distribution $H_\nu^{\mathrm{S}}(x) = \mathbb{P}_\nu(\hat{t}_{A_M,n} \leqslant x)$.

3.5.1. Further assumptions and preliminary results

We will require the following assumptions. Let $k \geqslant 2$ be a real number.

$\mathcal{H}_1(f, k, \nu)$. The small set S is such that

$$\sup_{x \in S} \mathbb{E}_x \left(\left(\sum_{i=1}^{\tau_S} |f(X_i)| \right)^k \right) < \infty \quad \text{and} \quad \mathbb{E}_\nu \left(\left(\sum_{i=1}^{\tau_S} |f(X_i)| \right)^k \right) < \infty.$$

$\mathcal{H}_2(k, \nu)$. The set S is such that $\sup_{x \in S} \mathbb{E}_x(\tau_S^k) < \infty$ and $\mathbb{E}_\nu(\tau_S^k) < \infty$.

For a sequence of nonnegative real numbers $(\alpha_n)_{n \in \mathbf{N}}$ converging to 0, we have:

$\mathcal{H}_3$. $p(x, y)$ is estimated by $p_n(x, y)$ at the rate α_n for the MSE when error is measured by the L^∞ loss over $S \times S$:

$$\mathbb{E}_\nu \left(\sup_{(x,y) \in S \times S} |p_n(x, y) - p(x, y)|^2 \right) = O(\alpha_n), \qquad \text{as } n \to \infty.$$

$\mathcal{H}_4$. The density ϕ is such that $\inf_{x \in S} \phi(x) > 0$.

$\mathcal{H}_5$. The transition density $p(x, y)$ and its estimate $p_n(x, y)$ are bounded by a constant $R < \infty$ over S^2.

We point out that assumptions $\mathcal{H}_1(f, k, \nu)$ and $\mathcal{H}_1(k, \nu)$ do not depend on the choice of the small set S (if it is verified for some accessible small set S, it is also fulfilled for all accessible small sets of the chain). Note also that when $\mathcal{H}_1(k, \nu)$ is satisfied, $\mathcal{H}_1(f, k, \nu)$ is verified for any bounded function f.

Remark 3.1. Numerous estimators of the transition density have been proposed in the literature and their estimation rates have been established under various smoothness assumptions (see, for instance, Athreya and Atuncar 1998; Clémençon 2000; and the references therein). Under classical Hölder constraints of order s, the typical rate in this set-up is $\alpha_n \sim (\ln n/n)^{s/(s+1)}$.

The next result justifies the use of our estimates in an asymptotic sense.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 are fulfilled by the chain, as well as* $\mathcal{H}_1(f, \rho, \nu)$, $\mathcal{H}_2(\rho, \nu)$ *with* $\rho \geqslant 4$, $\mathcal{H}_3$, $\mathcal{H}_4$ *and* $\mathcal{H}_5$. *Then we have as* $n \to \infty$, $\hat{\sigma}_n^2(f) \to \sigma_f^2$ *in* $\mathbb{P}_\nu$*-probability and*

$$\hat{\boldsymbol{n}}_{A_\mathcal{M}}^{1/2}(\hat{\boldsymbol{\mu}}_n(f) - \mu(f))/\hat{\sigma}_n(f) \to \mathcal{N}(0, 1) \text{ in distribution under } \mathbb{P}_\nu.$$

We recall that condition $\mathcal{H}_1(f, \rho, \nu)$ may be more easily checked in practice by using test function methods (see Kalashnikov 1978). In particular, it is well known that 'block' moment conditions may be replaced by drift criteria of Lyapunov's type; see Chapter 11 in Meyn and Tweedie (1996) for further details on such conditions and many examples, as well as Douc *et al.* (2004) for recent results. We also point out that assumptions $\mathcal{H}_1(f, \rho, \nu)$ and $\mathcal{H}_2(\rho, \nu)$ classically imply that the block-moment conditions (iii) and (iv) are satisfied by the split chain for $s = \rho$.

3.5.2. Main asymptotic theorem

We now define the bootstrap counterparts of the statistics introduced above. Let $\mathcal{B}_1^*, \ldots, \mathcal{B}_{l_n^*-1}^*$ be the bootstrapped pseudo-regenerative blocks and $n_{A_\mathcal{M}}^* = \sum_{j=1}^{l_n^*-1} l(\mathcal{B}_j^*)$ be the length of the ARBB data series, then set

$$\mu_n^*(f) = n_{A_\mathcal{M}}^{*-1} \sum_{j=1}^{l_n^*-1} f(\mathcal{B}_j^*), \qquad \sigma_n^{*2}(f) = n_{A_\mathcal{M}}^{*-1} \sum_{j=1}^{l_n^*-1} \{f(\mathcal{B}_j^*) - \mu_n^*(f) l(\mathcal{B}_j^*)\}^2.$$

The unstandardized version of the *ARBB* distribution is given by

$$H_{\mathrm{ARBB}}^{\mathrm{U}}(x) = \mathbb{P}^*(n_{A_\mathcal{M}}^{*1/2} \hat{\sigma}_n^{-1}(f)(\mu_n^*(f) - \hat{\boldsymbol{\mu}}_n(f)) \leqslant x | X^{(n+1)}).$$

Define also the bootstrap version of the pseudo-regeneration based studentized sample mean by

$$t_{A_\mathcal{M},n}^* = \frac{n_{A_\mathcal{M}}^{*1/2}(\mu_n^*(f) - \hat{\boldsymbol{\mu}}_n(f))}{\sigma_n^*(f)}$$

and the studentized ARBB distribution estimate $H_{\mathrm{ARBB}}^{\mathrm{S}}(x) = \mathbb{P}^*(t_{A_\mathcal{M},n}^* \leqslant x | X^{(n+1)})$. This is the same construction as in the atomic case, except that it uses the approximate blocks instead of the true regenerative ones. We have the following result.

Theorem 3.3. *Under the hypotheses of Theorem 3.2, we have the following convergences in probability under* $\mathbb{P}_\nu$*:*

$$\Delta_n^{\mathrm{U}} = \sup_{x\in\mathbb{R}} |H_{\mathrm{ARBB}}^{\mathrm{U}}(x) - H_\nu^{\mathrm{U}}(x)| \to 0, \qquad \text{as } n \to \infty$$

$$\Delta_n^{\mathrm{S}} = \sup_{x\in\mathbb{R}} |H_{\mathrm{ARBB}}^{\mathrm{S}}(x) - H_\nu^{\mathrm{S}}(x)| \to 0, \qquad \text{as } n \to \infty.$$

3.5.3. Second-order properties of the ARBB in the stationary case

In consideration of technical difficulties, the study of the second-order properties of the ARBB distribution estimate is confined in this paper to the unstudentized case in a stationary framework. In the stationary case, one may actually use $\mu_n(f) = n^{-1}\sum_{i=1}^n f(X_i)$ as a natural unbiased estimate of $\mu(f)$, thus avoiding controlling the contributions of the first and last pseudo-regenerative blocks. In the following we deal with

$$H_\mu^{\mathrm{U}}(x) = \mathbb{P}_\mu(n^{1/2}\sigma_f^{-1}\{\bar{\mu}_n(f) - \mu(f)\} \leqslant x).$$

Because the last pseudo-regenerative block is dropped in the ARBB procedure, a bias problem appears in the stationary case, which can be easily handled by recentring the ARBB distribution. Hence, we now consider

$$H_{\mathrm{ARBB}}^{\mathrm{U,s}}(x) = \mathbb{P}^*(t^*_{A_{\mathcal{M}},n} - \mathbb{E}^*(t^*_{A_{\mathcal{M}},n}) \leqslant x | X^{(n+1)}).$$

We now state a result providing an explicit rate for the ARBB in this setting.

Theorem 3.4. *Assume that the Markov chain* X *is stationary (i.e.* $\mu = \nu$*). Under the hypotheses of Theorem 3.2 with* $\rho > 6$*, and if in addition the Cramér condition*

$$\overline{\lim_{|t|\to\infty}} \sup_{x\in S} \left| \mathbb{E}_x\left(\exp\left(\mathrm{i}t \left(\sum_{i=1}^{\tau_S} \{f(X_i) - \mu(f)\} \right) \right) \right) \right| < 1$$

is fulfilled, then we have

$$\Delta_n^{\mathrm{U}} = \sup_{x\in\mathbb{R}} |H_{\mathrm{ARBB}}^{\mathrm{U,s}}(x) - H_\mu^{\mathrm{U}}(x)| = O_{\mathbb{P}_\nu}(\alpha_n^{1/2} n^{-1/2}), \qquad \text{as } n \to \infty$$

Remark 3.2. Observe that only consistency of $p_n(x, y)$ over S^2 in the MSE sense is required for the ARBB to be second-order correct. Furthermore, in the geometrically recurrent case, the best rate that can be attained is typically of order $\alpha_n = n^{-1}\log(n)$ (see Clémençon 2000, for instance), yielding the validity of the ARBB up to the almost optimal rate $o_{\mathbb{P}_\nu}(n^{-1}\log(n)^{1/2})$, which clearly improves on the MBB even in the unstudentized case (see Lahiri 2003: Section 6.5).

Remark 3.3. What makes it very hard to carry over the ideas behind the RBB to the general ARBB case is that, by construction, pseudo-regeneration times $\hat{\tau}_{A_{\mathcal{M}}}(j)$ and the (dependent) data blocks $\hat{\mathcal{B}}_j$ they induce all depend on the whole trajectory, owing to the transition

estimation step. A possible construction to avoid this problem is to use a *double splitting trick* (in a semiparametric sense). The idea is to construct first the transition estimator using the first m_n observations (with $m_n \to \infty$, $m_n/n \to 0$ as $n \to \infty$), then to drop the next q_n observations (typically $q_n \ll m_n$, $q_n \to \infty$ as $n \to \infty$), allowing the split chain to regenerate with overwhelming probability, and finally to construct the pseudo-blocks from the $n - m_n - q_n$ remaining observations. It is easy to understand that these blocks are asymptotically i.i.d conditionally on the first m_n observations. One may then prove the second-order validity of the procedure in both the studentized and unstudentized case. However, this splitting trick entails some loss in rate of the ARBB distribution. By standard arguments, this rate is typically $O_{\mathbb{P}_\nu}(\alpha_{m_n}^{1/2}(n - q_n - m_n)^{-1/2})$ but can be optimized when α_n is known. In the regular case $\alpha_n = n^{-1}\log(n)$, it is easy to see that one can choose $m_n = n^{2/3}$ and the second-order validity of the studentized ARBB distribution holds up to $O(n^{-5/6}\log(n))$. This is still a significant improvement on the MBB rate.

3.6. Practical choice of the small set

On many examples of real-valued positive recurrent chains X, one may easily check that any compact interval $V_{x_0}(\varepsilon) = [x_0 - \varepsilon, x_0 + \varepsilon]$ for a well-chosen x_0 (typically close to its asymptotic mean) and $\varepsilon > 0$ small enough, is a small set with the uniform distribution $\mathcal{U}_{V_{x_0}(\varepsilon)}$ on $V_{x_0}(\varepsilon)$ as a minorization measure. To illustrate this point, consider the general heteroscedastic autoregressive model

$$X_{n+1} = m(X_n) + \sigma(X_n)\varepsilon_{n+1}, \qquad n \in \mathbb{N},$$

where $m : \mathbb{R} \to R$ and $\sigma : \mathbb{R} \to \mathbb{R}_+^*$ are measurable functions, $(\varepsilon_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of random variables drawn from $g(x)\mathrm{d}x$ such that, for all $n \in \mathbb{N}$, ε_{n+1} is independent of the X_k, $k \leqslant n$, with $\mathbb{E}(\varepsilon_{n+1}) = 0$ and $\mathrm{var}(\varepsilon_{n+1}) = 1$. The transition density is given by $p(x, y) = \sigma(x)^{-1} g(\sigma(x)^{-1}(y - m(x)))$, $(x, y) \in \mathbb{R}^2$. Assume further that g, m and σ are continuous functions and there exists $x_0 \in \mathbb{R}$ such that $p(x_0, x_0) > 0$. Then the transition density is uniformly bounded from below over some neighbourhood $V_{x_0}(\varepsilon)^2 = [x_0 - \varepsilon, x_0 + \varepsilon]^2$ of (x_0, x_0) in $\mathbb{R}^2$: there exists $\delta = \delta(\varepsilon) \in]0, 1[$ such that

$$\inf_{(x,y)\in V_{x_0}^2} p(x, y) \geqslant \delta(2\varepsilon)^{-1}. \tag{1}$$

Thus the chain X satisfies the minorization condition $\mathcal{M}(1, V_{x_0}(\varepsilon), \delta, \mathcal{U}_{V_{x_0}(\varepsilon)})$. Hence, in the case when one knows x_0, ε and δ such that the bound (1) holds (this simply amounts to knowing a uniform lower bound estimate for the probability of returning to $V_{x_0}(\varepsilon)$ in one step), one may effectively apply the ARBB methodology to X. The major point is essentially to choose $V_{x_0}(\varepsilon)$ and estimate the corresponding $\delta(\varepsilon)$.

The number of pseudo-regenerative blocks to resample actually depends on how large the small set chosen is (or more exactly, on how often it is visited by the chain in a trajectory of finite length) and how accurate the lower bound (1) is (the larger δ is, the larger is the probability of drawing pseudo-regenerative times $\hat{Y}_k = 1$ at randomization steps, that is, when $X_k \in V_{x_0}(\varepsilon)$). And since the larger ε is, the smaller $\delta(\varepsilon)$ is, it is intuitive to think (and

theoretically supported by the empirical EE of the bootstrap distribution) that better numerical results for the block-resampling procedure can be obtained in practice for some specific choices of the size ε, namely for choices corresponding to a maximum number of regenerative data blocks given the trajectory.

When no prior information about the structure is available, a possible selection rule relies on searching for $\varepsilon > 0$ and $x_0 \in \mathbb{R}$ so as to maximize the expected number of data blocks conditioned on the observed trajectory:

$$N_n(\varepsilon) = \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{I}\{X_i \in V_{x_0}(\varepsilon),\ \ Y_i = 1\} \mid X^{(n+1)}\right) \tag{2}$$

$$= \frac{\delta(\varepsilon)}{2\varepsilon} \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in V_{x_0}(\varepsilon)^2\}}{p(X_i, X_{i+1})}.$$

Since the transition density p and its minimum over $V_{x_0}(\varepsilon)^2$ are unknown, an empirical criterion $\hat{N}_n(\varepsilon)$ to optimize is obtained by replacing p by an estimate p_n and $\delta(\varepsilon)/2\varepsilon$ by a lower bound $\hat{\delta}_n(\varepsilon)/2\varepsilon$ for p_n over $V_{x_0}(\varepsilon)^2$ in (2).

It should be noticed that, if one uses the double splitting trick mentioned in Remark 3.3 and computes the estimates of ar $N_n(\varepsilon)$, $\delta(\varepsilon)$, etc. using an estimator of p based on the first $m_n$observations, then the second-order properties of the ARBB with estimated parameters still hold (provided that we have almost sure convergence of this estimator uniformly over the small set chosen). However, giving an exact rate in that case is more difficult.

Observe, finally, that other approaches may be used for the choice of the minorization condition; for instance, one may refer to Roberts and Rosenthal (1996) in the case of diffusion Markov processes. Some convincing simulation studies are available in Bertail and Clémençon (2003) as well as in a forthcoming companion paper.

4. Technical proofs

In the following, C, C_1, ... and K are constants which are not necessarily the same at each appearance. We denote by $\mathbb{E}^*(\cdot|X^{(n)})$ the conditional expectation given $X^{(n)}$.

Proof of Theorem 2.1. Observe first that it suffices to consider the case $\mathbb{E}_A(\tau_A) > 1$. Indeed, if $\mathbb{E}_A(\tau_A) = 1$, then $\tau_A = 1$ a.s. and the X_i are thus i.i.d. In this case the RBB exactly corresponds to the naive bootstrap and is naturally second-order correct both in the unstandardized and standardized cases up to $O(n^{-1})$ (see Hall 1992). The proof is standard and relies on checking that conditions for the validity of the EEs, established in Malinovskii (1987) (see Theorem 1 therein) and in Bertail and Clémençon (2004) (see Theorem 5.1 therein) respectively, are fulfilled (uniformly in n) for the RBB reconstructed series.

To check condition (i), observe first that the characteristic function $\mathbb{E}^*(e^{it\{f(\mathcal{B}_j^*)-\mu_n(f)l(\mathcal{B}_j^*)\}}|X^{(n)}) = (l_n - 1)^{-1}\sum_{j=1}^{l_n-1} \exp(it\{f(\mathcal{B}_j) - \mu_n(f)l(\mathcal{B}_j)\})$ converges to $\mathbb{E}_A(\exp(it\{f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j)\}))$, $\mathbb{P}_\nu$-a.s. uniformly over any compact set. The bootstrap

Cramér condition then follows from standard arguments (see Datta and McCormick 1993, for instance).

Conditions (ii)–(v) follow from the strong law of large numbers for Markov chains. It follows from Bertail and Clémençon (2004) (see their equations (8), (9)) that, as $n \to \infty$, $H^{\mathrm{U}}_{\mathrm{RBB}}(x)$ and $H^{\mathrm{S}}_{\mathrm{RBB}}(x)$ admit an EE up to order $O_{\mathbb{P}_\nu}(n^{-1})$, replacing the true quantities by the empirical ones. Here the empirical skewness $k_3^{(n)}(f)$ and the bias $b^{(n)}(f)$ are given by

$$k_3^{(n)}(f) = (\mathbb{E}^*(l(\mathcal{B}_j^*)|X^{(n)}))^{-1}\{M_{3,A}^{(n)}(f) - 3\sigma_n^2(f)\beta^{(n)}(f)\}/\sigma_n^3(f)$$

$$b^{(n)}(f) = -(E^*(l(\mathcal{B}_j^*)|X^{(n)}))^{-1}\beta^{(n)}(f)/\sigma_n(f),$$

with $M_{3,A}^{(n)}(f) = \mathbb{E}^*(\{f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)\}^3|X^{(n)})$ and $\beta^{(n)}(f) = \mathbb{E}^*(\{f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)\}l(\mathcal{B}_j^*)|X^{(n)})$. Now it is easy to show by standard CLT arguments that each of these terms converges at the rate $n^{-1/2}$ to the corresponding terms in the EE of $\mu_n(f)$. The proof is then concluded by observing that the EE of the true distributions and that of the RBB distribution match up to $O_{\mathbb{P}_\nu}(n^{-1})$ in the unstandardized case and in the standardized case as well. □

Proof of Theorem 3.1. In the following, denote by $\tau_S = \tau_S(1) = \inf\{n \geqslant 1, X_n \in S\}$ and $\tau_S(j) = \inf\{n > \tau_S(j-1), X_n \in S\}$, $j \geqslant 2$, the times of the successive visits to S. We consider the joint distribution such that, conditionally on the sample path $X^{(n+1)} = (X_1, \ldots, X_{\tau_S(1)}, \ldots, X_{\tau_S(L_n)}, \ldots, X_{n+1})$, denoting by $L_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S\}$ the number of visits of X to the small set S between time 1 and time n, the $(Y_i, \hat{Y}_i)$ are drawn independently for $1 \leqslant i \leqslant n$, so that

$$Y_{\tau_S(k)} \sim \mathrm{Ber}\big(\delta\phi\big(X_{\tau_S(k)+1}\big)/p(X_{\tau_S(k)}, X_{\tau_S(k)+1})\big),$$

$$\hat{Y}_{\tau_S(k)} \sim \mathrm{Ber}\big(\delta\phi\big(X_{\tau_S(k)+1}\big)/p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})\big),$$

and, if $p(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \leqslant p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$,

$$\mathbb{P}(\hat{Y}_{\tau_S(k)} = 1, Y_{\tau_S(k)} = 0|X^{(n+1)}) = p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1}) - p(X_{\tau_S(k)}, X_{\tau_S(k)+1}),$$

$$\mathbb{P}(\hat{Y}_{\tau_S(k)} = 0, Y_{\tau_S(k)} = 1|X^{(n+1)}) = 0,$$

and, if $p(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \geqslant p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$,

$$\mathbb{P}(\hat{Y}_{\tau_S(k)} = 0, Y_{\tau_S(k)} = 1|X^{(n+1)}) = p(X_{\tau_S(k)}, X_{\tau_S(k)+1}) - p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1}),$$

$$\mathbb{P}(\hat{Y}_{\tau_S(k)} = 1, Y_{\tau_S(k)} = 0|X^{(n+1)}) = 0,$$

for $k \in \{1, \ldots, L_n\}$, and that for all $i \in \{1, \ldots, n\}\backslash\{\tau_S(k), 1 \leqslant k \leqslant L_n\}$, $Y_i = \hat{Y}_i \sim \mathrm{Ber}(\delta)$. Hence, we deduce that, for $1 \leqslant k \leqslant L_n$,

$$\mathbb{P}(\hat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)}|X^{(n+1)}) = \left|\frac{\delta\phi\big(X_{\tau_S(k)+1}\big)}{p(X_{\tau_S(k)}, X_{\tau_S(k)+1})} - \frac{\delta\phi\big(X_{\tau_S(k)+1}\big)}{p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})}\right| \quad \text{a.s.},$$

$$l_1(P^{(n)}, \hat{P}^{(n)}) = \sum_{k=1}^{n-1} 2^{-k}\mathbb{E}\left[\mathbb{I}\{X_k \in S\}\left|\frac{\delta\phi(X_{k+1})}{p(X_k, X_{k+1})} - \frac{\delta\phi(X_{k+1})}{p_n(X_k, X_{k+1})}\right|\right].$$

Observe that we almost surely have

$$\left|\frac{\delta\phi(X_{k+1})}{p(X_k, X_{k+1})} - \frac{\delta\phi(X_{k+1})}{p_n(X_k, X_{k+1})}\right| \leqslant \mathbb{I}\{X_{k+1} \in S\}\frac{|p(X_k, X_{k+1}) - p_n(X_k, X_{k+1})|}{p(X_k, X_{k+1})}.$$

Consequently, using the fact that $p(X_k, X_{k+1}) \geqslant \delta\phi(X_{k+1})$, we may write

$$l_1(P^{(n)}, \hat{P}^{(n)}) \leqslant \sum_{k=1}^{n-1} 2^{-k}\mathbb{E}\left[\mathbb{I}\{X_k \in S,\ \ X_{k+1} \in S\}\frac{|p(X_k, X_{k+1}) - p_n(X_k, X_{k+1})|}{\delta\phi(X_{k+1})}\right].$$

Hence, under (i), we have

$$l_1(P^{(n)}, \hat{P}^{(n)}) \leqslant \left(\delta \inf_{x\in S}\phi(x)\right)^{-1}\sum_{k=1}^{n-1} 2^{-k}\mathbb{E}\left[\sup_{(x,y)\in S^2}|p(x, y) - p_n(x, y)|\right].$$

Thus, the desired bound results from the asymptotic properties of p_n. □

Proof of Theorem 3.2. In order to make the exposition of the proof much simpler, we only consider the case when f is bounded. The same argument applies to the general unbounded case with only slight and obvious modifications. The proof is based on the study of the closeness between the distribution of the blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{l_n}$ dividing the segment $X^{(n+1)} = (X_1, \ldots, X_{n+1})$ according to the l_n consecutive visits of (X_i, Y_i) to the atom $A_{\mathcal{M}} = S \times \{1\}$ between time 1 and time n, and the distribution of the blocks $\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_{\hat{l}_n}$ dividing $X^{(n+1)}$ according to the $\hat{l}_n$ successive visits of $(X_i, \hat{Y}_i)$ to $S \times \{1\}$ conditioned on $X^{(n+1)}$. Let us assume that, conditionally on $X^{(n+1)}$, the $(Y_i, \hat{Y}_i)$ are drawn as supposed in the proof of Theorem 3.1. We use the notation $l_n = \sum_{i=1}^{n}\mathbb{I}\{X_i \in S,\ Y_i = 1\}$, $\tau_{A_{\mathcal{M}}} = \tau_{A_{\mathcal{M}}}(1) = \inf\{n \geqslant 1, (X_n, Y_n) \in A_{\mathcal{M}}\}$, $\tau_{A_{\mathcal{M}}}(j+1) = \inf\{n > \tau_{A_{\mathcal{M}}}(j), \quad (X_n, Y_n) \in A_{\mathcal{M}}\}$, $l(\mathcal{B}_j) = \tau_{A_{\mathcal{M}}}(j) - \tau_{A_{\mathcal{M}}}(j)$ for $j \geqslant 1$. Set $n_{A_{\mathcal{M}}} = \tau_{A_{\mathcal{M}}}(l_n) - \tau_{A_{\mathcal{M}}}(1)$ and let $\mu_n(f) = n_{A_{\mathcal{M}}}^{-1}\sum_{j=1}^{l_n} f(\mathcal{B}_j)$ and $\sigma_n^2(f) = n_{A_{\mathcal{M}}}^{-1}\sum_{j=1}^{l_n}\{f(\mathcal{B}_j) - \mu_n(f)\mathrm{l}(\mathcal{B}_j)\}^2$ be the respective counterparts of $\hat{\mu}_n(f)$ and $\hat{\sigma}_n^2(f)$ based on the regenerative blocks.

For the sake of the simplicity, we introduce further notation and denote by $\tau_1 = \tau_{A_{\mathcal{M}}}$ (or $\hat{\tau}_1 = \hat{\tau}_{\mathrm{A}_{\mathcal{M}}}$) the (random) time corresponding to the first visit to $S \times \{1\}$ of $(X_i, Y_i)_{1\leqslant i\leqslant n}$ (or of $(X_i, \hat{Y}_i)_{1\leqslant i\leqslant n}$) as well as the time $\tau_2 = \tau_{A_{\mathcal{M}}}(l_n)$ (or $\hat{\tau}_2 = \hat{\tau}_{A_{\mathcal{M}}}(\hat{l}_n)$) corresponding to the last visit.

Lemma 4.1. *Let $\gamma \geqslant 2$. Under $\mathcal{H}_2(2\gamma, \nu)$ and $\mathcal{H}_3$, there exists a constant C such that*

$$\mathbb{E}_\nu(|\hat{\tau}_i - \tau_i|^\gamma) \leqslant C\alpha_n^{1/2}, \qquad \text{for } i \in \{1, 2\}.$$

Proof. Under $\mathcal{H}_4$ and $\mathcal{H}_5$, note that $\delta\phi\big(X_{\tau_S(k)+1}\big)/p(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ and $\delta\phi\big(X_{\tau_S(k)+1}\big)/p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ are bounded from below by $q = \delta\inf_{x\in S}\phi(x)/R$. Given the joint distribution of the $(Y_i, \hat{Y}_i)$ (refer to the proof of Theorem 3.1) and in particular that, for any $k \in \{1, \ldots, L_n\}$,

$$\mathbb{P}(Y_{\tau_S(k)} \neq \hat{Y}_{\tau_S(k)} | X^{(n+1)}) \leqslant (\delta \inf_{x \in S} \phi(x))^{-1} \sup_{(x,y) \in S^2} |p_n(x, y) - p(x, y)|, \tag{3}$$

one obtains the following bound for the conditional expectation:

$$\mathbb{E}(|\hat{\tau}_1 - \tau_1|^\gamma | X^{(n+1)}) \leqslant C \sum_{1 \leqslant l < k \leqslant L_n} (\tau_S(k) - \tau_S(l))^\gamma q(1-q)^{k-1} \times \sup_{(x,y) \in S^2} |p_n(x, y) - p(x, y)|.$$

Using the Cauchy–Schwarz inequality and assumption $\mathcal{H}_3$, easy calculations yield the following bound for the (unconditional) expectation

$$\mathbb{E}_\nu(|\hat{\tau}_i - \tau_i|^\gamma) \leqslant C \alpha_n^{1/2} \left(\sum_{k=1}^{\infty} k^2 (1-q)^k \mathbb{E}_\nu(\tau_S(k)^{2\gamma}) \right)^{1/2}.$$

Furthermore, it straightforwardly follows from the identity $\tau_S(k) = \tau_S + \sum_{l=1}^{k-1} \{\tau_S(l+1) - \tau_S(l)\}$ that $\mathbb{E}_\nu(\tau_S(k)^{2\gamma})$ is bounded by $2^{2\gamma}\{\mathbb{E}_\nu(\tau_S^{2\gamma}) + (k-1)^{2\gamma} \times \sup_{x \in S} \mathbb{E}_x(\tau_S^{2\gamma})\}$ for all k. Under $\mathcal{H}_2(2\gamma, \nu)$ the bound is thus established when $i = 1$.

The case $i = 2$ follows from a similar argument. □

Let $g : (E, \mathcal{E}) \to \mathbb{R}$ be measurable and bounded. Set $g(\mathcal{B}_j) = \sum_{i=1+\tau_{A_\mathcal{M}}(j)}^{\tau_{A_\mathcal{M}}(j+1)} g(X_i)$ and consider $T_n^{(M)}(g) = n^{-1} \sum_{j=1}^{l_n - 1} g(\mathcal{B}_j)^M$, $\hat{T}_n^{(M)}(g) = n^{-1} \sum_{j=1}^{\hat{l}_n - 1} g(\hat{\mathcal{B}}_j)^M$ for $M \geqslant 0$, where by convention $T_n^{(M)}(g) = 0$ ($\hat{T}_n^{(M)}(g) = 0$) when $l_n \leqslant 1$ (when $\hat{l}_n \leqslant 1$). The following lemma provides an asymptotic bound for $D_n^{(M)}(g) = |T_n^{(M)}(g) - \hat{T}_n^{(M)}(g)|$, when $M \in \{1, 2, 3\}$.

Lemma 4.2. *Let $\gamma \geqslant 6$. Under $\mathcal{H}_2(\gamma, \nu)$ and $\mathcal{H}_3$ we have, as $n \to \infty$,*

$$D_n^{(1)}(g) = O_{\mathbb{P}_\nu}(n^{-1} \alpha_n^{1/2}), \tag{4}$$

$$D_n^{(k)}(g) = O_{\mathbb{P}_\nu}(\alpha_n), \qquad \text{for } k = 2, 3. \tag{5}$$

Proof. Bound (4) immediately follows from Lemma 4.1.

Let $N_n = \sum_{k=\max(\hat{\tau}_1, \tau_1)}^{\min(\hat{\tau}_2, \tau_2)} \mathbb{I}\{Y_{\tau_S(k)} \neq \hat{Y}_{\tau_S(k)}\}$ be the number of times when X_i visits S and $\hat{Y}_i$ differs from Y_i simultaneously between time $\max(\hat{\tau}_1, \tau_1)$ and time $\min(\hat{\tau}_2, \tau_2)$ (with the usual convention with regard to empty summation). We introduce the corresponding successive random times

$$t_1 = \inf\{\tau_S(k);\ \max(\hat{\tau}_1, \tau_1) \leqslant \tau_S(k) \leqslant \min(\hat{\tau}_2, \tau_2),\ Y_{\tau_S(k)} \neq \hat{Y}_{\tau_S(k)}\},$$

$$t_{j+1} = \inf\{\tau_S(k);\ t_j < \tau_S(k) \leqslant \min(\hat{\tau}_2, \tau_2),\ Y_{\tau_S(k)} \neq \hat{Y}_{\tau_S(k)}\},$$

with $j = 1, \ldots, N_n - 1$. And for $1 \leqslant j \leqslant N_n$, we denote by $t_j^{(1)}$ ($t_j^{(2)}$) the last time before (the first time after) t_j when, simultaneously, X_i visits S and Y_i or $\hat{Y}_i$ is equal to one, between time 0 and time n. We can check that

$$|D_n^{(2)}(g)| \leqslant \frac{\|g\|_\infty^2}{n}\left\{(\hat{\tau}_1-\tau_1)^2+(\hat{\tau}_2-\tau_2)^2+2\sum_{j=1}^{N_n}(t_j^{(2)}-t_j)(t_j-t_j^{(1)})\right\}$$

$$\leqslant \frac{\|g\|_\infty^2}{n}\left\{(\hat{\tau}_1-\tau_1)^2+(\hat{\tau}_2-\tau_2)^2+\sum_{j=1}^{N_n}\{(t_j^{(2)}-t_j)^2+(t_j-t_j^{(1)})^2\}\right\}.$$

Set $t_j = t_j^{(2)} = 0$ for $j > N_n$. By arguments similar to those used before, one can easily show that there exist constants $c_r(q)$ depending only on q, such that $\mathbb{E}_\nu((t_j^{(2)}-t_j)^r) \leqslant c_r(q)\sup_x \mathbb{E}_x(\tau_S^r)$ for any $j \geqslant 1$, $r \leqslant \gamma$. By the Cauchy–Schwarz inequality we have, for any deterministic sequence of positive integers m_n,

$$\mathbb{E}_\nu\left(\sum_{j=1}^{N_n}(t_j^{(2)}-t_j)^2\right) \leqslant \sum_{j=1}^{m_n}\mathbb{E}_\nu((t_j^{(2)}-t_j)^2)+\mathbb{E}_\nu\left(\sum_{j=1}^{n}(t_j^{(2)}-t_j)^2\mathbb{I}\{N_n > m_n\}\right)$$

$$\leqslant C_1 m_n + C_2 n(\mathbb{P}(N_n > m_n))^{1/2},$$

where $C_1 = \sup_{x\in S}\mathbb{E}_x(\tau_S^2)c_2(q)$ and $C_2 = \sup_{x\in S}\mathbb{E}_x(\tau_S^4)c_4(q)$. The probability that Y_i differs from $\hat{Y}_i$ is bounded by $q_n = (\delta \inf_{x\in S}\phi(x))^{-1}\alpha_n$ and the $(Y_{\tau_S(k)}, \hat{Y}_{\tau_S(k)})$ are drawn independently conditionally on $X^{(n+1)}$. Hence, by using the Chebyshev exponential inequality, we derive that

$$\mathbb{P}(N_n > m_n) \leqslant \mathrm{e}^{-m_n}\mathbb{E}(e^{N_n}) \leqslant \mathrm{e}^{-m_n}(1+q_n\mathrm{e})^n. \tag{6}$$

Now by choosing $m_n \uparrow \infty$ such that $m_n/n \to 0$ and $n\alpha_n/m_n \to 0$ as $n \to \infty$, we deduce from (6) that $n^{-1}\mathbb{E}_\nu(\sum_{j=1}^{N_n}(t_j^{(2)}-t_j)^2) = O(\alpha_n)$, as $n \to \infty$. By an analogous argument, one can show that $n^{-1}\mathbb{E}_\nu(\sum_{j=1}^{N_n}(t_j-t_j^{(1)})^2) = O(\alpha_n)$, as $n \to \infty$. $D_n^{(3)}(g)$ may be treated similarly. □

From these lemmas, we deduce the following:

Lemma 4.3. *Under the assumptions of Theorem 3.2, we have the following convergences in $\mathbb{P}_\nu$-probability as $n \to \infty$:*

$$n^{1/2}(\hat{\mu}_n(f)-\mu_n(f)) \to 0, \tag{7}$$

$$\alpha_n^{-1/2}|\hat{l}_n/n - l_n/n| \to 0. \tag{8}$$

Moreover, we have for $k = 2, 3$,

$$n^{-1}\sum_{j=1}^{l_n-1} l(\mathcal{B}_j)^k - n^{-1}\sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)^k = O_{\mathbb{P}_\nu}(\alpha_n), \tag{9}$$

$$n^{-1}\sum_{j=1}^{l_n-1} f(\mathcal{B}_j)^k - n^{-1}\sum_{j=1}^{\hat{l}_n-1} f(\hat{\mathcal{B}}_j)^k = O_{\mathbb{P}_\nu}(\alpha_n), \tag{10}$$

$$n^{-1}\sum_{j=1}^{l_n-1} l(\mathcal{B}_j)f(\mathcal{B}_j) - n^{-1}\sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)f(\hat{\mathcal{B}}_j) = O_{\mathbb{P}_\nu}(\alpha_n). \tag{11}$$

Proof. From (4) in Lemma 4.2 with $g \equiv 1$, it follows that $|\hat{n}_{A_\mathcal{M}}/n - n_{A_\mathcal{M}}/n| = O_{\mathbb{P}_\nu}(\alpha_n^{1/2} n^{-1})$ as $n \to \infty$. Given that $n_{A_\mathcal{M}}/n \to 1$ $\mathbb{P}_\nu$-a.s. as $n \to \infty$, this combined with (4) again with $g = f$ obviously yields (7).

Now observe that $|\hat{l}_n/n - l_n/n| \leqslant n^{-1}\sum_{k=1}^{L_n} \mathbb{I}\{Y_{\tau_S(k)} \neq \hat{Y}_{\tau_S(k)}\}$. Using again the fact that conditionally on $X^{(n+1)}$ the $(Y_{\tau_S(k)}, \hat{Y}_{\tau_S(k)})$ are drawn independently for $k = 1, \ldots, L_n$ and (3), we have

$$\mathbb{E}\Big(|\hat{l}_n/n - l_n/n|\,|X^{(n+1)}\Big) \leqslant n^{-1} L_n (\delta \inf_{x\in S}\phi(x))^{-1} \sup_{(x,y)\in S^2} |p_n(x, y) - p(x, y)|.$$

Since $L_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S\} \leqslant n$, by taking the expectation one obtains that $\mathbb{E}_\nu(|\hat{l}_n/n - l_n/n|) = O(\alpha_n^{1/2})$, as $n \to \infty$.

Now (9) straightforwardly results from Lemma 4.2 with $g \equiv 1$ (as does (10) with $g = f$). And (11) may be proved by simply noticing that

$$2\mu(f)\sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)f(\hat{\mathcal{B}}_j) = \sum_{j=1}^{\hat{l}_n-1} \bar{f}(\hat{\mathcal{B}}_j)^2 - \sum_{j=1}^{\hat{l}_n-1} f(\hat{\mathcal{B}}_j)^2 - \mu(f)^2 \sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)^2,$$

and applying Lemma 4.2 to each component on the right-hand side (by successively taking $g(x)$ equal to $\bar{f}(x) = f(x) - \mu(f)$, $f(x)$ and 1). □

Now one can easily deduce from these results that $\hat{\sigma}_n^2(f) - \sigma_n^2(f) \to 0$ in $\mathbb{P}_\nu$-probability, as $n \to \infty$. Hence, given that $\sigma_n^2(f) \to \sigma_f^2$ in $\mathbb{P}_\nu$-probability, as $n \to \infty$ (see the preliminary remarks in Section 2.4.1) the consistency of $\hat{\sigma}_n^2(f)$ is established. Finally, combining this with (7) and the CLT for the sample mean $\mu_n(f)$ related to the split chain (see Section 2.4.1) proves that $\tilde{n}_{A_\mathcal{M}}^{1/2}\hat{\sigma}_n(f)^{-1}(\hat{\mu}_n(f) - \mu(f)) \to \mathcal{N}(0, 1)$ in $\mathbb{P}_\nu$-distribution, as $n \to \infty$. The proof of theorem 3.2 is complete. □

Proof of Theorem 3.3. In what follows, we write $Z_n \overset{\mathbb{P}^*}{\to} Z$ in $\mathbb{P}_\nu$-probability ($\mathbb{P}_\nu$-a.s.) along the sample when

$$\mathbb{P}^*(|Z_n - Z| > \varepsilon | X^{(n+1)}) \overset{n\to\infty}{\to} 0 \text{ in } \mathbb{P}_\nu\text{-probability } (\mathbb{P}_\nu\text{-a.s.}).$$

The unstudentized case. The result is proved by following line by line the classical argument establishing the CLT for regenerative processes (see Theorem 17.2.2 in Meyn and

Tweedie 1996). The latter relies on approximating the summation over a random number of regenerative blocks by a sum involving a deterministic number of blocks. Note first that, conditioned on $X^{(n+1)}$, the ARBB sequence (though not Markovian) defines a regenerative process with i.i.d. segments $\hat{\mathcal{B}}_j^*$, $j \geqslant 1$. By the law of large numbers, we have $(l_n - 1)^{-1} \sum_{j=1}^{l_n - 1} l(\mathcal{B}_j) \to \mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}})$, $\mathbb{P}_\nu$-a.s. Lemma 4.3 thus entails that as $n \to \infty$,

$$\mathbb{E}^*(l(\hat{\mathcal{B}}_1^*) X^{(n+1)}) = \frac{1}{\hat{l}_n - 1} \sum_{j=1}^{\hat{l}_n - 1} l(\hat{\mathcal{B}}_j) \to \mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}}) \qquad \text{in } \mathbb{P}_\nu\text{-probability.}$$

We also have $\mathbb{E}^*(l(\hat{\mathcal{B}}_1^*)^2 | X^{(n+1)}) \to \mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}}^2) < \infty$ in $\mathbb{P}_\nu$-probability as $n \to \infty$. This implies in particular that, as $n \to \infty$,

$$\frac{l(\hat{\mathcal{B}}_j^*)}{n} \xrightarrow{\mathbb{P}^*} 0 \text{ and } \frac{n_{A_\mathcal{M}}^*}{n} \xrightarrow{\mathbb{P}^*} 1 \text{ in } \mathbb{P}_\nu\text{-probability along the sample.} \tag{12}$$

The law of large numbers and Lemma 4.3 yield $l_n^{*-1} \sum_{i=1}^{l_n^*} l(\hat{\mathcal{B}}_j^*) - (\hat{l}_n - 1)^{-1} \sum_{i=1}^{\hat{l}_n - 1} l(\hat{\mathcal{B}}_j) \xrightarrow{\mathbb{P}^*} 0$, $\mathbb{P}_\nu$-a.s., as $n \to \infty$. This entails that $l_n^{*-1} \sum_{i=1}^{l_n^*} l(\hat{\mathcal{B}}_j^*) - \mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}}) \xrightarrow{\mathbb{P}^*} 0$ in $\mathbb{P}_\nu$-probability along the sample. We deduce that $l_n^*/n - \mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}})^{-1} \xrightarrow{\mathbb{P}^*} 0$ in $\mathbb{P}_\nu$-probability along the sample. Now with the same arguments as in Theorem 17.2.2 in Meyn and Tweedie (1996), that is, combining (12) and the Markov inequality, we obtain, for n large enough,

$$\frac{\mu_n^*(f) - \hat{\mu}_n(f)}{n_{A_\mathcal{M}}^{*-1/2} \hat{\sigma}_n(f)^{1/2}} = \frac{\sum_{j=1}^{1 + \lfloor n \mathbb{E}_{A_\mathcal{M}}(\tau_A)^{-1} \rfloor} \{ f(\mathcal{B}_j^*) - \hat{\mu}_n(f) l(\mathcal{B}_j^*) \}}{n^{1/2} \hat{\sigma}_n(f)^{1/2}} + o_{\mathbb{P}*}(1)$$

along the sample in $\mathbb{P}_\nu$-probability, as $n \to \infty$. Now it is sufficient to apply the classical bootstrap CLT (see Bickel and Freedmann 1981) to the i.i.d. random variables $\{ f(\mathcal{B}_j^*) - \hat{\mu}_n(f) l(\mathcal{B}_j^*) \}_{j \geqslant 1}$. These random variables are centred with variance $\tilde{n}_{A_\mathcal{M}} \hat{\sigma}_n^2(f)/(\hat{l}_n - 1)$, which converges to $\mathbb{E}_{A_\mathcal{M}}(\tau_{A_\mathcal{M}}) \sigma_f^2$ in $\mathbb{P}_\nu$-probability under the hypotheses of Theorem 3.3 (see Theorem 3.2 and Lemma 4.3).

The studentized case. We essentially have to prove that, as $n \to \infty$, $\sigma_n^*(f) - \hat{\sigma}_n(f) \xrightarrow{\mathbb{P}^*} 0$ in $\mathbb{P}_\nu$-probability along the sample. With arguments similar to those used in the unstudentized case, one can easily show that, in $\mathbb{P}_\nu$-probability along the sample,

$$\sigma_n^*(f)^2 = n^{-1} \sum_{j=1}^{1 + \lfloor n / \mathbb{E}_{A_\mathcal{M}} \tau_A) \rfloor} \{ f(\mathcal{B}_j^*) - \hat{\mu}_n(f) l(\mathcal{B}_j^*) \}^2 + o_{\mathbb{P}^*}(1),$$

and the result follows also from standard bootstrap results in the i.i.d case.

Proof of Theorem 3.4. The EE proved in Malinovskii (1987) – along with that in Bertail and Clémençon (2004) in the studentized case – for atomic chains straightforwardly extends to the case of a general positive recurrent chain X by applying the latter to the split chain (X, Y) constructed via the Nummelin technique from a minorization condition $\mathcal{M}$: it is noteworthy that, though expressed using the parameters of condition $\mathcal{M}$, the coefficients in the EE are independent of the latter. In the stationary setting the EE of the sample mean in the unstudentized case is given in Malinovskii (1987), with $b^{(n)}(f)$ replaced by 0. Because

the reconstructed ARBB series is regenerative, just as in the atomic case, it is sufficient to check the hypotheses ensuring the validity of the EE for the ARBB statistic. The proof is thus similar to the proof of Theorem 2.1 except that the true regenerative blocks are replaced by the approximated ones. Lemma 4.3 allows us to control the rate of convergence of all the empirical terms appearing in the EE of the ARBB distribution up to $O_{\mathbb{P}}(\alpha_n^{1/2} n^{-1/2})$. The recentring ensures that the bias vanishes so that the two EEs match up to $O_{\mathbb{P}}(\alpha_n^{1/2} n^{-1/2})$. □

Acknowledgements

We would like to thank a referee, an associate editor and Professor McCullagh for their suggestions which greatly improved the presentation of this paper.

References

Asmussen, S. (1987) *Applied Probability and Queues*. Chichester: Wiley.

Athreya, K. and Atuncar, G. (1998) Kernel estimation for real-valued Markov chains. *Sankhya Ser. A*, **60**, 1–17.

Athreya, K. and Fuh, C. (1989) Bootstrapping Markov chains: countable case. Technical Report B-89-7, Institute of statistical Science, Academia Sinica, Taiwan.

Bertail, P. and Clémençon, S. (2003) Regenerative block-bootstrap for Markov chains (Revised version). CREST preprint no. 2004-47. http://www.crest.fr/doctravail/document/2004-47.pdf (accessed 10 February 2006).

Bertail, P. and Clémençon, S. (2004) Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. *Probab. Theory Related Fields*, **130**, 388–414.

Bertail, P. and Clémençon, S. (2005) Note on the regeneration-based bootstrap for atomic Markov chains. *Test*. To appear.

Bickel, P. and Freedman, D. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Bolthausen, E. (1982) The Berry–Esseen theorem for strongly mixing Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **60**, 283–289.

Clémençon, S. (2000) Adaptive estimation of the transition density of a regular Markov chain. *Math. Methods Statist.*, **9**, 323–357.

Clémençon, S. (2001) Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.*, **55**, 227–238.

Datta, S. and McCormick W. (1993) Regeneration-based bootstrap for Markov chains. *Canad. J. Statist.*, **21**, 181–193.

Douc, R., Fort, G., Moulines E. and Soulier P. (2004) Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, **14**, 1353–1377.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.

Franke, J., Kreiss, J.P. and Mammen, E. (2002) Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*, **8**, 1–37.

Götze, F. and Künsch, H. (1996) Second order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.*, **24**, 1914–1933.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hobert, J.P. and Robert, C.P. (2004) A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *Ann. Appl. Probab.*, **14**, 1295–1305.

Horowitz, J. (2003) Bootstrap methods for Markov processes. *Econometrica*, **71**, 1049–1082.

Jain, J. and Jamison, B. (1967) Contributions to Doeblin's theory of Markov processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **8**, 19–40.

Kalashnikov, V. (1978) *The Qualitative Analysis of the Behavior of Complex Systems by the Method of Test Functions*. Moscow: Nauka.

Lahiri, S. (2003) *Resampling Methods for Dependent Data*. New York: Springer-Verlag.

Malinovskii, V. (1987) Limit theorems for Harris Markov chains I. *Theory Probab. Appl.*, **31**, 269–285.

Malinovskii, V. (1989) Limit theorems for Harris Markov chains II. *Theory Probab. Appl.*, **34**, 252–265.

Meyn, S. and Tweedie, R. (1996) *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Nummelin, E. (1978) A splitting technique for Harris recurrent chains. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **43**, 309–318.

Orey, S (1971) *Limit Theorems for Markov Chain Transition Probabilities*. London: Van Nostrand Reinhold.

Rachev, S. and Rüschendorf, L. (1998) *Mass Transportation Problems. Vol. II: Applications*. New York: Springer-Verlag.

Roberts, G. and Rosenthal, J. (1996) Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.*, **9**, 1–21.

Smith, W. (1955) Regenerative stochastic processes. *Proc. Roy. Soc., Lond. Ser. A*, **232**, 6–31.

Thorisson, H. (2000) *Coupling, Stationarity, and Regeneration*. New York: Springer-Verlag.

Received September 2004 and revised September 2005