

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2004-47

**Regenerative Block-bootstrap
for Markov Chains**

P. BERTAIL¹
S. CLÉMENÇON²

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ CREST, Laboratoire de Statistiques.

² MODAL'X, Université Paris X, LPMA, Universités Paris VI et Paris VII.

Regenerative block-bootstrap for Markov chains¹

(revised September 2004)

Patrice Bertail,
CREST- Laboratoire de Statistiques
Stéphan Clémenton,
MODAL'X, Université Paris X,
LPMA, Universités Paris VI et Paris VII

Abstract : This paper introduces a specific Bootstrap method for positive recurrent Markov chains, based on the regenerative method and the Nummelin splitting technique. The main idea underlying this construction consists in generating a sequence of approximate pseudo-renewal times for a Harris chain X from data X_1, \dots, 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 & McCormick (1993) for bootstrapping additive functionals of type $n^{-1} \sum_{i=1}^n f(X_i)$ when the chain possesses an atom. We prove that, in the atomic case, our method inherits the accuracy of the Bootstrap in the i.i.d. case up to $O_P(n^{-1})$ under weak conditions. In the general (non 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 (up to a rate close to $O_P(n^{-1})$ for regular stationary Markov Chains) is obtained in the stationary case. Applications to specific Markovian models are discussed, together with some simulation results.

Résumé: Cet article étudie une méthode de Bootstrap spécifique pour les chaînes de Markov Harris récurrente positive, basée sur les propriétés de régénération et la technique de découpage introduite par Nummelin. L'idée principale sous-jacente à cette construction est 1) de générer une séquence de temps de renouvellement approché pour la chaîne, à partir d'observation X_1, \dots, X_n et de la connaissance des paramètres d'une condition de minoration satisfaite par la densité de transition 2) d'appliquer une variante de la méthodologie proposée par Datta et McCormick (1993) pour bootstrapper des fonctionnelles additives de chaînes de Markov atomiques. Nous montrons que, dans le cas atomique, notre méthode hérite des propriétés au second ordre du bootstrap i.i.d, jusqu'à l'ordre $O_P(n^{-1} \log(n))$ sous des conditions faibles. Dans le cas général, nous montrons la validité asymptotique de la procédure, sous l'hypothèse de l'existence d'un estimateur adéquat du noyau de transition. Nous discutons des applications à des modèles spécifiques et présentons quelques résultats de simulations.

¹**Keywords :** Bootstrap, Markov Chain, Regenerative Processes, Nummelin splitting technique.

Primary Class : 62G09, **Secondary Class :** 62G20, 60J10

1 Introduction

Many researchers have been working on transposing the naive Bootstrap method (Efron, 1979) introduced in the i.i.d. setting to a dependent setting. In most situations, stationary time series or homogeneous random fields are considered. 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 (refer to Lahiri (2003) for a recent survey and exhaustive references). However, as noticed by many authors, the results obtained by using such an approach are not completely satisfactory for the following reasons. First, the MBB approach usually requires stationarity for the observations and generally fails in a general nonstationary 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 & Künsch (1996), Lahiri (2003)). From a theoretical viewpoint, the rate of convergence of the MBB distribution is slower than the one of the Bootstrap in the i.i.d. case: at best it is of order $O_P(n^{-3/4})$ 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 $O_P(n^{-1})$ in the i.i.d. setting. Finally, the choice of the size of the blocks is a key point to get an accurate estimation. In some very particular case (the sample mean or functions of the sample mean, for which the Bootstrap may appear of lesser use), it is possible to give some indications concerning the adequate size of the blocks (see Götze & Künsch (1996), Lahiri (2003)), but this practical problem still remains open in the general case.

Recently, several authors have been interested in bootstrapping some particular types of Markov chain. On the one hand, if a Markovian model is *a priori* specified (for instance an ARMA model or a nonlinear model with a finite number of lags and i.i.d residuals, such as a GARCH model), the problem reduces then to the random sampling of estimated centered residuals in the stationary case. The properties of such a semiparametric Bootstrap are well understood since Bose (1988) (see the references therein). In the same spirit, Franke *et al* (2002) proposed a residual resampling scheme based on a preliminary nonparametric estimation of the linking function in nonlinear autoregressive models, that is asymptotically justified. Based on residual resampling ideas, Bühlmann (1997) also considered a sieve bootstrap method based on the approximation of time series by some $AR(p)$ model with a large p , eventually depending on n . This kind of Bootstrap, which presents promising results and good practical performance at the same time, is well suited to linear stationary time series rather than to general Markov chains. On the other hand, most of the recent works on the Bootstrap for Markov chains (refer to Horowitz (2003) for a comprehensive survey) follows the proposal

of Rajarshi (1990) in the case of finite state chains and Datta & McCormick (1995), which uses a nonparametric estimate of the transition probability in the Bootstrap procedure, so as to mimic the markovian underlying structure of the chain. Paparoditis & Politis (2002) have introduced a *local Markov Bootstrap*, which avoids the use of an explicit (smooth) nonparametric estimate of the transition kernel by using a local resampling scheme, but is nevertheless based on an implicit estimation of the transition probability. Unfortunately, the results obtained in that direction are weakened by the form of the hypotheses made on the chains or models considered (see Horowitz (2003)).

In this paper, we focus on a method originating from Athreya & Fuh (1989) and Datta & McCormick (1993), which exploits the renewal properties of Markov chains when a (recurrent) state is infinitely often visited. The main idea underlying this method consists in resampling a deterministic number of data blocks corresponding to regeneration cycles. However, because of some inadequate standardization, the *regeneration-based bootstrap* method proposed in Datta & McCormick (1993) is not second order correct (its rate is $O_P(n^{-1/2})$). Bertail & Cléménçon (2003) have proposed a modification of the procedure introduced by Datta & McCormick (1993), which is second order correct up to $O_P(n^{-1} \log(n))$ 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 nonstationary case, as a careful examination of the second order properties of the sample mean statistic of a positive recurrent chain based on its Edgeworth expansion (see Malinovskii (1987), Bertail & Cléménçon (2004)) shows. Our proposal consists here 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, so as to approximate the distribution of the (random) number of regeneration blocks in a series of length n and remove some bias terms. In this paper we study in particular the higher order properties of this resampling method, which we call the *regenerative block-bootstrap* (RBB), for suitably standardized functionals. We also 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.

The outline of the paper is as follows. In section 2, notations are set out and a few definitions concerning the communication structure and the stochastic stability of Markov chains are given. We then consider the particular case of Markov chains with an atom (also called *Markov models with regeneration times*, which find many applications in the field of operational research for modeling queuing/storage systems, see Asmussen (1987) and Feller (1971)). We demonstrate here the power of our method for suitably

standardized statistics: the RBB has in particular an uniform rate of convergence of order $O_P(n^{-1})$, that is the optimal rate in the i.i.d case. Unlike the MMB, 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, some basics about the regenerative method and the Nummelin splitting technique are recalled. We then extend our bootstrap methodology to general positive recurrent chains. We establish the asymptotic validity of this procedure, even in a nonstationary framework, that 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 the one 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 article. 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. These results are illustrated by some examples and simulation results in section 4. And technical proofs are detailed in section 5.

2 Bootstrapping Markov chains with an atom

2.1 Notation and definitions

We shall introduce some notation that will be needed throughout the paper (see Revuz (1984) for further detail on key notions of the Markov chain theory). Let $X = (X_n)_{n \in \mathbb{N}}$ be an aperiodic ψ -irreducible Markov chain on a countably generated state space (E, \mathcal{E}) , with transition probability Π , and initial probability distribution ν . Thus for any $B \in \mathcal{E}$ and $n \in \mathbb{N}$, we have

$$X_0 \sim \nu \text{ and } P(X_{n+1} \in B \mid X_0, \dots, X_n) = \Pi(X_n, B) \text{ a.s..}$$

In what follows, P_ν (respectively P_x for x in E) will denote the probability measure on the underlying probability space such that $X_0 \sim \nu$ (resp. $X_0 = x$), $E_\nu(\cdot)$ the P_ν -expectation (resp. $E_x(\cdot)$ the P_x -expectation), and $I\{\mathcal{A}\}$ will denote the indicator function of the event \mathcal{A} .

Assume that the chain is ψ -irreducible and possesses an accessible atom, that is to say a measurable set A such that for all x, y in A :

$$\Pi(x, \cdot) = \Pi(y, \cdot) \text{ and } \psi(A) > 0.$$

Denote by $\tau_A = \tau_A(1) = \inf \{n \geq 1, X_n \in A\}$ the hitting time on A , by $\tau_A(j) = \inf \{n > \tau_A(j-1), X_n \in A\}$ for $j \geq 2$, the successive return times to A , and by $E_A(\cdot)$ the expectation conditionally to $X_0 \in A$. Assume further that the chain is Harris recurrent, hence the probability of returning infinitely often to the atom A is equal to one, no matter the starting point: $\forall x \in E, P_x(\tau_A < \infty) = 1$. Then, it follows from the *strong Markov property* that, for any initial distribution ν , the sample paths of the chain may be divided into i.i.d. blocks of random length corresponding to consecutive visits to A

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

taking their values in the torus $T = \cup_{n=1}^{\infty} E^n$. The sequence $(\tau_A(j))_{j \geq 1}$ defines successive times at which the chain forgets its past, such random times are called *regeneration times*. When an accessible atom exists, the *stochastic stability* properties of the chain amount to properties concerning the speed of return time to the atom only. For instance, in this framework, the chain X is positive recurrent iff $E_A(\tau_A) < \infty$ (this result is known as Kac's theorem, see Theorem 10.2.2 in Meyn & Tweedie (1996)). In such a case the unique invariant probability distribution μ is the occupation measure given by:

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

For such chains, limit theorems can be derived from the application of the corresponding results to the i.i.d. blocks $(\mathcal{B}_n)_{n \geq 1}$ (see Smith (1955) for an introduction to the *regenerative method*). One may refer to Meyn & Tweedie (1996) for the LLN, CLT, LIL, Bolthausen (1980) for the Berry-Esseen theorem, and Malinovskii (1987, 89) for other refinements of the CLT. The same technique can also be applied to establish moment and probability inequalities, which are not asymptotic results (cf Cléménçon (2001)).

2.2 Preliminary remarks

Let $X^{(n)} = (X_1, \dots, X_n)$ be observations drawn from a Markov chain X valued in a state space E equipped with a countably generated σ -field \mathcal{E} , with unknown transition probability Π and initial probability distribution ν . Assume further that the chain X is positive recurrent with unknown stationary probability μ and admits an *a priori* known accessible atom A (see Example 1 in section 4 for a practical case). In the following we denote by $l_n = \sum_{i=1}^n I\{X_i \in A\}$ the number of successive visits to the atom, giving rise to $l_n + 1$ data blocks

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

with the convention $\mathcal{B}_{l_n}^{(n)} = \emptyset$ when $\tau_A(l_n) = n$. We denote by $l(\mathcal{B}_j) = \tau_A(j+1) - \tau_A(j)$, $j \geq 1$, the lengths of the regeneration blocks (note that $E(l(\mathcal{B}_j)) = E_A(\tau_A) = \mu(A)^{-1}$ for $j \geq 1$).

Let $f : E \rightarrow \mathbb{R}$ be a μ -integrable function and consider the estimator $\bar{\mu}_n(f) = n^{-1} \sum_{i=1}^n f(X_i)$ of the unknown mean $\mu(f) = E_\mu(f(X_1))$ constructed from the whole data segment $X^{(n)}$. In Bertail & Cléménçon (2004) (see Proposition 3.1) it is shown that in the case when the chain is not stationary (*i.e.* when the initial distribution ν differs from μ), the first data block \mathcal{B}_0 induces a significant bias, which cannot be estimated from a single realization $X^{(n)}$ of the chain starting from ν . It is thus impossible to approximate the second order properties of such a statistic in the nonstationary case by using a resampling method. Hence, when the matter is to consider estimators, for which one can obtain accurate bootstrap distribution estimates, it is preferable to construct them using the data collected from the first regeneration time (*i.e.* the first visit to A), so as to get rid of the first order term induced by \mathcal{B}_0 in the bias. The last (non regenerative) data block $\mathcal{B}_{l_n}^{(n)}$ induces a first order term in the bias too. And although it seems possible to estimate accurately its sampling distribution, we shall consider in what follows statistics based on the observations $X_{\tau_A}^{\tau_A(l_n)} = (\mathcal{B}_1, \dots, \mathcal{B}_{l_n-1})$ collected between the first and last visits to the atom only (the use of $\mathcal{B}_{l_n}^{(n)}$ would make the resampling method we introduce below slightly more complex on the one hand, and would make its technical study much more difficult on the other hand). In the case of the estimation of $\mu(f)$, this prevails to consider the sample mean statistic based on the data segment $(X_{\tau_A+1}, \dots, X_{\tau_A(l_n)})$, $\mu_n(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{i=1}^{l_n-1} f(X_i)$ with the convention that $\mu_n(f) = 0$ when $l_n < 2$. In the case of $\mu_n(f)$ an appropriate standardization $S_n = S_n(X_{\tau_A}^{\tau_A(l_n)})$ has been exhibited in Bertail & Cléménçon (2004) (see § 2.4.1 below).

2.3 The regenerative block-bootstrap algorithm

Although asymptotic results are only stated for the case of the sample mean in this paper, we present here a valid algorithm for general statistics T_n , for which there exists an adequate standardisation S_n : this covers the case of nondegenerate 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 shall consider are constructed from the regenerative data blocks $\mathcal{B}_1, \dots, \mathcal{B}_{l_n-1}$ only (see Bertail & Cléménçon (2004) for other types of standardization). We are thus interested in estimating accurately its sampling distribution under P_ν , say $H_{P_\nu}^{(n)}(x) = P_\nu(S_n^{-1}(T_n - \theta) \leq x)$. Given a specified parameter b_n controlling the maximum length of the bootstrap

data segment (a typical choice in practice is $b_n = n$), the RBB procedure is performed in four steps as follows.

1. Count the number of visits l_n to the atom A up to time n . And divide the observed sample path $X^{(n)} = (X_1, \dots, X_n)$ into $l_n + 1$ blocks, $\mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_{l_n-1}, \mathcal{B}_{l_n}^{(n)}$ 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 . Drop the first and last (nonregenerative) blocks.
2. Draw sequentially bootstrap data blocks $\mathcal{B}_{1,n}^*, \dots, \mathcal{B}_{k,n}^*$ independently from the empirical distribution $F_n = (l_n - 1)^{-1} \sum_{j=1}^{l_n-1} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leq j \leq l_n-1}$ conditioned on $X^{(n)}$, until the length $l^*(k) = \sum_{j=1}^k l(\mathcal{B}_{j,n}^*)$ of the bootstrap data series is larger than b_n . Let $l_n^* = \inf\{k \geq 1, l^*(k) > b_n\}$.
3. From the bootstrap data blocks generated at step 2, reconstruct a pseudo-trajectory of size b_n by binding the blocks together

$$X_{b_n}^{*(n)} = (\mathcal{B}_{1,n}^*, \dots, \mathcal{B}_{l_n^*-1,n}^*).$$

And compute the *RBB statistic* based on the bootstrap data blocks

$$T_{n,b_n}^* = T_{b_n}(X_{b_n}^{*(n)}) = T(\mathcal{B}_{1,n}^*, \dots, \mathcal{B}_{l_n^*-1,n}^*).$$

4. If $S_n = S(\mathcal{B}_1, \dots, \mathcal{B}_{l_n-1})$ is an appropriate standardization of the original statistic T_n , compute

$$S_{n,b_n}^* = S_n(X_{b_n}^{*(n)}) = S(\mathcal{B}_{1,n}^*, \dots, \mathcal{B}_{l_n^*-1,n}^*).$$

The *RBB distribution* is then given by

$$H_{RBB}(x) = P^*(S_{n,b_n}^{*-1}(T_{n,b_n}^* - T_n) \leq x \mid X^{(n)})$$

where $P^*(\cdot \mid X^{(n)})$ denotes the conditional probability given $X^{(n)}$.

Remark 2.1 In the next sections, the maximum length for the bootstrap series is chosen to be $b_n = n$. As will be shown below, the RBB is then second order correct for regular functionals. However, in the i.i.d. case, it is known that subsampling (*i.e.* choosing $b_n = o(n)$ in our case) without replacement may yield a general second-order asymptotic validity at the cost of some efficiency : see Bertail (1997) and Politis, Romano & Wolf (2000) for a recent survey. The study of the properties of such a regenerative subsampling distribution estimate goes beyond the scope of this paper and will be the subject of further investigation.

Remark 2.2 One may naturally compute a Monte-Carlo approximation to $H_{RBB}(x)$ by repeating independently the procedure above B times.

Remark 2.3 We point out that the RBB differs from the regeneration-based bootstrap proposed by Datta & McCormick (1993), which is not second order correct up to $O_P(n^{-1/2})$, (and from its modified version in Bertail & Cl  men  on (2003) too) in which the number of resampled blocks is held fixed to $l_n - 1$, conditionally to the sample. By generating this way a random number $l_n^* - 1$ of bootstrap regenerative blocks, we get a data series that somehow mimics the renewal properties of the chain, although it is not markovian. Consequently, the usual properties of the i.i.d. Bootstrap cannot be directly used for studying the RBB method, contrary to the regeneration-based bootstrap studied in Bertail & Cl  men  on (2003).

Remark 2.4 We also emphasize that the principles underlying the RBB may be applied to any (eventually continuous time) regenerative process (and not necessarily markovian).

2.4 Second order accuracy of the RBB

Pursuing the work of Datta & McCormick (1993) in which the regeneration-based bootstrap is introduced, Bertail & Cl  men  on (2003) showed that a suitably modified version of their bootstrap methodology is second order correct up to $O_P(n^{-1} \log(n))$ for the unstudentized mean, in the stationary case only. We study here 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 (refer to Hall (1992)). Since we know from Bertail (1997) that extrapolating subsampling distribution (or m out of n bootstrap distribution) yields second order correctness (at least up to $o_P(n^{-1/2})$) in an automatic way under quite weak conditions, it is also of theoretical and practical importance to determine accurately the rate of convergence (see G  tze & K  unsch (1996) for the case of the MBB). We prove that for a broad class of Markov chains (including chains with strong mixing coefficients decreasing at a polynomial rate), the accuracy reached by the RBB is of order $O_P(n^{-1})$ for the unstandardized sample mean as well as for the studentized sample mean. The rate obtained is thus similar to the optimal rate of the Bootstrap distribution in the i.i.d. case, contrary to the MBB (see G  tze & K  unsch (1996)). The proof relies on the E.E. for the studentized sample mean established in Bertail & Cl  men  on (2004), which result mainly derives from the methods used in Malinovskii (1987) to obtain the E.E. for the unstandardized sample mean (see also Malinovskii (1987, 1989) and Bolthausen (1980)).

2.4.1 Further notations and preliminary remarks

We set

$$n_A = \tau_A(l_n) - \tau_A(1) = \sum_{j=1}^{l_n-1} l(\mathcal{B}_j) \text{ and } f(\mathcal{B}_j) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i), \text{ for any } j \geq 1.$$

With these notations, we may 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, $\{f(\mathcal{B}_j) - l(\mathcal{B}_j) \mu(f)\}_{j \geq 1}$ are i.i.d. r.v.'s with mean 0 and variance

$$\sigma_F^2 = E(\{f(\mathcal{B}_j) - l(\mathcal{B}_j) \mu(f)\}^2) \quad (2)$$

In the following, we also set

$$\alpha = E_A(\tau_A) \text{ and } \beta = \text{cov}(l(\mathcal{B}_j), f(\mathcal{B}_j) - l(\mathcal{B}_j) \mu(f)).$$

Under the assumption that the expectations $E_A((\sum_{i=1}^{\tau_A} f(X_i))^2)$, $E_A(\tau_A^2)$, $E_\nu((\sum_{i=1}^{\tau_A} |f(X_i)|))$ and $E_\nu(\tau_A)$ are finite, the CLT holds (see Theorem 17.2.2 in Meyn & Tweedie (1996)). We have as $n \rightarrow \infty$

$$n^{-1/2}(\mu_n(f) - \mu(f)) \rightarrow \mathcal{N}(0, \sigma_f^2) \text{ in distribution under } P_\nu,$$

with the asymptotic variance $\sigma_f^2 = \alpha^{-1} \sigma_F^2$.

For the MBB, the choice of the standardization, the bias it induces and the definition of its Bootstrap counterpart are key points to obtain the second order validity of the method. In our regenerative setting, we avoid this problem with the RBB, since the following estimate of the asymptotic variance $\sigma_f^2 = \alpha^{-1} \sigma_F^2$ based on the expression (1) 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 & Cléménçon (2004). A straightforward application of the SLLN for positive recurrent Markov chains shows it is strongly consistent. Under some further regularity conditions, Bertail & Cléménçon (2004) have also shown that its bias is of order $O(n^{-1})$ and it is asymptotically normal. As will be shown below, this standardization does not weaken the performance of the RBB, while the

standardization of the MBB distribution in the strong mixing case is the main barrier to achieve good performance (as shown by Götze & Künsch (1996)). Moreover, in opposition to the MBB, the Bootstrap counterparts in the studentized case may be defined straightforwardly in our regenerative setting. With $n_A^* = \sum_{j=1}^{l_n^*-1} l(\mathcal{B}_j^*)$, we define

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

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 fulfills the following conditions,*

- (i) (Cramer condition) $\overline{\lim}_{|t| \rightarrow \infty} |E_A(\exp(it(\sum_{i=1}^{\tau_A} \{f(X_i) - \mu(f)\}))| < 1$.
- (ii) (Nondegeneracy of the asymptotic variance) $\sigma_f^2 > 0$.
- (iii) ("Block moment conditions") For some integer $s \geq 2$,

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

- (iv) ("Block moment conditions" for the initial law ν)

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

- (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 $(f(\mathcal{B}_j) - l(\mathcal{B}_j)\mu(f))^2$'s.

Then, under assumptions (i)-(iv) with $s = 6 + \varepsilon$, the RBB distribution estimate for the unstandardized sample mean is second order accurate in the following sense

$$\Delta_n^U = \sup_{x \in \mathbb{R}} |H_{RBB}^U(x) - H_\nu^U(x)| = O_{P_\nu}(n^{-1}), \text{ as } n \rightarrow \infty,$$

with $H_{RBB}^U(x) = P^*(n_A^{1/2} \sigma_f^{-1} \{\mu_n^*(f) - \mu_n(f)\} \leq x \mid X^{(n)})$ and $H_\nu^U(x) = P_\nu(n_A^{1/2} \sigma_f^{-1} \{\mu_n(f) - \mu(f)\} \leq x)$.

Under the assumptions (i)-(v) with $s = 8 + \varepsilon$, the RBB distribution estimate for the studentized sample mean is also second order correct

$$\Delta_n^S = \sup_{x \in \mathbb{R}} |H_{RBB}^S(x) - H_\nu^S(x)| = O_{P_\nu}(n^{-1}), \text{ as } n \rightarrow \infty,$$

with $H_{RBB}^S(x) = P^*(n_A^{*-1/2} \sigma_n^{*-1}(f) \{ \mu_n^*(f) - \mu_n(f) \} \leq x \mid X^{(n)})$ and $H_\nu^S(x) = P_\nu(n_A^{1/2} \sigma_n^{-1}(f) \{ \mu_n(f) - \mu(f) \} \leq x)$.

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 models widely used in practice (see § 2.4 in Meyn & Tweedie (1996), Feller (1971) and Asmussen (1987) for more details on such models and for instance the popular storage model in Example 1 of section 4). We point out that the relationship between the "block moment" condition (iii) and the rate of decay of mixing coefficients has been investigated in Bolthausen (1982): for instance condition (iii) is typically fulfilled when f is bounded as soon as the strong mixing coefficients sequence decreases at an arithmetic rate $n^{-\rho}$, for some $\rho > s - 1$. Condition (v) is a technical assumption used in Bertail & Cléménçon (2004) to obtain an E.E. in the studentized case. As may be shown by a straightforward calculation, if the density of the $\sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} (f(X_i) - \mu(f))$'s is bounded then it holds for $N = 2$.

Remark 2.5 The same results holds a.s. up to $O(n^{-1} \log \log(n)^{1/2})$, just like in the i.i.d. case under the same moment conditions. This straightforwardly results from the LIL applied to the empirical moments of the blocks appearing in the E.E. of the RBB distribution. Notice also that if one is interested in getting the second order validity up to $o(n^{-1/2})$ only, then a careful examination of Malinovskii (1987) and Bertail & Cléménçon (2004) shows that $s > 3$ (respectively $s > 4$) is sufficient in the standardized case (resp. the studentized case).

3 Approximate regenerative block-bootstrap for positive recurrent chains

3.1 General Harris chains and the splitting technique

For clarity's sake, we now briefly recall the *splitting technique* introduced in Nummelin (1978), which allows 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. First, consider the following notion.

Definition 3.1 For a Markov chain valued in a state space (E, \mathcal{E}) with transition probability Π , a set $S \in \mathcal{E}$ is said to be small if there exist an integer $m > 0$, a probability measure Φ supported by S , and $\delta > 0$ such that

$$\forall x \in S, \forall A \in \mathcal{E}, \quad \Pi^m(x, A) \geq \delta\Phi(A),$$

denoting by Π^m the m -th 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 & Jamison (1967)). Let X be a Harris recurrent chain. Let us now make precise how to construct the atomic chain onto which the initial chain X is embedded, from a set on which an iterate Π^m of the transition probability is uniformly bounded below. Suppose that the chain X satisfies $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ for some measurable set S such that $\psi(S) > 0$. Even if it entails to replace the chain $(X_n)_{n \in \mathbb{N}}$ by the chain $((X_{nm}, \dots, X_{n(m+1)-1}))_{n \in \mathbb{N}}$, we suppose $m = 1$. 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 $P_{\nu, \mathcal{M}}$ whose construction relies on the following randomization of the transition probability Π each time the chain $(X_n)_{n \in \mathbb{N}}$ hits the set S (note that it happens almost surely 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 $Y_n = 0$ (that happens with probability $1 - \delta$), then X_{n+1} is distributed according to the probability measure $(1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta\Phi(\cdot))$.

We now have constructed a bivariate Markov chain $X^{\mathcal{M}} = ((X_n, Y_n))_{n \in \mathbb{N}}$, called the *split chain*, taking its values in $E \times \{0, 1\}$ and which possesses an atom, namely $S \times \{1\}$. The whole point of this construction consists in the fact that $X^{\mathcal{M}}$ inherits all the communication and stochastic stability properties from X (irreducibility, Harris recurrence,...), in particular (for the case $m = 1$) the blocks constructed for the split chain are i.i.d.. Hence the splitting method enables to extend the regenerative method and to establish all of the results known for atomic chains, to general Harris chains. It should be noticed that if X satisfies $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ for $m > 1$, the resulting blocks are not independent anymore but 1-dependent, a form of dependence which may be also easily handled. For simplicity's sake, we shall omit the subscript \mathcal{M} in what follows and abusively denote by P_ν the extensions of the underlying probability we consider.

3.2 ”Approximating the regenerative blocks”

Our method for bootstrapping positive recurrent chains relies on approximating the splitting construction and is based on the knowledge of the parameters of a minorization condition \mathcal{M} fulfilled by the chain. In the following, we suppose that condition \mathcal{M} is fulfilled with $m = 1$ for the sake of the simplicity. We assume further that the family of the conditional distributions $\{\Pi(x, dy)\}_{x \in E}$ and the initial distribution ν are dominated by a σ -finite measure λ of reference, so that $\nu(dy) = f(y)\lambda(dy)$ and $\Pi(x, dy) = p(x, y)\lambda(dy)$, for all $x \in E$. Notice that the minorization condition entails that Φ is absolutely continuous with respect to λ too, and that $p(x, y) \geq \delta\phi(y)$, $\lambda(dy)$ a.s. for any $x \in S$, with $\Phi(dy) = \phi(y)dy$. 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 $Y^{(n)} = (Y_1, \dots, Y_n)$ conditionally to $X^{(n+1)} = (x_1, \dots, x_{n+1})$ is the tensor product of Bernoulli distributions given by: for all $\beta^{(n)} = (\beta_1, \dots, \beta_n) \in \{0, 1\}^n$, $x^{(n+1)} = (x_1, \dots, x_{n+1}) \in E^{n+1}$,

$$P_\nu(Y^{(n)} = \beta^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = \prod_{i=1}^n P_\nu(Y_i = \beta_i \mid X_i = x_i, X_{i+1} = x_{i+1}),$$

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

- if $x_i \notin S$,

$$P_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) = \delta,$$

- if $x_i \in S$,

$$\begin{aligned} P_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) &= \delta\phi(x_{i+1})/p(x_i, x_{i+1}), \\ P_\nu(Y_i = 0 \mid X_i = x_i, X_{i+1} = x_{i+1}) &= 1 - \delta\phi(x_{i+1})/p(x_i, x_{i+1}). \end{aligned}$$

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 by $\mathcal{L}^{(n)}(p, S, \delta, \phi, x^{(n+1)})$ this probability distribution. If we were able to generate Y_1, \dots, Y_n , so that $X^{\mathcal{M}(n)} = ((X_1, Y_1), \dots, (X_n, Y_n))$ be a realization of the split chain $X^{\mathcal{M}}$ described in § 3.1, then we could apply the RBB procedure to the sample path $X^{\mathcal{M}(n)}$. Unfortunately, as shown above, knowledge of the transition density $p(x, y)$ is required to draw practically Y_1, \dots, Y_n this way. Hence, 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, \dots, X_{n+1} (note that we may choose the estimate $p_n(x, y)$ of the transition density such

that $p_n(x, y) \geq \delta\phi(y)$, $\lambda(dy)$ a.s., and $p_n(X_i, X_{i+1}) > 0$, $1 \leq i \leq n$), and then drawing a binary random vector $(\widehat{Y}_1, \dots, \widehat{Y}_n)$ conditionally to $X^{(n+1)} = (X_1, \dots, X_{n+1})$, 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, \dots, Y_n) for given $X^{(n+1)}$. Our method for bootstrapping Harris chains, which we call *approximate regenerative block-bootstrap* (ARBB), simply amounts then to apply the RBB procedure to the data $((X_1, \widehat{Y}_1), \dots, (X_n, \widehat{Y}_n))$ as if they were drawn from the atomic chain $X^{\mathcal{M}}$. We shall prove that, even if it requires to use a consistent estimate of the "nuisance parameter" p and the corresponding approximate blocks it induces, this bootstrap method still remains asymptotically valid.

In the next subsection, we show that the accuracy of this approximation in the sense of the Mallows distance (which metric is a crucial tool for proving asymptotic validity of bootstrap methods, see Bickel & Freedman (1981) and technical proofs in section 5) mainly depends on the rate of the uniform convergence of $p_n(x, y)$ to $p(x, y)$ over $S \times S$.

3.3 Mallows distance between $(X_i, \widehat{Y}_i)_{1 \leq i \leq n}$ and $(X_i, Y_i)_{1 \leq i \leq n}$

Let us show that the distribution of $(X_i, \widehat{Y}_i)_{1 \leq i \leq n}$ gets closer and closer to the distribution of $(X_i, Y_i)_{1 \leq i \leq n}$ in the sense of the Mallows distance (also known as the Kantorovich or Wasserstein metric in the probability literature) as $n \rightarrow \infty$. Hence, we express here the distance between the distributions P^Z and $P^{Z'}$ of two random sequences $Z = (Z_n)_{n \geq 0}$ and $Z' = (Z'_n)_{n \geq 0}$, taking their values in \mathbb{R}^k , by (see Rachev & Rüschemdorf (1998), p 76)

$$l_p(Z, Z') = l_p(P^Z, P^{Z'}) = \min \left\{ L_p(W, W'); W \sim P^Z, W' \sim P^{Z'} \right\},$$

with

$$(L_p(W, W'))^{1/q} = E[D^p(W, W')],$$

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 $Z^{(n)} = (X_k, Y_k)_{1 \leq k \leq n}$ and $\widehat{Z}^{(n)} = (X_k, \widehat{Y}_k)_{1 \leq k \leq n}$ as the beginning segments of infinite series, we evaluate the deviation between the distribution $P^{(n)}$ of $Z^{(n)}$ and the distribution $\widehat{P}^{(n)}$ of $\widehat{Z}^{(n)}$ by using this definition

$$l_1(P^{(n)}, \widehat{P}^{(n)}) = \min_{\substack{Z^{(n)} \sim P^{(n)} \\ \widehat{Z}^{(n)} \sim \widehat{P}^{(n)}}} \sum_{k=1}^n 2^{-k} E(|Y_k - \widehat{Y}_k|).$$

Theorem 3.1 *Assume that*

- (i) S is chosen so that $\inf_{x \in S} \phi(x) > 0$,
- (ii) p is estimated by p_n at the rate α_n for the MSE when error is measured by the L^∞ loss over S^2 , then

$$l_1(P^{(n)}, \widehat{P}^{(n)}) \leq (\delta \inf_{x \in S} \phi(x))^{-1} \alpha_n^{1/2}.$$

This results 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. This gives us some hints on how to choose the small set with a *data driven method* to obtain better finite sample results (see section 3.6).

3.4 The ARBB algorithm.

It is now easy to see how we can perform an *approximate regenerative block-bootstrap* (ARBB) algorithm to obtain the sample distribution of some statistic T_n . Given the parameter b_n , it is simply performed in six steps as follows.

1. From the data $X^{(n+1)} = (X_1, \dots, X_{n+1})$, compute an estimate $p_n(x, y)$ of the transition density such that $p_n(x, y) \geq \delta \phi(y)$, $\lambda(dy)$ a.s., and $p_n(X_i, X_{i+1}) > 0$, $1 \leq i \leq n$.
2. Conditionally to $X^{(n+1)}$, draw a binary data vector $(\widehat{Y}_1, \dots, \widehat{Y}_n)$ 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 \widehat{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 thus \widehat{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 $\widehat{l}_n = \sum_{i=1}^n I\{X_i \in S, \widehat{Y}_i = 1\}$ to the set $A_{\mathcal{M}} = S \times \{1\}$ up to time n . And divide the observed sample path $X^{(n)}$ into $\widehat{l}_n + 1$ blocks, valued in the torus $T = \cup_{n=1}^\infty E^n$, corresponding to the pieces of the sample path between consecutive visits to $A_{\mathcal{M}}$,

$$\begin{aligned} \widehat{\mathcal{B}}_0 &= (X_1, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(1)}), \widehat{\mathcal{B}}_1 = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(1)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(2)}), \dots, \\ \widehat{\mathcal{B}}_{\widehat{l}_n-1} &= (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n-1)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)}), \widehat{\mathcal{B}}_{\widehat{l}_n}^{(n)} = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)+1}, \dots, X_n), \end{aligned}$$

with

$$\begin{aligned}\widehat{\tau}_{A,\mathcal{M}}(1) &= \inf \left\{ n \geq 1, X_n \in S, \widehat{Y}_n = 1 \right\}, \\ \widehat{\tau}_{A,\mathcal{M}}(j+1) &= \inf \left\{ n > \widehat{\tau}_{A,\mathcal{M}}(j), X_n \in S, \widehat{Y}_n = 1 \right\}, \\ l(\widehat{\mathcal{B}}_j) &= \widehat{\tau}_{A,\mathcal{M}}(j+1) - \widehat{\tau}_{A,\mathcal{M}}(j), \text{ for } j \geq 1.\end{aligned}$$

4. Draw sequentially bootstrap data blocks $\mathcal{B}_1^*, \dots, \mathcal{B}_k^*$ independently from the empirical distribution $F_n = (\widehat{l}_n - 1)^{-1} \sum_{j=1}^{\widehat{l}_n - 1} \delta_{\widehat{\mathcal{B}}_j}$ of the blocks $\widehat{\mathcal{B}}_1, \dots, \widehat{\mathcal{B}}_{\widehat{l}_n - 1}$, conditioned on $X^{(n)}$ until the length of the bootstrap data series $l^*(k) = \sum_{j=1}^k l(\mathcal{B}_j^*)$ is larger than b_n . Let $l_n^* = \inf\{k \geq 1, l^*(k) > b_n\}$.
5. From the bootstrap data blocks generated at step 4, reconstruct a pseudo-trajectory by binding the blocks together, getting the reconstructed *ARBB sample path*

$$X_{b_n}^{*(n)} = (\mathcal{B}_1^*, \dots, \mathcal{B}_{l_n^* - 1}^*).$$

Then compute the *ARBB statistic* and the *ARBB standardization*

$$T_{n,b_n}^{*(n)} = T(X_{b_n}^{*(n)}), \quad S_{n,b_n}^{*(n)} = S(X_{b_n}^{*(n)}).$$

6. The *ARBB distribution* is then given by

$$H_{ARBB}(x) = P^*(S_{n,b_n}^{*-1}(T_{n,b_n}^* - T_n) \leq x \mid X^{(n+1)}).$$

In the following we choose $b_n = n$.

Remark 3.1 A Monte-Carlo approximation to $H_{ARBB}(x)$ may be straightforwardly computed by repeating independently N times the steps 4-6 of the procedure above.

3.5 Asymptotic validity of the ARBB

As explained in Malinovskii (1987) in the unstandardized case, the Edgeworth expansion (E.E. in abbreviated form) proved in Bertail & Cléménçon (2004) for an atomic chain in the studentized case straightforwardly extends to a general positive recurrent chain 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 E.E. are independent from these latter, in particular the asymptotic variance, which may be written

$\sigma_f^2 = E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} E_{A_{\mathcal{M}}}((\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} \{f(X_i) - \mu(f)\})^2)$, where $\tau_{A_{\mathcal{M}}} = \inf\{n \geq 1, (X_n, Y_n) \in S \times \{1\}\}$ and $E_{A_{\mathcal{M}}}(\cdot)$ denotes the expectation conditionally to $(X_0, Y_0) \in A_{\mathcal{M}} = S \times \{1\}$. However in the studentized case, one cannot use the standardization defined in 2.4.1 in the atomic setting for the split chain, since the times when the split chain regenerates 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 3.4,

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

of $\mu(f)$ and σ_f^2 respectively, with $\hat{n}_{A_{\mathcal{M}}} = \hat{\tau}_{A_{\mathcal{M}}}(\hat{l}_n) - \hat{\tau}_{A_{\mathcal{M}}}(1) = \sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)$ and $f(\hat{\mathcal{B}}_j) = \sum_{i=1+\hat{\tau}_{A_{\mathcal{M}}}(j)}^{\hat{\tau}_{A_{\mathcal{M}}}(j+1)} f(X_i)$. By convention, $\hat{\mu}_n(f)$ (respectively $\hat{\sigma}_n^2(f)$, $\hat{n}_{A_{\mathcal{M}}}$) equals to 0, when $\hat{l}_n \leq 1$. Note that, analogously to the way we proceeded in the atomic case to avoid bias terms that cannot be approximated by using a resampling method (see subsection 2.2), eventual data collected before the first (respectively, after the last) pseudo-regeneration time are not used to construct these estimators.

Let us define the unstandardized distribution by

$$H_{\nu}^U(x) = P_{\nu}(\hat{n}_{A_{\mathcal{M}}}^{1/2} \sigma(f)^{-1} (\hat{\mu}_n(f) - \mu(f)) \leq x).$$

We also define the *pseudo-regeneration based studentized sample mean*

$$\hat{t}_{A_{\mathcal{M}},n} = \frac{\sum_{j=1}^{\hat{l}_n-1} f(\hat{\mathcal{B}}_j) - \mu(f) \sum_{j=1}^{\hat{l}_n-1} l(\hat{\mathcal{B}}_j)}{\left(\sum_{j=1}^{\hat{l}_n-1} \left\{ f(\hat{\mathcal{B}}_j) - \hat{\mu}_n(f) l(\hat{\mathcal{B}}_j) \right\}^2 \right)^{1/2}} = \hat{n}_{A_{\mathcal{M}}}^{1/2} \frac{\hat{\mu}_n(f) - \mu(f)}{\hat{\sigma}_n(f)},$$

with sampling distribution $H_{\nu}^S(x) = P_{\nu}(\hat{t}_{A_{\mathcal{M}},n} \leq x)$.

3.5.1 Further assumptions and preliminary results

In this setting we use the following assumptions to establish the asymptotic validity of the ARBB approach. Let $k \geq 2$ be a real number.

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

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

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

$$\sup_{x \in S} E_x(\tau_S^k) < \infty \text{ and } E_{\nu}(\tau_S^k) < \infty.$$

Remark 3.2 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 checked for some accessible small set S , it is also fulfilled for all accessible small sets of the chain). Note also that in the case when assumption $\mathcal{H}_1(k, \nu)$ is satisfied, $\mathcal{H}_1(f, k, \nu)$ is fulfilled for any bounded measurable function f .

For a sequence of nonnegative real numbers $(\alpha_n)_{n \in \mathbf{N}}$ converging to 0 as $n \rightarrow \infty$,

\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 $S \times S$:

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

Remark 3.3 Numerous estimators of the transition density of positive recurrent Markov chains have been proposed in the literature and their estimation rates have been established under various smoothness assumptions on the density of the joint distribution $\mu(dx)\Pi(x, dy)$ and the one of $\mu(dx)$ (refer to Roussas (1969), Rosenblatt (1970), Birgé (1983), Doukhan & Ghindès (1983), Athreya & Atunçar (1998) or Cléménçon (2000) for instance). For instance, under classical Hölder constraints of order s , the typical rate for the risk in this setup is $\alpha_n \sim (\ln n/n)^{s/(s+1)}$.

\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 have the following result.

Theorem 3.2 *Suppose that conditions of Theorem 3.1 are fulfilled by the chain. Assume moreover that X satisfies $\mathcal{H}_1(f, \rho, \nu)$ and $\mathcal{H}_2(\rho, \nu)$ with $\rho \geq 4$, \mathcal{H}_3 , \mathcal{H}_4 and \mathcal{H}_5 , then we have as $n \rightarrow \infty$*

$$\begin{aligned} \hat{\sigma}_n^2(f) &\rightarrow \sigma_f^2 \text{ in } P_\nu\text{-probability,} \\ \hat{n}_{A, \mathcal{M}}^{1/2} \frac{\hat{\mu}_n(f) - \mu(f)}{\hat{\sigma}_n(f)} &\rightarrow \mathcal{N}(0, 1) \text{ in distribution under } P_\nu. \end{aligned}$$

We recall that condition $\mathcal{H}_1(f, \rho, \nu)$ may be more easily checked in practice by using test functions methods (cf Kalashnikov (1978)). In particular, it is well known that "block" moment conditions may be replaced by drift criteria of Lyapounov's type (see Chapter 11 in Meyn & 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

The bootstrap counterparts of the statistics introduced above are then defined as follows. Let $\mathcal{B}_1^*, \dots, \mathcal{B}_{l_n^*}^*$ be the bootstrapped pseudo-regenerative data 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

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

The unstandardized version of the approximate-regenerative bootstrap distribution is given by

$$H_{ARBB}^U(x) = P^*(n_{A\mathcal{M}}^{*1/2} \hat{\sigma}_n^{-1}(f) (\mu_n^*(f) - \hat{\mu}_n(f)) \leq x \mid X^{(n+1)}).$$

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

$$t_{A\mathcal{M},n}^* = \frac{\sum_{j=1}^{l_n^*-1} f(\mathcal{B}_j^*) - \hat{\mu}_n(f) \sum_{j=1}^{l_n^*-1} l(\mathcal{B}_j^*)}{\left(\sum_{j=1}^{l_n^*-1} \{f(\mathcal{B}_j^*) - \mu_n^*(f)l(\mathcal{B}_j^*)\}^2\right)^{1/2}} = n_{A\mathcal{M}}^{*1/2} \frac{\mu_n^*(f) - \hat{\mu}_n(f)}{\sigma_n^*(f)}$$

and the studentized ARBB distribution estimate $H_{ARBB}^S(x) = P^*(t_{A\mathcal{M},n}^* \leq x \mid X^{(n+1)})$. Note that this is the same construction as in the atomic case, except that one uses the approximated blocks instead of the true regenerative ones.

Theorem 3.3 *Under the hypotheses of Theorem 3.2, we have the following convergences in probability under P_ν*

$$\begin{aligned} \Delta_n^U &= \sup_{x \in \mathbb{R}} |H_{ARBB}^U(x) - H_\nu^U(x)| \rightarrow 0, \text{ as } n \rightarrow \infty \\ \Delta_n^S &= \sup_{x \in \mathbb{R}} |H_{ARBB}^S(x) - H_\nu^S(x)| \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

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. Although such a result is clearly not satisfactory regarding practical applications, we restricted ourselves here to this simplistic setting with the aim to show that the ARBB may yield second order correctness at a better rate than the MBB does, by

using simple arguments only. In the stationary case, one may actually use $\mu_n(f) = n^{-1} \sum_{i=1}^n X_i$ as a natural unbiased estimate of $\mu(f)$, avoiding this way controlling the behaviour of the first and last pseudo-regenerative blocks, a lengthy task that could be accomplished at the price of a much more technical study). In the following we thus deal with

$$H_\mu^U(x) = P_\mu(n^{1/2}\sigma_f^{-1}\{\bar{\mu}_n(f) - \mu(f)\} \leq x).$$

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

$$H_{ARBB}^{U,s}(x) = P^*(t_{A_{\mathcal{M}},n}^* - E^*(t_{A_{\mathcal{M}},n}^*) \leq x \mid X^{(n+1)}).$$

We then have the following result.

Theorem 3.4 *Assume that the chain is stationary. Under the hypotheses of Theorem 3.2 with $\rho > 6$, and if in addition the Cramer condition $\overline{\lim}_{|t| \rightarrow \infty} \sup_{x \in S} |E_x(\exp(it(\sum_{i=1}^{\tau_S} \{f(X_i) - \mu(f)\}))| < 1$ is fulfilled, then we have*

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

Remark 3.4 Observe that, even if one uses a nonoptimal estimator p_n , the ARBB is still second order correct. Of course the better the estimator is, the sharper the approximation is. In the geometrically recurrent case, the best rate that can be attained is typically $\alpha_n = n^{-1} \log(n)$ (see Clémençon (2000) for instance), yielding the validity of the ARBB up to the almost optimal rate $o_{P_\nu}(n^{-1} \log(n)^{1/2})$, which clearly improves over the MBB even in the unstudentized case (see Lahiri (2003), chap. 6.5).

Remark 3.5 The second order properties of the ARBB in the studentized general case cannot be straightforwardly deduced from the E.E. of the RBB version by the standard Chibisov lemma, nor from the argument used in the atomic case. As a careful examination of the proof of Theorem 5.1 in Bertail & Clémençon (2004) shows, second order asymptotic results for the RBB when a known atom A exists are established by partitioning the probability space according to the number $l_n - 1$ of regenerative blocks and the values taken by the successive regeneration times $\tau_A(1), \dots, \tau_A(l_n)$ up to time n , and applying then non uniform limit theorems for sample mean statistics based on 1-lattice i.i.d. random vectors on each subset of the partition (see also Malinovskii (1987, 89)). What makes this approach very hard to transpose in the ARBB case is that, by construction, pseudo-regeneration times $\widehat{\tau}_{A_{\mathcal{M}}}(j)$, and the data blocks $\widehat{\mathcal{B}}_j$ they induce, depend on the

whole trajectory, owing to the transition probability estimation step. In spite of this, we nevertheless expect that the ARBB properties are comparable to the RBB properties.

3.6 Practical choice of the small set

One may easily check at hand on many examples of real valued Markov chains X , that any compact interval $V_{x_0}(\varepsilon) = [x_0 - \varepsilon, x_0 + \varepsilon]$ for a well chosen x_0 (typically its mean) and $\varepsilon > 0$ small enough, is a small set with the uniform distribution on $V_{x_0}(\varepsilon)$, denoted by $\mathcal{U}_{V_{x_0}(\varepsilon)}$ as a minorization measure (see example 2 in § 4.2). Hence, in the case when one knows x_0 , ε and δ such that (2) holds (this simply amounts to knowing a uniform lower bound estimate for the probability to return to $V_{x_0}(\varepsilon)$ in one step), one may effectively apply the ARBB methodology to X .

The number of pseudo-regenerative blocks to resample 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 is the lower bound (2) (the larger δ is, the larger is the probability to draw pseudo regenerative times $\widehat{Y}_k = 1$ at randomization steps, *i.e.* when $X_k \in V_{x_0}(\varepsilon)$). And since the larger ε is, the smaller $\delta(\varepsilon)$ is, it is intuitive to think that better numerical results for the block-resampling procedure can be obtained in practice for some specific choices of the size ε , likely (but with no argument except empirical evidence to support this point) for choices corresponding to a maximum number of regenerative data blocks given the trajectory. Thus, from a practical viewpoint, when no prior information about the structure of the chain is available, a data driven method for choosing empirically the minorization condition parameters could be as follows.

A possible ideal selection rule could rely on searching for $\varepsilon > 0$ so as to maximize the expected number of data-blocks conditionally to the observed trajectory, that is

$$\begin{aligned} N_n(\varepsilon) &= E\left(\sum_{i=1}^n I\{X_i \in V_{x_0}(\varepsilon), Y_i = 1\} \mid X^{(n+1)}\right) \\ &= \frac{\delta(\varepsilon)}{2\varepsilon} \sum_{i=1}^n I\{(X_i, X_{i+1}) \in V_{x_0}(\varepsilon)^2\} \frac{1}{p(X_i, X_{i+1})}. \end{aligned} \quad (4)$$

Since the transition density p and its minimum over $V_{x_0}(\varepsilon)^2$ are unknown, a practical empirical criterion $\widehat{N}_n(\varepsilon)$ to optimize is obtained by replacing p by an estimate p_n and $\delta(\varepsilon)/2\varepsilon$ by a lower bound $\widehat{\delta}_n(\varepsilon)/2\varepsilon$ for p_n over $V_{x_0}(\varepsilon)^2$ in expression (4). Observe finally that other approaches may be used for the choice of the minorization condition, for instance one may refer to Roberts & Rosenthal (1996) in the case of diffusion Markov processes.

4 Illustrative examples

We now give two examples, only with a view to illustrate the scope of applications of our methodology. The first example presents a regenerative Markov chain described and studied at greater length in Harrison & Resnick (1976) (see also Brockwell, Resnick & Tweedie (1982) and Browne & Sigman (1992)) for modeling storage systems. In consideration of the recent emphasis on nonlinear models in the time series literature, our second example shows to what extent the ARBB method may apply to a general nonlinear AR model (see Franke *et al.* (2002)). Further, we point out that the principles exposed in this paper are by no means restricted to the markovian setting, but may apply to any process for which a regenerative extension can be constructed and simulated from the data available (see chapter 10 in Thorisson (2000)).

4.1 Example 1 : content-dependent storage systems

We consider a general model for storage, evolving through a sequence of *input times* $(T_n)_{n \in \mathbb{N}}$ (with $T_0 = 0$ by convention), at which the storage system is replenished. Let S_n be the amount of input into the storage system at the n^{th} input time T_n and C_t be the amount of contents of the storage system at time t . When possible, there is withdrawal from the storage system between these input times at the constant rate r and the amount of stored contents that drops in a time period $[T, T + \Delta T]$ since the latter input time is equal to $C_T - C_{T+\Delta T} = r\Delta T$, and when the amount of contents reaches zero, it continues to take the value zero until it is replenished at the next input time. If X_n denotes the amount of contents immediately before the input time T_n (*i.e.* $X_n = C_{T_n} - S_n$), we have for all $n \in \mathbb{N}$,

$$X_{n+1} = (X_n + S_n - r\Delta T_{n+1})_+,$$

with $(x)_+ = \sup(x, 0)$, $X_0 = 0$ by convention and $\Delta T_n = T_n - T_{n-1}$ for all $n \geq 1$. Let $K(x, ds)$ be a transition probability kernel on \mathbb{R}_+ . Assume that, conditionally to X_1, \dots, X_n , the amounts of input S_1, \dots, S_n are independent from each other and independent from the inter-arrival times $\Delta T_1, \dots, \Delta T_n$ and that the distribution of S_i is given by $K(X_i, \cdot)$, for $0 \leq i \leq n$. Under the further assumption that $(\Delta T_n)_{n \geq 1}$ is an i.i.d. sequence with common distribution G , independent from $X = (X_n)_{n \in \mathbb{N}}$, the storage process X is a Markov chain with transition probability kernel Π given by

$$\begin{aligned} \Pi(X_n, \{0\}) &= \Gamma(X_n, [X_n, \infty[), \\ \Pi(X_n,]x, \infty[) &= \Gamma(X_n,]-\infty, X_n - x]) \end{aligned}$$

for all $x > 0$, where the transition probability Γ is given by the convolution product $\Gamma(x,]-\infty, y]) = \int_{t=0}^{\infty} \int_{z=0}^{\infty} G(dt)K(x, dz)I\{rt - z < y\}$.

One may check that the chain Π is δ_0 -irreducible as soon as $K(x, \cdot)$ has infinite tail for all $x \geq 0$. In this case, $\{0\}$ is an accessible atom for X and it can be shown that it is positive recurrent if and only if there exists $b > 0$ and a test function $V : \mathbb{R}_+ \rightarrow [0, \infty]$ such that $V(0) < \infty$ and for all $x \geq 0$:

$$\int \Pi(x, dy)V(y) - V(x) \leq -1 + bI\{x = 0\}.$$

The times at which the storage process X reaches the value 0 are thus regeneration times, and allow to define regeneration blocks dividing the sample path, as shown in Fig. 1. Figure 2 below shows a reconstructed RBB data series, generated by a sequential sampling of the regeneration blocks (as described in 2.3), on which RBB statistics may be based.

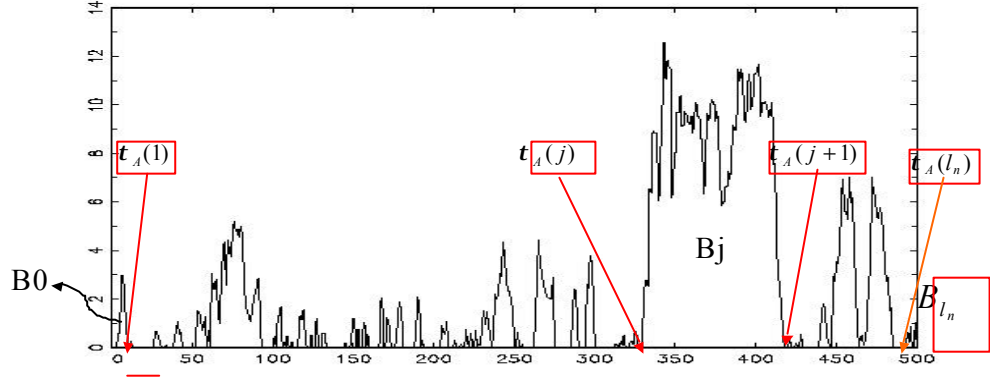


Fig. 1 : Dividing the trajectory of the storage process into data blocks corresponding to the regeneration times $\tau_A(j)$

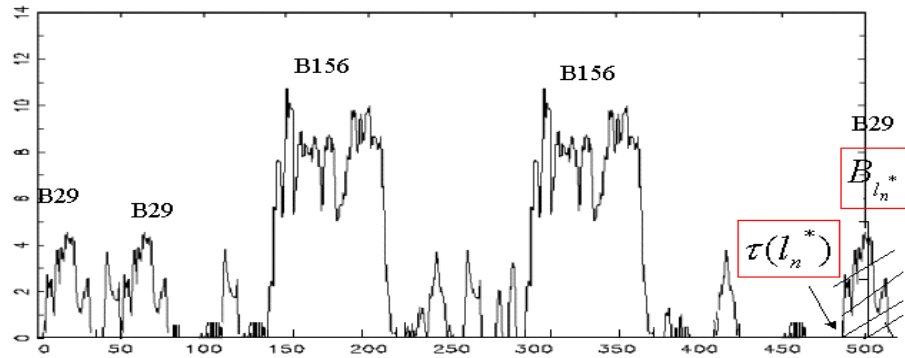


Fig. 2 : Reconstruction of a storage process data series using the RBB resampling procedure

Simulation results We simulated two trajectories of respective length $n = 100$ and $n = 200$ drawn from this Markov chain with $r = 1$, $K(x, dy) = Exp_3(dy)$ and $G(dy) = Exp_1(dy)$, denoting by $Exp_\lambda(dy)$ the exponential distribution with mean $1/\lambda > 0$, which is a standard M/M/1 model (see Asmussen (1987) for instance). In Fig. 3 below, a Monte-Carlo estimate of the true distribution of the sample mean standardized by its estimated standard error (as defined in § 3.2.1) computed with 10000 simulated trajectories is compared to the RBB distribution (in both cases, Monte-Carlo approximations of RBB estimates are computed from $B = 2000$ repetitions of the RBB procedure, see remark 3.1)) and to the gaussian approximation.

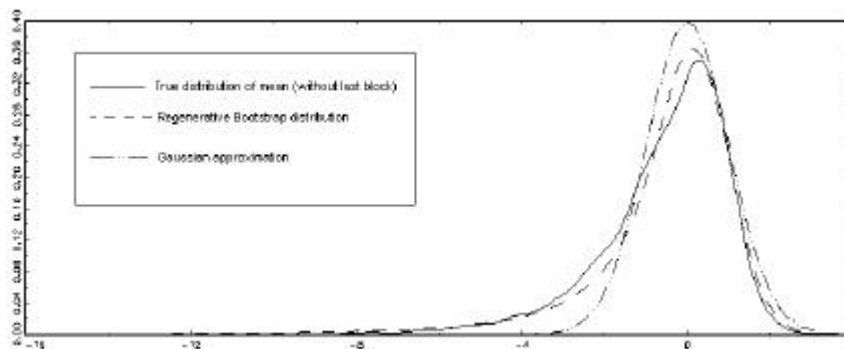


Fig. 3 : Comparison of true, RBB and gaussian distributions for $n = 200$.

With a view to constructing accurate confidence intervals, Table 1 compares the quantile of order γ of the true distribution, the one of the gaussian approximation (both estimated with 10000 simulated trajectories) and the mean of the quantile of order γ of the RBB distribution over 100 repetitions of the RBB procedure in the tail regions.

The left tail is clearly very well estimated, whereas the right tail gives a better approximation than the asymptotic distribution. The gain in term of coverage accuracy is quite enormous in comparison to the asymptotic distribution. For instance at the level 95%, for $n = 200$, the asymptotic distribution yields a bilateral coverage interval of level 71% whereas the RBB distribution yields in our simulation a level of 92%.

n=	100		200		∞	n=	100		200		∞
$\gamma\%$	TD	RBB	TD	RBB	ASY	$\gamma\%$	TD	RBB	TD	RBB	ASY
1	-7.733	-7.044	-5.492	-5.588	-2.326	90	1.041	1.032	1.029	1.047	1.282
5	-4.374	-4.134	-3.430	-3.477	-1.645	95	1.287	1.360	1.299	1.356	1.645
10	-3.184	-2.950	-2.472	-2.560	-1.282	99	1.762	1.970	1.839	1.916	2.326

Table 1 : Comparison of the quantiles of the true distribution (TD), RBB and gaussian distributions.

4.2 Example 2 : nonlinear AR models

Consider the general heteroskedastic autoregressive model

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

where $m : \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R} \rightarrow \mathbb{R}_+^*$ are measurable functions, $(\varepsilon_n)_{n \in \mathbb{N}}$ is a i.i.d. sequence of r.v.'s drawn from $g(x)dx$ such that, for all $n \in \mathbb{N}$, ε_{n+1} is independent from the X_k 's, $k \leq n$ with $E(\varepsilon_{n+1}) = 0$ and $var(\varepsilon_{n+1}) = 1$. The transition kernel density of the chain is given by $p(x, y) = g((y - m(x))/\sigma(x))$, $(x, y) \in \mathbb{R}^2$. 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 neighborhood $V_{x_0}(\varepsilon)^2 = [x_0 - \varepsilon, x_0 + \varepsilon]^2$ of (x_0, x_0) in \mathbb{R}^2 : there exists $\delta = \delta(\varepsilon) \in]0, 1[$ such that,

$$\inf_{(x,y) \in V_{x_0}^2} p(x, y) \geq \delta(2\varepsilon)^{-1}. \quad (5)$$

Thus the chain X satisfies the minorization condition $\mathcal{M}(1, V_{x_0}(\varepsilon), \delta, \mathcal{U}_{V_{x_0}(\varepsilon)})$. Here are empirical evidences for two specific models.

The $AR(1)$ model : $X_{i+1} = \alpha X_i + \varepsilon_{i+1}$, $i \in \mathbb{N}$, with $\varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\alpha = 0.95$, $X_0 = 0$ and for a trajectory of length $n = 200$.

The $AR(1)$ model with $ARCH(1)$ residuals called AR-ARCH model:

$$X_{i+1} = \alpha X_i + (1 + \beta X_i^2)^{1/2} \varepsilon_{i+1}, \quad i \in \mathbb{N},$$

with $\varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\alpha = 0.6$, $\beta = 0.1$, $X_0 = 0$ and for a trajectory of length $n = 200$.

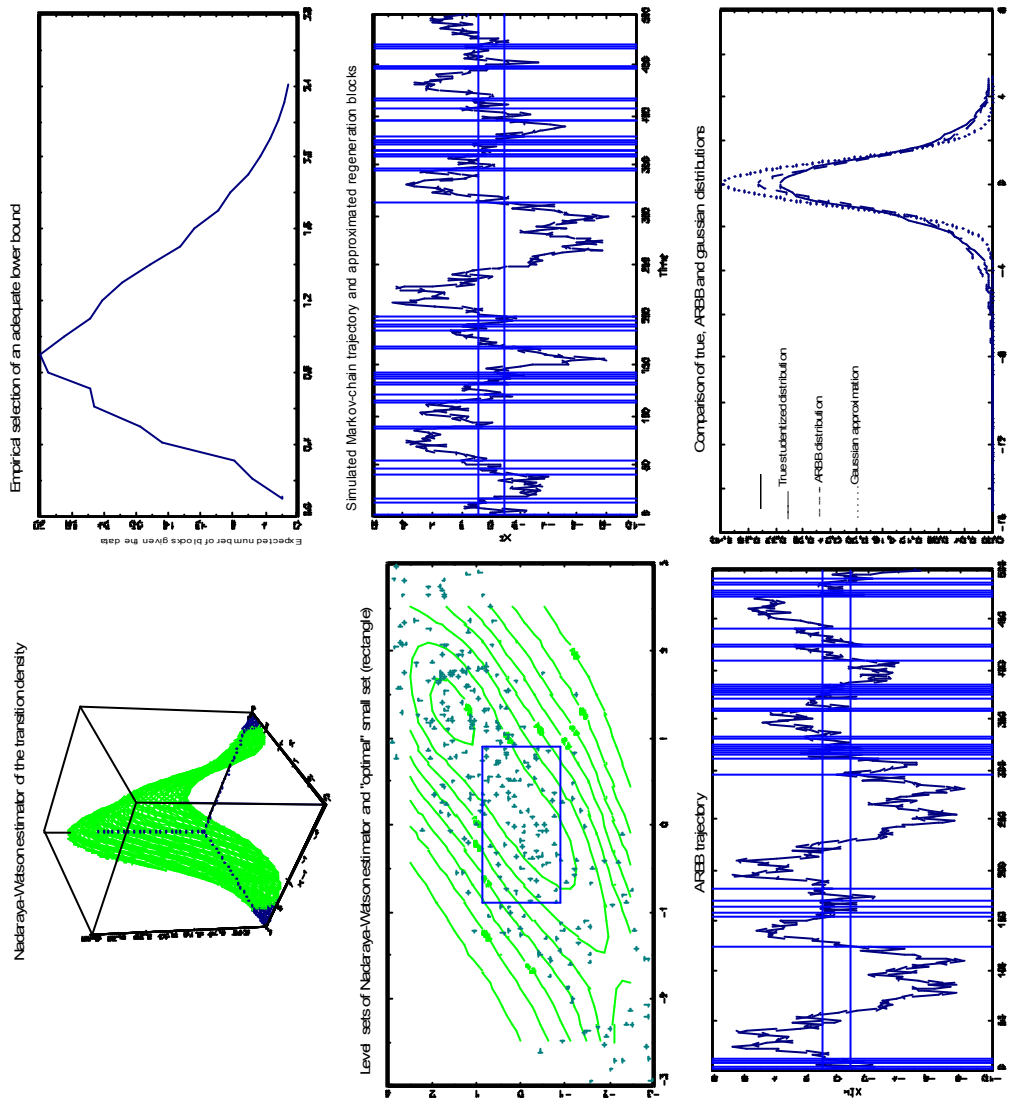
A much more detailed simulation study as well as applications involving different types of data generating markovian model will be carried out in a forthcoming paper. Here the true distribution of the sample mean is estimated with 10000 simulations, the ARBB distribution is approximated with $B = 2000$ iterations and the procedure is repeated 100 times (we simulated 100 trajectories and averaged the 100 quantiles of the corresponding ARBB distributions in Table 2). We estimated the transition density with a simple kernel estimator \hat{p}_n of Nadaraya-Watson's type (see Doukhan & Ghindès (1983) for instance). Note it is a nonparametric estimator and its computation requires no prior information about the underlying markovian model. The small set is selected by maximizing the empirical criterion $\hat{N}_n(\varepsilon)$ described above over $\varepsilon > 0$. The main steps of the procedure are summarized in graph panels 1 and 2. The first figure shows the Nadaraya-Watson (NW) estimator, the second one represents $\hat{N}_n(\varepsilon)$ as ε grows and clearly allows to identify an optimal value for the size of the small set. In the case of the AR model for instance, this selection rule leads to pick $\varepsilon = 0.90$ and $\hat{\delta} = 0.110$ (the minimum value of $p(x, y)$ over the corresponding square is

actually $\delta = 0.118$). In the second line of the panel, the level sets of the NW estimator, the data points (X_i, X_{i+1}) and the "optimal" small set are represented (this also shows that the small set chosen may be not that "small" if the transition density is flat around $(0, 0)$). The next figure shows a trajectory of the chain and indicates the pseudo-regenerative blocks obtained by applying the randomization rule with $Ber(1 - \widehat{\delta}(2\varepsilon)^{-1} / \widehat{p}_n(X_i, X_{i+1}))$ at times i when $(X_i, X_{i+1}) \in V_0(\varepsilon)^2$. The figure on the right hand shows how binded blocks form a typical ARBB trajectory. It is noteworthy that such a trajectory presents less artificial "jumps" than a trajectory reconstructed from a classical MBB procedure. Finally the true distribution, the ARBB distribution and the asymptotic gaussian distribution are compared in the last picture. Table 2 below gives the mean of some quantiles (of order γ) of the ARBB distribution over 100 replications of the procedure for the two models, compared to the true and asymptotic quantiles.

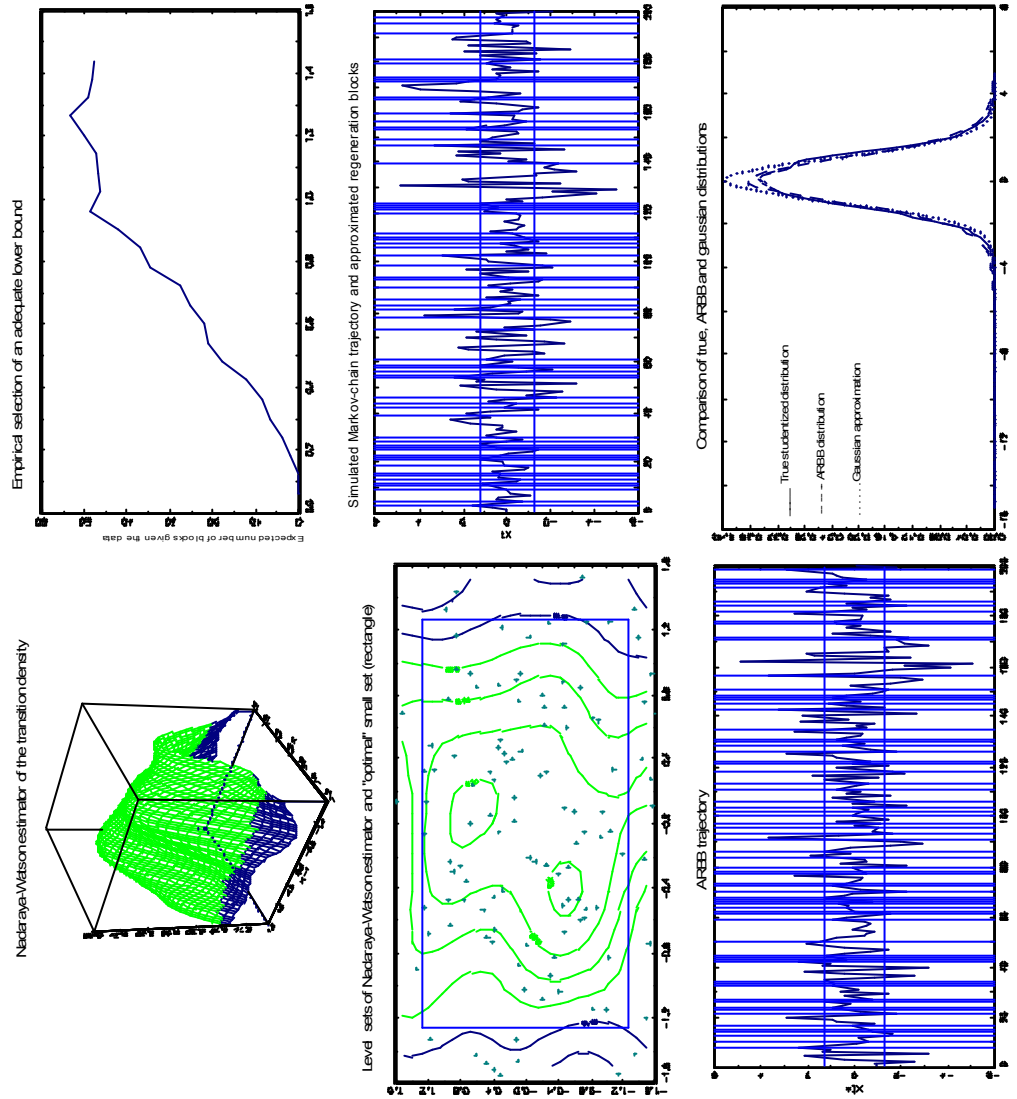
	AR		AR-ARCH				AR		AR-ARCH		
γ	TD	ARBB	TD	ARBB	ASY	γ	TD	RBB	TD	ARBB	ASY
1	-3.639	-3.754	-2.532	-2.683	-2.326	90	1.683	1.583	1.362	1.416	1.282
2.5	-2.772	-2.818	-2.025	-2.146	-1.960	95	2.160	1.934	1.732	1.826	1.645
5	-2.346	-2.388	-1.793	-1.866	-1.645	97.5	2.731	2.533	2.004	2.143	1.960
10	-1.741	-1.797	-1.429	-1.452	-1.282	99	3.627	3.572	2.533	2.693	2.326

Table 2: Comparison of the tails of the true, ARBB and gaussian distributions for the two models

These pictures speak volumes : for both models the true distribution is accurately approximated. But note in particular the difference in the size of the "optimal" small set and in the number of pseudo-regenerations between these models. We point out that, though remarkable when compared to the gaussian approximation, the gain in accuracy obtained by applying the ARBB methodology to the AR model is higher than the one obtained for the AR-ARCH type model. As may be confirmed by other simulations, the ARBB method provides less accurate results for a given (moderate) sample size, as one gets closer to a unit root model (*i.e.* as α tends to 1): one may simply get an insight for this phenomenon by noticing that the rate of the number of regenerations (respectively, of the number of visits to the small set) then decreases. Although generalizing successfully the ARBB method to the null recurrent case seems possible from a theoretical viewpoint, these empirical results suggest that further investigation is needed to elaborate a practical ARBB procedure with so good properties in this case. This goes beyond the scope of this article but will be studied elsewhere.



Graph panel 1 : AR-model, $\alpha = 0.95$, $n = 200$, $X_0 = 0$.



Graph panel 2 : AR(1) with ARCH(1) residual model with $\alpha = 0.2$, $\beta = 0.7$, $n = 200$, $X_0 = 0$

5 Proofs

5.1 Proof of Theorem 2.1.

In the following, C and K are constants which may be not necessarily the same at each appearance. We denote by $E^*(\cdot \mid X^{(n)})$ the conditional expectation given $X^{(n)}$ (recall that the $f(\mathcal{B}_j^*)$'s are i.i.d with distribution

$l_n^{-1} \sum_{j=1}^{l_n} \delta_{\mathcal{B}_j}$, conditionally to the trajectory $X^{(n)}$). Observe first that it suffices to consider the case of a non trivial regeneration set only, that is $E_A(\tau_A) > 1$. Indeed if $E_A(\tau_A) = 1$, then $\tau_A = 1$ a.s. and the X_i 's 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 $O(n^{-1})$ (cf Hall (1986)).

The proof now relies on checking that conditions for the validity of the Edgeworth expansions (E.E. in abbreviated form) established in Malinovskii (1987) (cf Theorem 1 therein) and in Bertail & Cl  men  on (2004) (cf Theorem 5.1 therein) respectively are fulfilled for the RBB reconstructed series. Because we only consider the statistics without the first and last block, the E.E. holds up to $O(n^{-1})$. These asymptotic results actually hold not only for Markov chains but also for any regenerative process, with regeneration times defining blocks satisfying the hypotheses required. It is easy to see that exactly the same argument may be used to derive the E.E. of the bootstrap counterpart (just like in the i.i.d. case), with the only difference that all the quantities are replaced by their empirical counterparts.

- Condition (i). Denote for $t > 0, x \in \mathbb{R}$,

$$C_n(t, x) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} \exp(it\{f(\mathcal{B}_j) - xl(\mathcal{B}_j)\}),$$

$$C(t) = E_A(\exp(it\{f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j)\})).$$

We have

$$\lim_{n \rightarrow \infty} \overline{\lim}_{|t| \rightarrow \infty} |E^*(\exp(it\{f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)\}) | X^{(n)})| = \lim_{n \rightarrow \infty} \overline{\lim}_{|t| \rightarrow \infty} |C_n(t, \mu_n(f))|.$$

On any compact set the following bound holds uniformly in t ,

$$|C_n(t, \mu_n(f)) - C_n(t, \mu(f))| \leq t|\mu_n(f) - \mu(f)| \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j),$$

and the term at the right hand side almost surely converges to 0 as $n \rightarrow \infty$ by virtue of the SLLN ($\mu_n(f) \rightarrow \mu(f)$, while $(l_n - 1)^{-1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j) \rightarrow E_A(\tau_A)$ as $n \rightarrow \infty$). Moreover the SLLN also yields that $|C_n(\mu(f)) - C(t)| \rightarrow 0$ a.s. as $n \rightarrow \infty$. Thus $C_n(t, \mu_n(f))$ converges to $C(t)$ uniformly over any compact set. Since $\lim_{t \rightarrow \infty} C(t) < 1$, we may choose $0 < \eta < 1$, such that $|C(t)| \leq 1 - \eta$ for any t large enough. Then, for any $A > 0$, for any $|t| \leq A$, there exists N such that, for all $n > N$, $|C_n(t, \mu_n(f))| \leq |C(t)| + \eta/2 \leq 1 - \eta/2$. Since this is true for any $A > 0$, the empirical Cramer condition is satisfied.

- Conditions (ii) and (iii). By virtue of the SLLN again, we have

$$E^*(f(\mathcal{B}_j^*)^p | X^{(n)}) = \frac{1}{l_n - 1} \sum_{k=1}^{l_n-1} f(\mathcal{B}_k)^p \rightarrow E(f(\mathcal{B}_j)^p) < \infty, \text{ as } n \rightarrow \infty$$

In particular,

$$\begin{aligned} \text{var}^*(f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)) &= \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} \{f(\mathcal{B}_j) - \mu_n(f)l(\mathcal{B}_j)\}^2 \\ &= \frac{\tau_A(l_n) - \tau_A(1)}{l_n - 1} \sigma_n^2(f) \rightarrow E_A(\tau_A) \sigma_f^2 > 0 \text{ a.s.} \end{aligned}$$

as $n \rightarrow \infty$.

- Conditions (iv) and (v). We have by definition of the blocks

$$E^*(l(\mathcal{B}_j^*) \mid X^{(n)}) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j)^s \rightarrow E_A(\tau_A^s) \text{ a.s., as } n \rightarrow \infty$$

Thus for n large enough,

$$E^*(l(\mathcal{B}_j^*) \mid X^{(n)}) > 1, \text{ a.s. .}$$

It follows from Bertail & Cl  men  on (2004) that, as $n \rightarrow \infty$,

$$\begin{aligned} H_{RBB}^U(x) &= E_{n,n}^{(2)}(x) + O_{P_\nu}(n^{-1}), \\ H_{RBB}^S(x) &= F_{n,n}^{(2)}(x) + O_{P_\nu}(n^{-1}), \end{aligned}$$

with

$$\begin{aligned} E_{n,n}^{(2)}(x) &= \Phi(x) - n^{-1/2} \frac{k_3^{(n)}(f)}{6} (x^2 - 1) \phi(x) - n^{-1/2} b^{(n)}(f) \phi(x), \\ F_{n,n}^{(2)}(x) &= \Phi(x) + n^{-1/2} \frac{k_3^{(n)}(f)}{6} (2x^2 + 1) \phi(x) - n^{-1/2} b^{(n)}(f) \phi(x). \end{aligned} \quad (6)$$

Here the empirical skewness $k_3^{(n)}(f)$ and the bias $b^{(n)}(f)$ are given by

$$\begin{aligned} k_3^{(n)}(f) &= (E^*(l(\mathcal{B}_j^*) \mid X^{(n)}))^{-1} \{M_{3,A}^{(n)}(f) - 3\sigma_n^2(f)\beta^{(n)}(f)\} / \sigma_n^3(f), \\ b^{(n)}(f) &= -(E^*(l(\mathcal{B}_j^*) \mid X^{(n)}))^{-1} \beta^{(n)}(f) / \sigma_n(f), \end{aligned}$$

with

$$\begin{aligned} M_{3,A}^{(n)}(f) &= E^*(\{f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)\}^3 \mid X^{(n)}), \\ \beta^{(n)}(f) &= E^*(\{f(\mathcal{B}_j^*) - \mu_n(f)l(\mathcal{B}_j^*)\}l(\mathcal{B}_j^*) \mid X^{(n)}). \end{aligned}$$

It suffices thus to show that each of these terms converges at the rate $n^{-1/2}$ to the corresponding terms in the E.E. of $\mu_n(f)$. By developping first the cubic term and applying a CLT to each term, we clearly have as $n \rightarrow \infty$,

$$M_{3,A}^{(n)}(f) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} \{f(\mathcal{B}_j) - \mu_n(f)l(\mathcal{B}_j)\}^3 = M_{3,A}(f) + O_{P_\nu}(n^{-1/2}),$$

where

$$M_{3,A}(f) = E(\{f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j)\}^3)$$

In a similar fashion one may easily prove that, as $n \rightarrow \infty$,

$$\sigma_n^2(f) = \sigma_f^2 + O_{P_\nu}(n^{-1/2}) \text{ and } \beta^{(n)}(f) = \beta + O_{P_\nu}(n^{-1/2}),$$

with $\beta = \text{cov}(l(\mathcal{B}_j), f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j))$, provided that condition (ii) is fulfilled with $s = 6 + \varepsilon$. Note in addition that, as $n \rightarrow \infty$,

$$E^*(l(\mathcal{B}_j^*) | X^{(n)}) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j) = E_A(\tau_A) + O_{P_\nu}(n^{-1/2}).$$

The proof is then finished by observing that the E.E. of the true distributions (see expressions (8) and (9) in Bertail & Cléménçon (2004) and the one of the RBB distribution match up to $O_{P_\nu}(n^{-1})$ in the unstandardised case and in the standardized case as well, as $n \rightarrow \infty$.

5.2 Proof of Theorem 3.1

In what follows, we denote by $\tau_S = \tau_S(1) = \inf \{n \geq 1, X_n \in S\}$ and $\tau_S(j) = \inf \{n > \tau_S(j-1), X_n \in S\}$, $j \geq 2$, the times of the successive visits to the small set S . We consider the joint distribution such that, conditionally on the sample path $X^{(n+1)} = (X_1, \dots, X_{\tau_S(1)}, \dots, X_{\tau_S(L_n)}, \dots, X_{n+1})$, denoting by $L_n = \sum_{i=1}^n 1_S(X_i)$ the number of visits of X to the small set S between time 1 and time n , the (Y_i, \hat{Y}_i) 's are drawn independently for $1 \leq i \leq n$ so that

$$\begin{aligned} Y_{\tau_S(k)} &\sim \text{Ber}(\delta\phi(X_{\tau_S(k)+1}) / p(X_{\tau_S(k)}, X_{\tau_S(k)+1})), \\ \hat{Y}_{\tau_S(k)} &\sim \text{Ber}(\delta\phi(X_{\tau_S(k)+1}) / p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})), \end{aligned}$$

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

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

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

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

for $k \in \{1, \dots, L_n\}$, and that for all $i \in \{1, 1, \dots, n\} \setminus \{\tau_S(k), 1 \leq k \leq L_n\}$, $Y_i = \hat{Y}_i \sim \text{Ber}(\delta)$. Hence, we deduce that, for $1 \leq k \leq L_n$,

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

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

Observe that, we almost surely have

$$\left| \frac{\delta\phi(X_{k+1})}{p(X_k, X_{k+1})} - \frac{\delta\phi(X_{k+1})}{p_n(X_k, X_{k+1})} \right| \leq 1_S(X_{k+1}) \frac{|p(X_k, X_{k+1}) - p_n(X_k, X_{k+1})|}{p(X_k, X_{k+1})}.$$

Consequently, we may write

$$\begin{aligned} l_1(P^{(n)}, \widehat{P}^{(n)}) &\leq \sum_{k=1}^{n-1} 2^{-k} E[I\{X_k \in S, X_{k+1} \in S\} \\ &\quad \times \frac{|p(X_k, X_{k+1}) - p_n(X_k, X_{k+1})|}{p(X_k, X_{k+1})}] \\ &\leq \sum_{k=1}^{n-1} 2^{-k} E[I\{X_k \in S, X_{k+1} \in S\} \\ &\quad \times \frac{|p(X_k, X_{k+1}) - p_n(X_k, X_{k+1})|}{\delta\phi(X_{k+1})}] \end{aligned}$$

Hence, under (i), we have

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

Thus, given the asymptotic properties of p_n we assumed, we get the wished bound $(\delta \inf_{x \in S} \phi(x))^{-1} \alpha_n^{1/2}$.

5.3 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, since the same argument applies to the general unbounded case except for light and obvious modifications. The proof is based on the study of the closeness between the distribution of the blocks $\mathcal{B}_1, \dots, \mathcal{B}_{l_n}$ dividing the segment $X^{(n+1)} = (X_1, \dots, X_{n+1})$ 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 $\widehat{\mathcal{B}}_1, \dots, \widehat{\mathcal{B}}_{\widehat{l}_n}$ dividing $X^{(n+1)}$ according to the \widehat{l}_n successive visits of (X_i, \widehat{Y}_i) to $S \times \{1\}$ conditionally to $X^{(n+1)}$. Let us assume that, conditionally to $X^{(n+1)}$, the (Y_i, \widehat{Y}_i) 's are drawn as supposed in subsection 5.2. We shall use the notations $l_n = \sum_{i=1}^n I\{X_i \in S, Y_i = 1\}$, $\tau_{A_{\mathcal{M}}} = \tau_{A_{\mathcal{M}}}(1) = \inf\{n \geq 1, (X_n, Y_n) \in A_{\mathcal{M}}\}$, $\tau_{A_{\mathcal{M}}}(j+1) = \inf\{n > \tau_{A_{\mathcal{M}}}(j), (X_n, Y_n) \in A_{\mathcal{M}}\}$, $l(\mathcal{B}_j) = \tau_{A_{\mathcal{M}}}(j) - \tau_{A_{\mathcal{M}}}(j-1)$ for $j \geq 1$.

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 $\sigma_n^2(f) = n_{A_{\mathcal{M}}}^{-1} \sum_{j=1}^{l_n} \{f(\mathcal{B}_j) - \mu_n(f)l(\mathcal{B}_j)\}^2$ be the respective counterparts of $\hat{\mu}_n(f)$ and $\hat{\sigma}_n^2(f)$ based on the regenerative blocks.

We first begin by controlling the difference between the first (resp. the last) pseudo regeneration time and the true one. For the sake of the simplicity, we introduce further notation and denote by $\tau_1 = \tau_{A_{\mathcal{M}}}$ (respectively, $\hat{\tau}_1 = \hat{\tau}_{A_{\mathcal{M}}}$) the (random) time corresponding to the first visit to $S \times \{1\}$ of $(X_i, Y_i)_{1 \leq i \leq n}$ (resp., of $(X_i, \hat{Y}_i)_{1 \leq i \leq n}$) as well as the time $\tau_2 = \tau_{A_{\mathcal{M}}}(l_n)$ (resp., $\hat{\tau}_2 = \hat{\tau}_{A_{\mathcal{M}}}(\hat{l}_n)$) corresponding to the last visit.

Lemma 5.1 *Let $\gamma \geq 2$. Under $\mathcal{H}_2(2\gamma)$ and \mathcal{H}_3 , there exists a constant C such that for $i \in \{1, 2\}$,*

$$E_\nu(|\hat{\tau}_i - \tau_i|^\gamma) \leq C\alpha_n^{1/2}.$$

Proof. Let C denote a constant that is not necessarily be the same at each appearance in what follows. Given assumptions \mathcal{H}_4 and \mathcal{H}_5 , note that $\delta\phi(X_{\tau_S(k)+1})/p(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ and $\delta\phi(X_{\tau_S(k)+1})/p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ 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) 's (refer to 5.2 above) and in particular that

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

for any $k \in \{1, \dots, L_n\}$, one may derive the following bound for the conditional expectation

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

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

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

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 $E_\nu(\tau_S(k)^{2\gamma})$ is bounded by $2^{2\gamma} \{E_\nu(\tau_S^{2\gamma}) + (k-1)^{2\gamma} \sup_{x \in S} E_x(\tau_S^{2\gamma})\}$ for all k . Under $\mathcal{H}_2(2\gamma)$ the bound is thus established when $i = 1$.

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

Let $g : (E, \mathcal{E}) \rightarrow \mathbf{R}$ be a bounded measurable function and set $g(\mathcal{B}_j) = \sum_{i=1+\tau_{A, \mathcal{M}}(j)}^{\tau_{A, \mathcal{M}}(j+1)} g(X_i)$. Now consider the functionals $T_n^{(M)}(g) = n^{-1} \sum_{j=1}^{l_n-1} g(\mathcal{B}_j)^M$ and $\widehat{T}_n^{(M)}(g) = n^{-1} \sum_{j=1}^{\widehat{l}_n-1} g(\widehat{\mathcal{B}}_j)^M$ for $M \in \{1, 2\}$, with by convention $T_n^{(M)}(g) = 0$ (respectively, $\widehat{T}_n^{(M)}(g) = 0$) when $l_n \leq 1$ (resp., when $\widehat{l}_n \leq 1$). The following lemma provides an asymptotic bound for

$$D_n^{(M)}(g) = \left| T_n^{(M)}(g) - \widehat{T}_n^{(M)}(g) \right|,$$

with $M = 1, \dots, 3$.

Lemma 5.2 *Let $\gamma \geq 6$. Under $\mathcal{H}_2(\gamma)$ and \mathcal{H}_3 , we have as $n \rightarrow \infty$*

$$D_n^{(1)}(g) = O_{P_\nu}(n^{-1}\alpha_n^{1/2}), \quad (7)$$

$$D_n^{(k)}(g) = O_{P_\nu}(\alpha_n), \quad k = 2, 3 \quad (8)$$

as $n \rightarrow \infty$.

Proof. Bound (7) immediately follows from lemma 5.1.

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

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

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

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

$$\begin{aligned} |D_n^{(2)}(g)| &\leq \frac{\|g\|_\infty^2}{n} \left\{ (\widehat{\tau}_1 - \tau_1)^2 + (\widehat{\tau}_2 - \tau_2)^2 + 2 \sum_{j=1}^{N_n} (t_j^{(2)} - t_j)(t_j - t_j^{(1)}) \right\} \\ &\leq \frac{\|g\|_\infty^2}{n} \left\{ (\widehat{\tau}_1 - \tau_1)^2 + (\widehat{\tau}_2 - \tau_2)^2 + \sum_{j=1}^{N_n} \left\{ (t_j^{(2)} - t_j)^2 + (t_j - t_j^{(1)})^2 \right\} \right\}. \end{aligned}$$

Set $t_j = t_j^{(2)} = 0$ for $j > N_n$. By proceeding analogously as we did previously, one easily shows that there exist constants $c_r(q)$ depending only on q , such that $E_\nu((t_j^{(2)} - t_j)^r) \leq c_r(q) \sup_x E_x(\tau_S^r)$ for any $j \geq 1$, $r \leq \gamma$. By Cauchy-Schwarz's inequality we have for any deterministic sequence of positive integers m_n ,

$$\begin{aligned} E_\nu\left(\sum_{j=1}^{N_n} (t_j^{(2)} - t_j)^2\right) &\leq \sum_{j=1}^{m_n} E_\nu((t_j^{(2)} - t_j)^2) + E_\nu\left(\sum_{j=1}^n (t_j^{(2)} - t_j)^2 I\{N_n > m_n\}\right) \\ &\leq C_1 m_n + C_2 n (P(N_n > m_n))^{1/2}, \end{aligned}$$

where $C_1 = \sup_{x \in S} E_x(\tau_S^2) c_2(q)$ and $C_2 = \sup_{x \in S} E_x(\tau_S^4) c_4(q)$. As shown in 5.2, the probability that Y_i differs from \widehat{Y}_i is bounded by $q_n = (\delta \inf_{x \in S} \phi(x))^{-1} \alpha_n$ and the $(Y_{\tau_S(k)}, \widehat{Y}_{\tau_S(k)})$'s are drawn independently conditionally to $X^{(n+1)}$. Hence, by using Chebyshev's exponential inequality, we derive that

$$P(N_n > m_n) \leq e^{-m_n} E(e^{N_n}) \leq e^{-m_n} (1 + q_n e)^n. \quad (9)$$

Now by choosing $m_n \uparrow \infty$ such that $m_n/n \rightarrow 0$ and $n\alpha_n/m_n \rightarrow 0$ as $n \rightarrow \infty$, we deduce from (9) that $n^{-1} E_\nu(\sum_{j=1}^{N_n} (t_j^{(2)} - t_j)^2) = O(\alpha_n)$, as $n \rightarrow \infty$. By an analogous argument, one shows that $n^{-1} E_\nu(\sum_{j=1}^{N_n} (t_j - t_j^{(1)})^2) = O(\alpha_n)$, as $n \rightarrow \infty$. $D_n^{(3)}(g)$ may be treated similarly. ■

We deduce from this result that the following empirical quantities based on the pseudo-blocks converge to their respective counterparts based on the regenerative blocks.

Lemma 5.3 *Under the assumptions of Theorem 3.2, we have the following convergences in P_ν -probability as $n \rightarrow \infty$,*

$$n^{1/2}(\widehat{\mu}_n(f) - \mu_n(f)) \rightarrow 0, \quad (10)$$

$$\left| \widehat{l}_n/n - l_n/n \right| = O(\alpha_n^{1/2}) \quad (11)$$

Moreover we have for $k = 2, 3$

$$n^{-1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j)^k - n^{-1} \sum_{j=1}^{\widehat{l}_n-1} l(\widehat{\mathcal{B}}_j)^k = O_{P_\nu}(\alpha_n), \quad (12)$$

$$n^{-1} \sum_{j=1}^{l_n-1} f(\mathcal{B}_j)^k - n^{-1} \sum_{j=1}^{\widehat{l}_n-1} f(\widehat{\mathcal{B}}_j)^k = O_{P_\nu}(\alpha_n), \quad (13)$$

$$n^{-1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j) f(\mathcal{B}_j) - n^{-1} \sum_{j=1}^{\widehat{l}_n-1} l(\widehat{\mathcal{B}}_j) f(\widehat{\mathcal{B}}_j) = O_{P_\nu}(\alpha_n). \quad (14)$$

Proof. From (7) in lemma 5.2 with $g \equiv 1$, it follows that $\left| \widehat{n}_{A_{\mathcal{M}}}/n - n_{A_{\mathcal{M}}}/n \right| = O_{P_\nu}(\alpha_n^{1/2} n^{-1})$ as $n \rightarrow \infty$. Given that $n_{A_{\mathcal{M}}}/n \rightarrow 1$ P_ν a.s. as $n \rightarrow \infty$, this combined to (7) again with $g = f$ obviously yield (10).

Now observe that

$$\left| \widehat{l}_n/n - l_n/n \right| \leq n^{-1} \sum_{k=1}^{L_n} I\{Y_{\tau_S(k)} \neq \widehat{Y}_{\tau_S(k)}\},$$

Using again the fact that conditionally to $X^{(n+1)}$ the $(Y_{\tau_S(k)}, \widehat{Y}_{\tau_S(k)})$'s are drawn independently for $k = 1, \dots, L_n$ and $P(\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)} \mid X^{(n+1)}) \leq (\delta \inf_{x \in S} \phi(x))^{-1} \sup_{(x,y) \in S^2} |p_n(x,y) - p(x,y)|$, this entails

$$E\left(\left| \widehat{l}_n/n - l_n/n \right| \mid X^{(n+1)}\right) \leq n^{-1} L_n (\delta \inf_{x \in S} \phi(x))^{-1} \sup_{(x,y) \in S^2} |p_n(x,y) - p(x,y)|.$$

Since $L_n = \sum_{i=1}^n I\{X_i \in S\} \leq n$, taking the expectation implies that $E_\nu\left(\left| \widehat{l}_n/n - l_n/n \right|\right) = O(\alpha_n^{1/2})$, as $n \rightarrow \infty$.

Now, (12) (respectively, (13)) straightforwardly results from lemma 5.2 with $g \equiv 1$ (resp., with $g = f$).

And (14) may be proved by simply noticing that

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

and applying lemma 5.2 to each component on the right hand side (by taking successively $g(x)$ equal to $\overline{f}(x) = f(x) - \mu(f)$, $f(x)$ and 1). ■

Now, one easily deduce from these results that $\widehat{\sigma}_n^2(f) - \sigma_n^2(f) \rightarrow 0$ in P_ν -*pr.*, as $n \rightarrow \infty$. Hence, given that $\sigma_n^2(f) - \sigma_f^2$ in P_ν -*pr.*, as $n \rightarrow \infty$ (see the preliminary remarks in §.2.4.1) the consistency of $\widehat{\sigma}_n^2(f)$ is established.

Finally, combining this to (10) and the CLT for the sample mean $\mu_n(f)$ relating to the atomic split chain (cf § 2.4.1) proves that $\widehat{n}_{A_{\mathcal{M}}}^{1/2} \widehat{\sigma}_n(f)^{-1} (\widehat{\mu}_n(f) - \mu(f)) \rightarrow \mathcal{N}(0, 1)$ in P_ν -distribution, as $n \rightarrow \infty$.

5.4 Proof of Theorem 3.3

We first recall the "Bootstrap mode of convergence". In what follows, we shall write $Z_n \xrightarrow{P^*} Z$ in P_ν -probability (resp., P_ν a.s.) along the sample when

$$P^*(|Z_n - Z| > \varepsilon \mid X^{(n+1)}) \xrightarrow{n \rightarrow \infty} 0 \text{ in } P_\nu\text{-probability (resp., } P_\nu \text{ a.s.)}.$$

The unstudentized case Note that one may write

$$n_{A_{\mathcal{M}}}^{*1/2} \frac{(\mu_n^*(f) - \widehat{\mu}_n(f))}{\widehat{\sigma}_n(f)} = \frac{\sum_{j=1}^{l_n^*-1} \{f(\mathcal{B}_j^*) - \widehat{\mu}_n(f)l(\mathcal{B}_j^*)\}}{n_{A_{\mathcal{M}}}^{*1/2} \widehat{\sigma}_n(f)}.$$

The result is proved by following line by line the classical argument establishing the CLT for regenerative process (cf Theorem 17.2.2 in Meyn & 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 conditionally to $X^{(n+1)}$ the ARBB sequence (although not Markovian) defines a regenerative process with i.i.d. segments $\widehat{\mathcal{B}}_j^*$, $j \geq 1$. By the L.L.N., we have $(l_n - 1)^{-1} \sum_{j=1}^{l_n-1} l(\mathcal{B}_j) \rightarrow E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})$, P_ν a.s. . Lemma 5.3 thus entails that, as $n \rightarrow \infty$,

$$E^*(l(\widehat{\mathcal{B}}_1^*) | X^{(n+1)}) = \frac{1}{\widehat{l}_n - 1} \sum_{j=1}^{\widehat{l}_n - 1} l(\widehat{\mathcal{B}}_j) \rightarrow E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}}) \text{ in } P_\nu\text{-probability.}$$

In a similar fashion, we have as $n \rightarrow \infty$,

$$E^*(l(\widehat{\mathcal{B}}_1^*)^2 | X^{(n+1)}) \rightarrow E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}}^2) < \infty \text{ in } P_\nu\text{-probability.}$$

This implies in particular that, as $n \rightarrow \infty$,

$$\frac{l(\widehat{\mathcal{B}}_j^*)}{n} \xrightarrow{P^*} 0 \text{ and } \frac{n_{A_{\mathcal{M}}}^*}{n} \xrightarrow{P^*} 1 \text{ in } P_\nu\text{-probability along the sample.} \quad (16)$$

Now note that by definition of l_n^* we have

$$\frac{l_n^*}{\sum_{j=1}^{l_n^*} l(\widehat{\mathcal{B}}_j^*)} \leq \frac{l_n^*}{n} \leq \frac{l_n^*}{\sum_{j=1}^{l_n^*-1} l(\widehat{\mathcal{B}}_j^*)}. \quad (17)$$

Combining the L.L.N. to lemma 5.3 we have as $n \rightarrow \infty$,

$$l_n^{*-1} \sum_{i=1}^{l_n^*} l(\widehat{\mathcal{B}}_i^*) - (\widehat{l}_n - 1)^{-1} \sum_{i=1}^{\widehat{l}_n - 1} l(\widehat{\mathcal{B}}_i) \xrightarrow{P^*} 0 \text{ } P_\nu \text{ a.s.,}$$

which entails that $l_n^{*-1} \sum_{i=1}^{l_n^*} l(\widehat{\mathcal{B}}_i^*) - E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}}) \xrightarrow{P^*} 0$, in P_ν -probability along the sample. We deduce that

$$\frac{l_n^*}{n} - E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} \xrightarrow{P^*} 0, \text{ in } P_\nu\text{-probability along the sample.}$$

Thus for $\epsilon > 0$, if we set $n_1 = \lfloor (1 - \epsilon)E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1}n \rfloor + 1$ and $n_2 = \lfloor (1 + \epsilon)E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1}n \rfloor$ (denoting by $\lfloor x \rfloor$ the integer part of $x \in \mathbf{R}$), there exists $n_0 = n_0(\epsilon) \in \mathbf{N}$ such that, for $n > n_0$

$$P^*(n_1 \leq l_n^* - 1 \leq n_2 | X^{(n+1)}) \geq 1 - \epsilon \text{ in } P_\nu\text{-probability} \quad (18)$$

Combining (16), (18) and Markov's inequality implies that for any $\eta > 0$,

$$\begin{aligned}
& P^* \left(\left| \frac{\mu_n^*(f) - \hat{\mu}_n(f)}{n_{A_{\mathcal{M}}}^{*-1/2} \hat{\sigma}_n(f)} - \frac{\sum_{j=1}^{1+\lfloor nE_{A_{\mathcal{M}}}(\tau_A)^{-1} \rfloor} \{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\}}{n_{A_{\mathcal{M}}}^{*1/2} \hat{\sigma}_n(f)} \right| > \eta \mid X^{(n+1)} \right) \\
& \leq \varepsilon + P^* (\hat{\sigma}_n(f)^{-1} \max_{n_1 \leq l \leq n_2} \left| \sum_{j=l}^{1+\lfloor nE_{A_{\mathcal{M}}}(\tau_A)^{-1} \rfloor} \{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\} > \eta n_{A_{\mathcal{M}}}^{*1/2} \mid X^{(n+1)} \right) \\
& \leq \varepsilon + P^* (\hat{\sigma}_n(f)^{-1} \max_{n_1 \leq l \leq n_2} \left| \sum_{j=l}^{1+\lfloor nE_{A_{\mathcal{M}}}(\tau_A)^{-1} \rfloor} \{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\} > \eta n^{1/2}/2 \mid X^{(n+1)} \right) \\
& + P^* (n_{A_{\mathcal{M}}}^{*1/2} n^{-1/2} < 1/2 \mid X^{(n+1)}) \\
& \leq \varepsilon + 4\hat{\sigma}_n(f)^{-2} 2\varepsilon E^* (\{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\}^2) / \eta^2 + \varepsilon,
\end{aligned}$$

for n large enough. As this holds for any $\eta, \varepsilon > 0$, we get (using again (16))

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

along the sample in P_ν - probability, as $n \rightarrow \infty$. Now it is sufficient to apply the classical bootstrap CLT (see Bickel & Freedmann (1981)) to the i.i.d. r.v.'s $\{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\}_{j \geq 1}$. These r.v.'s are centered with variance

$$E^* (\{f(\mathcal{B}_j^*) - \hat{\mu}_n(f)l(\mathcal{B}_j^*)\}^2 \mid X^{(n+1)}) = \hat{n}_{A_{\mathcal{M}}} \hat{\sigma}_n^2(f) / (\hat{l}_n - 1),$$

which converges to $E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})\sigma_f^2$ in P_ν - probability under the hypotheses of Theorem 3.3 (cf Theorem 3.2 and lemma 5.3).

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

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

in P_ν - probability along the sample and the result follows also from standard bootstrap results in the i.i.d case.

6 Proof of Theorem 3.4

Under our assumptions, the E.E. of the sample mean in the unstudentized case is given in Malinowskii (1987), see 6 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 E.E. for the ARBB. The proof is thus similar to the proof of Theorem 2.1 except for replacing the true regenerative blocks by the approximated ones. Lemma 5.3 allows to control the rate of convergence of all the empirical terms appearing in the E.E. of the ARBB distribution up to $O_P(\alpha_n^{1/2}n^{-1/2})$. The recentering ensures that the bias vanishes so that the two E.E. match up to $O_P(\alpha_n^{1/2}n^{-1/2})$.

References

- [1] Asmussen, S. (1987). *Applied Probabilities and Queues*. Wiley.
- [2] Athreya, K., Atuncar, G. (1998). Kernel estimation for real-valued Markov chains. *Sankhya*, **60**, series A, No 1, 1-17.
- [3] Athreya, K., Fuh, C. (1989). Bootstrapping Markov chains: countable case. *Tech. Rep.* B-89-7, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC..
- [4] Bertail, P. (1997). Second order properties of an extrapolated bootstrap without replacement: the i.i.d. and the strong mixing cases. *Bernoulli*, **3**, 149-179.
- [5] Bertail, P., Cléménçon, S. (2004). Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. to appear *Prob. Th. Rel. Fields*, 129, n°4.
- [6] Bertail, P., Cléménçon, S. (2003). Note on the regeneration-based bootstrap for atomic Markov chains. *Submitted*.
- [7] Bickel, P., Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Stat.* **9**, 1196-1217.
- [8] Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahr. verw. Gebiete*, **65**, 181-237.
- [9] Bolthausen, E. (1980). The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains. *Z. Wahr. Verw. Gebiete*, **54**, 59-73.
- [10] Bolthausen, E. (1982). The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains. *Z. Wahr. Verw. Gebiete*, **60**, 283-289.
- [11] Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Stat.*, **16**, 1709-1722.

- [12] Brockwell, P., Resnick, S., Tweedie, R. (1982). Storage processes with general release rules and additive inputs. *Adv. Appl. Prob.*, **14**, 392-433.
- [13] Browne, S., Sigman, K. (1992). Work-modulated queues with applications to storage processes. *J. Appl. Prob.*, **29**, 699-712.
- [14] Bühlmann, P. (1997). Sieve Bootstrap for time series. *Bernoulli*, **3**, 123-148.
- [15] Cléménçon, S. (2000). Adaptive estimation of the transition density of a regular Markov chain. *Math. Meth. Stat.*, **9**, No. 4, 323-357.
- [16] Cléménçon, 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.
- [17] Datta, S., McCormick W. (1993). Regeneration-based bootstrap for Markov chains. *Can. J. Stat.*, **21**, No.2, 181-193.
- [18] Datta, S., McCormick, W. (1995). Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *J. Mult. Analysis*, **52**, 83-106.
- [19] Douc, R. , Fort, G., Moulines E. and Soulier P. (2004). Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability*, **14**, 1353-1377.
- [20] Doukhan, P., Ghindès, M. (1983). Estimation de la transition de probabilité d'une chaîne de Markov Doeblin récurrente. *Stochastic Process. Appl.*, **15**, 271-293.
- [21] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Stat.*, **7**, 1-26.
- [22] Feller, W. (1971). *An Introduction to Probability Theory and its Applications: vol. I and II*. John Wiley & Sons, NY, 3rd edition.
- [23] Franke, J. , Kreiss, J. P., Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*, **8**, 1-37.
- [24] Götze, F., Künsch, H. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *Ann. Stat.*, **24**, 1914-1933.
- [25] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.
- [26] Harrison, J., Resnick, S. (1976). The stationary distribution and first exit probabilities of a storage process with general release rule. *Math. Oper. Res.*, **1**, 347-358.
- [27] Horowitz, J. (2003) Bootstrap methods for Markov processes. *Econometrica*, **71**, 1049-1082 .
- [28] Jain, J., Jamison, B. (1967). Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.*, **8**, 19-40.

- [29] Kalashnikov, V. (1978). *The Qualitative Analysis of the Behavior of Complex Systems by the Method of Test Functions*. Nauka, Moscow.
- [30] Lahiri, S. (2003). *Resampling methods for Dependant Data*. Springer Series in Statistics. Springer Verlag, N.Y.
- [31] Malinovskii, V. (1987). Limit theorems for Harris Markov chains I. *Theory Prob. Appl.*, **31**, 269-285.
- [32] Malinovskii, V. (1989). Limit theorems for Harris Markov chains II. *Theory Prob. Appl.*, **34**, 252-265.
- [33] Meyn, S., Tweedie, R., (1996). *Markov chains and stochastic stability*. Springer.
- [34] Nummelin, E. (1978). A splitting technique for Harris recurrent chains. *Z. Wahrsch. Verw. Gebiete*, **43**, 309-318.
- [35] Paparoditis, E. and Politis, D. (2002). The local bootstrap for Markov processes. *J. Statist. Plan. Infer.*, **108**, 301–328.
- [36] Politis, D.N., Romano, J.P., Wolf, T. (2000). *Subsampling*. Springer Series in Statistics, Springer, N.Y.
- [37] Rachev, S., Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. II: Applications*. Springer.
- [38] Rajarshi, M.B. (1990). Bootstrap in Markov sequences based on estimates of transition density. *Ann. Inst. Stat. Math.*, **42**, 253-268.
- [39] Revuz, D. (1984). *Markov chains*. North-Holland, 2nd edition.
- [40] Roberts, G., Rosenthal, J. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. *Electr. Journ. Prob.*, **9**, 1-21.
- [41] Rosenblatt, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference*, Ed. M. Puri, 199-210.
- [42] Roussas, G. (1969). Nonparametric Estimation in Markov Processes. *Ann. Inst. Stat. Math.*, 73-87.
- [43] Smith, W. (1955). Regenerative stochastic processes. *Proc. Royal Stat. Soc., A*, **232**, 6-31.
- [44] Thorisson, H. (2000). *Coupling, Stationarity and Regeneration*. Springer.