Regenerative block-bootstrap for Markov chains*

Patrice Bertail,
CREST- Laboratoire de Statistiques

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

Abstract

This paper is devoted to the construction and the study of 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 $X_1, ..., X_n$ 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. This methodology mainly consists in dividing the trajectory of the chain into i.i.d. blocks corresponding to the successive visits to the atom and resampling the blocks. We prove that in the atomic case our method inherits the $O_P(n^{-1})$ accuracy of the Bootstrap in the i.i.d. case (up to a logarithmic factor). In the general case (including the nonstationary case), asymptotic validity for this resampling procedure is established, provided that a consistent estimator of the transition kernel may be computed. Applications to specific Markovian models are discussed, together with some simulation results.

*Keywords : Bootstrap, Markov Chain, Regenerative Processes, Nummelin splitting technique.
Primary Class : 62G09, **Secondary Class** : 62G20, 60J10

1 Introduction

1.1 Bootstrap methods for Time Series

In the last few years, 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 (for an introduction refer to Chapter 9 in Shao & Tu (1995) and the references therein). The main idea underlying these generalizations is to resample blocks of observations to mimic the dependence of the data. The idea of the *moving-block bootstrap* (MBB) (see Hall (1985), Carlstein (1986), Künsch (1989) and Liu & Singh (1992)) is to resample (overlapping or disjoint) blocks of observations to capture the dependence structure of the observations. Refer to Bühlmann (2002), Politis (2003) for recent survey and more complete 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.

• 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 Lahiri (1992), Politis & Romano (1992), Götze & Künsch (1996)). 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)), but this practical problem still remains open in the general case.

Recently, several authors have been interested in bootstrapping some particular type of Markov chains. 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). Based on these ideas, Bühlmann (1997) has considered a sieve bootstrap method based

on the approximation of times 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 (2002) 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 (2001) 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 models considered. Most of the time, under these regularity assumptions, the conditions of Götze & Hipp (1983) may be checked directly on these models, so that Edgeworth expansions are immediately available and may be inverted straightforwardly, yielding even better results than what can be expected with these methods. 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. We call this method the *Regenerative Block Bootstrap* (RBB). The problem of extending the RBB methodology to general (eventually nonstationary) Harris recurrent Markov chains is also addressed in this article.

1.2 On the description of Markov chains via Renewal and Regenerative processes

Renewal theory plays a key role in the analysis of the asymptotic structure of many kinds of stochastic processes, and especially in the development of asymptotic properties of general irreducible Markov chains. The underlying ground consists in the fact that limit theorems proved for sums of independent random vectors may be easily extended to regenerative random processes, that is to say random processes that may be decomposed at random times, called regeneration times, into a sequence of mutually independent segments, namely regeneration cycles (see Smith (1955)). The method based on this principle is traditionally called the *regenerative method*. As will be recalled at length in subsection 3.1, Harris chains that possess an atom, i.e. a Harris set on which the transition probability kernel is constant, are special cases of regenerative processes and so fall into the range of application

of the regenerative method. The theory developed in Nummelin (1978) (and in parallel the closely related concepts introduced in Athreya & Ney (1978)) showed that general Markov chains could all be considered as regenerative in a broader sense (*i.e.* in the sense of the existence of a regenerative extension for the chain, see subsection 3.2), as soon as the Harris recurrence property is satisfied. Hence this theory made the regenerative method applicable to the whole class of Harris Markov chains and allowed to carry over many limit theorems to Harris chains. The aim of this paper is to reexamine and develop further the application of the regenerative method to construct a data-resampling procedure for Markov chains.

The powerful ideas introduced in Athreya & Fuh (1989) and Datta & McCormick (1993) do not seem to be widely known in the Bootstrap literature, nor used in practice. This may be partly due to the fact that they only deal with the restrictive case of Markov chains possessing a known atom under rather strong assumptions, so that the scope of applications is limited. Moreover, because of some inadequate standardization, the *regeneration-based bootstrap* method proposed in Datta & McCormick (1993) is not second order correct and performs very poorly in the applications (see Bertail & Clémençon (2003b)). The main idea underlying this method consists in resampling a deterministic number of data blocks corresponding to regeneration cycles. Bertail & Clémençon (2003b) have proposed a modification of the procedure introduced by Datta & McCormick (1993), which is second order correct in the unstudentized case (*i.e.* when the variance is known) when the chain is stationary. They showed that, provided the bootstrap statistic is suitably standardized and recentered, the *regeneration-based bootstrap* achieves the rate $O_P(n^{-1}\log(n))$ in the stationary case. 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 & Clémençon (2003a)) shows: nonstationarity induces a significant bias, that may be decomposed into three components, each of them being induced by a specific segment of the data, and cannot be estimated when the number of resampled blocks is held fixed conditionally to the original data, as for the regeneration-based bootstrap. Our proposal consists 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 the 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 and show how it may be extended to the much broader class of Harris Markov chains.

We first 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 (1968, 71)). We demonstrate here the power of this method for suitably standardized statistics: the RBB has in particular an uniform rate of convergence of order $O_P(n^{-1}\log(n))$, close to the optimal rate in the i.i.d case. Moreover, it is noteworthy that, 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 mixing conditions (polynomial rate for the decay of the strong mixing coefficients).

Then we extend this methodology to general positive recurrent 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. Unfortunately, this requires to compute a consistent estimate of the transition kernel of the chain. We establish the asymptotic validity of this procedure, even in a nonstationary framework, that is clearly more suitable for many applications. The study of the second order properties of this general method and of the optimal rate that may be attained presents severe technical difficulties and will be carried out at length in a forthcoming article. Here we essentially 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.

1.3 Outline

In section 2, notations are set out and a few definitions concerning the communication structure and the stochastic stability of Markov chains are given. Some basics about the regenerative method and the Nummelin splitting technique are then recalled. In section 3, our proposal for bootstrapping atomic chains is described. Beyond the actual RBB algorithm, an asymptotic result claiming the second order validity of the RBB method for studentized sample mean statistics is stated. Section 4 deals with the extension of the RBB procedure to general positive recurrent chains. Two illustrative examples are given in section 5. And technical proofs are detailed in section 6.

2 Theoretical background

2.1 Notation and definitions

We shall introduce some notation and recall key notions of the Markov chain theory that will be needed throughout the paper (for further detail, refer to Revuz (1984)). Let $X = (X_n)_{n\in\mathbb{N}}$ be an aperiodic 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 any $n \in \mathbb{N}$, we have

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

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. $X_0 = x$), $E_\mu(.)$ the P_ν-expectation (resp. $E_x(.)$ the P_x-expectation), and $I\{\mathcal{A}\}$ will denote the indicator function of the event $\mathcal{A}$.

A measurable set B is *Harris recurrent* for the chain if for any $x \in B$, $P_x(\sum_{n=1}^{\infty} I\{X_n \in B\} = \infty) = 1$. And the chain is said *Harris recurrent* if it is ψ-irreducible and every measurable set B such that $\psi(B) > 0$ is Harris recurrent. When the chain is Harris recurrent, we have the property that $P_x(\sum_{n=1}^{\infty} I\{X_n \in B\} = \infty) = 1$ for any $x \in E$ and any $B \in \mathcal{E}$ such that $\psi(B) > 0$.

And as it is obviously easier to deal with chains with time-invariant marginal distributions for statistical purposes, it is natural to require the existence of an invariant probability measure. A probability measure μ on E is said invariant for the chain when $\mu\Pi = \mu$, where $\mu\Pi(dy) = \int_{x\in E} \mu(dx)\Pi(x, dy)$. An irreducible chain is said *positive recurrent* when it admits an invariant probability (it is then unique).

Now we recall some basics concerning the regenerative method and its application to the analysis of the behaviour of general Harris chains via the Nummelin splitting technique (refer to Nummelin (1984) for further detail).

2.2 Chains possessing an atom

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:

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

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 $E_A(.)$ 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 what the starting point:

$$\forall x \in E,\ P_x\left(\tau_A < \infty\right) = 1.$$

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 the atom A

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

taking their values in the torus $T = \cup_{n=1}^{\infty} E^n$.

The sequence $\left(\tau_A(j)\right)_{j\geqslant 1}$ 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 following result, known as Kac's theorem, holds.

Theorem 2.1 *The chain $(X_n)_{n\in\mathbb{N}}$ is positive recurrent if and only if $E_A(\tau_A) < \infty$. In such a case the unique invariant probability distribution μ is the occupation measure given by:*

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

For such chains, limit theorems can be derived from the application of the corresponding results to the i.i.d. blocks $(\mathcal{B}_n)_{n\geqslant 1}$. One may refer for example 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 (see Clémençon (2001)).

2.3 General Harris chains

In this subsection, we recall the *splitting technique* introduced in Nummelin (1978), which allows to extend in some sense the probabilistic structure of the chain in order to artificially construct a regeneration set, an atom in the general Harris recurrent case. First, consider the following notion.

Definition 2.2 *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 (see Jain & Jamison (1967)). We assume that the chain is Harris recurrent. Now let us 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 $\big((X_{nm}, ..., X_{n(m+1)-1})\big)_{n\in\mathbb{N}}$, 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 $Y_n = 0$, (which happens with probability $1-\delta$), then X_{n+1} is distributed according to the probability measure $(1-\delta)^{-1}(\Pi(X_{n+1}, .) - \delta\Phi(.))$.

Set $Ber_\delta(\beta) = \delta\beta + (1-\delta)(1-\beta)$ for $\beta \in \{0, 1\}$. 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 the state space $E \times \{0, 1\}$ with a transition probability kernel $\Pi_{\mathcal{M}}$ defined by

- for any $x \in S^c$, $B \in \mathcal{E}$, β and β' in $\{0, 1\}$,

$$\Pi_{\mathcal{M}}((x, \beta), B \times \{\beta'\}) = Ber_\delta(\beta') \times \Pi(x, B),$$

- for any $x \in S$, $B \in \mathcal{E}$, β' in $\{0, 1\}$,

$$\Pi_{\mathcal{M}}((x, 1), B \times \{\beta'\}) = Ber_\delta(\beta') \times \Phi(B),$$
$$\Pi_{\mathcal{M}}((x, 0), A \times \{\beta'\}) = Ber_\delta(\beta') \times (1-\delta)^{-1}(\Pi(x, B) - \delta\Phi(B)).$$

The whole point of the construction consists in the fact that $S \times \{1\}$ is an atom for the split chain $X^{\mathcal{M}}$, which inherits all the communication and stochastic stability properties from X (irreducibility, Harris recurrence,...), in particular (for the case $m = 1$ here) the blocks constructed for the split chain are independent. Hence the splitting method enables to extend the regenerative method, and so to establish all of the results known for atomic chains, to general Harris chains. It should be noticed that if the chain 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 will omit the subscript $\mathcal{M}$ in what follows and abusively denote by P_ν the extensions of the underlying probability we shall consider.

3 Bootstrapping Markov chains with an atom

Let $X^{(n)} = (X_1, ..., X_n)$ 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 6 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

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

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

Let $f : E \to \Re$ be a μ-integrable function and consider the estimator $\overline{\mu}_n(f) = n^{-1} \sum_{i=1}^{n} f(X_i)$ of the unknown mean $\mu(f) = E_\mu(f(X_1))$ constructed from the whole data segment $X^{(n)}$. In Bertail & Clémençon (2003a) (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.* from the first visit to the atom 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 $X_{\tau_A}^{\tau_A(l_n)} = (\mathcal{B}_1, ..., \mathcal{B}_{l_n-1})$ 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 $(X_{\tau_A+1}, ..., X_{\tau_A(l_n)})$

$$\mu_n(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{i=1}^{l_n-1} f(X_i)$$

with the convention that $\mu_n(f) = 0$ when $l_n < 2$.

Hence, given a statistic $T_n = T_n(X_{\tau_A}^{\tau_A(l_n)})$ estimating some parameter θ based on $X_{\tau_A}^{\tau_A(l_n)}$ only, we are interested in estimating accurately its sampling distribution under P_ν, that is to say to obtain a sharp approximation of $H_{P_\nu}^{(n)}(x) = P_\nu(H_n(X_{\tau_A}^{\tau_A(l_n)}, \theta) \leqslant x)$, $x \in \Re$, where $H_n(.)$ is a root in the "Bootstrap literature" sense, that is either the difference $T_n - \theta$ or its absolute value $|T_n - \theta|$ (or $S_n^{-1}(T_n - \theta)$, when some adequate standardization is available) when θ is real, and $d(T_n, \theta)$ for some appropriate distance d in the general case. For the sake of the simplicity we only focus in the sequel on sampling distributions of type

$$H_{P_\nu}^{(n)}(x) = P_\nu(S_n^{-1}(T_n - \theta) \leqslant x)$$

for some real parameter θ and an adequate standardization S_n. 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 & Clémençon (2003a)) (see subsection 3.2). For the reasons mentioned above, the estimate of the asymptotic variance of T_n we shall consider is also constructed from the regenerative data blocks $\mathcal{B}_1, ..., \mathcal{B}_{l_n-1}$ only. This greatly simplifies the argument for establishing the Edgeworth expansion of the standardized sum (refer to Bertail & Clémençon (2003a) for further details).

3.1 The regenerative block-bootstrap algorithm

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 $X^{(n)} = (X_1, ..., X_n)$ into $l_n + 1$ blocks, $\mathcal{B}_0$,

$\mathcal{B}_1, ..., \mathcal{B}_{l_n-1}, \mathcal{B}_{l_n}^{(n)}$ 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 (non regenerative) blocks.

2. Draw sequentially bootstrap data blocks $\mathcal{B}_{1,n}^*, ..., \mathcal{B}_{k,n}^*$ independently from the empirical distribution $F_n = (l_n - 1)^{-1} \sum_{j=1}^{l_n-1} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leqslant j \leqslant l_n-1}$ 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 $l_n^* = \inf\{k \geqslant 1, l^*(k) > b_n\}$.

3. From the bootstrap data blocks generated at step 2, reconstruct a pseudo-trajectory of size b_n by binding the blocks together

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

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

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

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

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

The *RBB distribution* is then given by

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

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

Remark 3.1 *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 since Bretagnolle (1983) 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 (and knowledge on the convergence rate of the statistic) : see Politis & Romano (1994), 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 3.2 *One may naturally compute a Monte-Carlo approximation to* $H_{RBB}(x)$ *by repeating independently the procedure above* B *times.*

Remark 3.3 *We point out that the RBB differs from the regeneration-based bootstrap proposed by Datta & McCormick (1993) (and from its modified version in Bertail & Clémençon (2003b) as well) 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 (nor stationary). 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 & Clémençon (2003b).*

Remark 3.4 *We also emphasize that the principles underlying the RBB may be applied to any (eventually continuous time) regenerative process (and not necessarily markovian).*

3.2 Second order accuracy of the RBB

Prolongating the work of Datta & McCormick (1993), in which the regeneration-based bootstrap is introduced, Bertail & Clémençon (2003b) showed that a suitably modified version of their bootstrap methodology is second order order correct up to $O_P(n^{-1}\log(n))$ for the studentized mean, in the stationary case only. Given the necessary modifications (standardization and recentering) and the restrictive stationary framework required to obtain the second order accuracy, this Bootstrap method is of limited interest from a practical point of view. We study here the asymptotic validity of the RBB for the studentized mean 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 Götze & Künsch (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 and $O_P(n^{-1}\log(n))$ for the studentized sample mean. The logarithmic factor results from some crude approximation in establishing the Edgeworth expansion (E.E. in abbreviated form) but one may reasonably expect that some technical refinements could allow to assess the rate $O_P(n^{-1})$. The rate obtained is nevertheless comparable to the optimal rate of the Bootstrap distribution in the i.i.d. case, contrary to the MBB (see Götze & Künsch (1996)). The proof relies on the E.E. for the studentized sample mean established in Bertail & Clémençon (2003a), which result mainly derives from the methods used in Malinovskii (1987) to obtain the E.E. for the unstandardized sample mean (see also Malinovskii (1985, 89) and Bolthausen (1980)).

3.2.1 Further notations and preliminary remarks

We set

$$n_A = \tau_A(l_n) - \tau_A(1) = \sum_{j=1}^{l_n-1} l(\mathcal{B}_j),$$

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

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

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

In the following, we also set

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

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 $n \to \infty$

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

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 have been studied in Bertail & Clémençon (2003a). A straightforward application of the SLLN for positive recurrent Markov chains shows it is strongly consistent. Under some further regularity conditions, Bertail & Clémençon (2003a) have also shown that its bias is of order $O(n^{-1})$ and it is asymptotically normal. As will be shown below, the regenerative properties allow to exhibit a straightforward standardization, that 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)). In most practical situations (except for the very special case of m-dependence), positive moving-block based estimates of the variance with such good properties are not available. Moreover, in opposition to the MBB, the Bootstrap counterparts may be defined straightforwardly in our regenerative setting. With $n_A^* = \sum_{j=1}^{l_n^*-1} l(\mathcal{B}_j^*)$, we define

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

3.2.2 Main asymptotic result

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

Theorem 3.1 *Assume that the chain X fulfills the following conditions,*

(i) (Cramer condition)

$$\overline{\lim_{|t|\to\infty}} |E_A(\exp(it(\sum_{i=1}^{\tau_A}\{f(X_i) - \mu(f)\}))| < 1.$$

(ii) (Non degeneracy of the asymptotic variance)

$$\sigma_f^2 > 0.$$

(iii) ("Block moment conditions") For some integer $s \geq 2$,

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

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

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

(v) (Non trivial regeneration set)

$$E_A(\tau_A) > 1.$$

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

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

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 standardized sample mean is also second order correct*

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

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 for atomic chains the RBB has nearly 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 section 2.4 in Meyn & Tweedie (1996), Feller (1968, 71) and Asmussen (1987) for more details of such models and for instance the popular storage model in Example 1 of section 6). 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 $\rho > s - 1$.

4 Approximate regenerative block-bootstrap for positive recurrent chains

Consider now observations $X_1, ..., X_{n+1}$ drawn from a positive recurrent Markov chain X with transition probability Π, stationary law μ and initial distribution ν satisfying a minorization condition $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ with known parameters (m, S, δ, Φ) (we take $m = 1$ with regard to simplicity). If we were able to generate $Y_1, ..., Y_n$, so that $X^{\mathcal{M}\,(n)} = ((X_1, Y_1), ..., (X_n, Y_n))$ be a realization of the split chain $X^{\mathcal{M}}$ described in subsection 3.2, then we could apply the RBB procedure to the sample path $X^{\mathcal{M}\,(n)}$. Unfortunately, as will be shown below, knowledge of Π is required to draw practically $Y_1, ..., Y_n$ this way. Our proposal for bootstrapping positive recurrent chains relies on the splitting construction and is based on the knowledge of the parameters of a minorization condition. Precisely, the matter is to approximate this construction by computing an estimator $p_n(x, y)$ of the transition density $p(x, y)$ using data $X_1, ..., X_{n+1}$, and to generate a random vector $(\widehat{Y}_1, ..., \widehat{Y}_n)$ conditionally to $X^{(n+1)} = (X_1, ..., X_{n+1})$, drawn from a 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 $(Y_1, ..., Y_n)$ for given $X^{(n+1)}$. Our method, which we call *approximate regenerative block-bootstrap* (ARBB), amounts then to apply the RBB procedure to the data $((X_1, \widehat{Y}_1), ..., (X_n, \widehat{Y}_n))$ as if they were drawn from the atomic chain $X^{\mathcal{M}}$. Here we will show 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.

4.1 "Approximating the regenerative blocks"

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

$$p(x, y) \geq \delta\phi(y), \ \lambda(dy) \text{ a.s.}$$

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 $\mathcal{M}$. Note that the distribution of $Y^{(n)} = (Y_1, ..., Y_n)$ conditionally to $X^{(n+1)} = (x_1, ..., x_{n+1})$ is the tensor

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

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

with

- if $x_i \notin S$,

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

- if $x_i \in S$,

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

  for $1 \leqslant i \leqslant n$.

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. Our proposition for bootstrapping Harris chains consists thus in approximating this construction by computing an estimate $p_n(x, y)$ of $p(x, y)$ from data $X_1, ..., X_{n+1}$, and then drawing a random vector $(\widehat{Y}_1, ..., \widehat{Y}_n)$ conditionally to $X^{(n+1)} = (X_1, ..., X_{n+1})$, from the distribution $\mathcal{L}^{(n)}(p_n, S, \delta, \phi, X^{(n+1)})$. 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 $p_n(X_i, X_{i+1}) > 0$, $1 \leqslant i \leqslant n$.

In the next subsection, we show that the accuracy of this approximation (in the sense of the Mallows distance) mainly depends on the rate of the uniform convergence of $p_n(x, y)$ to $p(x, y)$ over $S \times S$.

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

Let us show that the distribution of $(X_i, \widehat{Y}_i)_{1 \leqslant i \leqslant n}$ gets closer and closer to the distribution of $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ in the sense of the Mallows distance (also known as the Kantorovich or Wasserstein metric in the probability literature) as n tends to infinity. Hence, we express here the distance between the distributions P^Z and $P^{Z'}$ of two random sequences $Z = (Z_n)_{n \geqslant 0}$ and $Z' =$

$(Z'_n)_{n\geqslant 0}$, taking their values in $\mathbb{R}^k$, by (see Rachev & Rüschendorf (1998), p 76)

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

with

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

where D denotes the metric on the space $\chi(\mathbb{R}^k) = (\mathbb{R}^k)^\infty$ defined by

$$D\left(w, w'\right) = \sum_{k=0}^{\infty} 2^{-k} \left\|w_k - w'_k\right\|_{\mathbb{R}^k}$$

for any w, w' in $\chi(\mathbb{R}^k)$ ($\|.\|_{\mathbb{R}^k}$ denoting the usual euclidian norm of $\mathbb{R}^k$). Thus, viewing the sequences $Z^{(n)} = (X_k, Y_k)_{1\leqslant k\leqslant n}$ and $\widehat{Z}^{(n)} = (X_k, \widehat{Y}_k)_{1\leqslant i\leqslant n}$ as the beginning segments of infinite series, we evaluate the deviation between the distribution $P^{(n)}$ of $Z^{(n)}$ and the distribution $\widehat{P}^{(n)}$ of $\widehat{Z}^{(n)}$ by using this definition

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

Theorem 4.1 *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

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

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 to obtain better finite sample results (see Example 2 in section 5).

4.3 The approximate regenerative block-bootstrap 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 performed in six steps as follows.

1. From the data $X^{(n+1)} = (X_1, ..., X_{n+1})$, 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 $p_n(X_i, X_{i+1}) > 0$, $1 \leqslant i \leqslant n$.

2. Conditionally to $X^{(n+1)}$, draw a vector $(\widehat{Y}_1, ..., \widehat{Y}_n)$ from the distribution estimate $\mathcal{L}^{(n)}(p_n, S, \delta, \phi, X^{(n+1)})$.

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 $A_{\mathcal{M}}$,

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

with

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

4. Draw sequentially bootstrap data blocks $\mathcal{B}_1^*, ..., \mathcal{B}_k^*$ 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 $\widehat{\mathcal{B}}_1, ...,$ $\widehat{\mathcal{B}}_{\widehat{l}_n-1}$, 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 $l_n^* = \inf\{k \geqslant 1,\ l^*(k) > b_n\}$.

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*

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

Then compute the *ARBB statistic*

$$T_{n,b_n}^{*(n)} = T(X_{b_n}^{*(n)})$$

and the *ARBB standardization*

$$S_{n,b_n}^{*(n)} = S(X_{b_n}^{*(n)}).$$

6. The *ARBB distribution* is then given by

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

In the following we choose $b_n = n$.

Remark 4.1 *From a practical point of view, it actually suffices to draw the $\widehat{Y}_i$'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 $\widehat{Y}_i$ according to the Bernoulli distribution with parameter $\delta\phi(X_{i+1})/p_n(X_i, X_{i+1})$).*

Remark 4.2 *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.*

4.4 Asymptotic validity of the ARBB

As explained in Malinovskii (1985, 87) in the unstandardized case, the Edgeworth expansion (E.E. in abbreviated form) proved in Bertail & Clémençon (2003a) 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 $\tau_{A_\mathcal{M}} = \inf\{n \geqslant 1, (X_n, Y_n) \in S \times \{1\}\}$ and $E_{A_\mathcal{M}}(.)$ 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 4.2.1 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 $(X_i, \widehat{Y}_i) \in S \times \{1\}$) generated by the procedure described in 4.1.

$$\widehat{\mu}_n(f) = \widehat{n}_{A_\mathcal{M}}^{-1} \sum_{j=1}^{\widehat{l}_n - 1} f(\widehat{\mathcal{B}}_j),$$

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

of $\mu(f)$ and σ_f^2 respectively, with $\widehat{n}_{A_\mathcal{M}} = \widehat{\tau}_{A_\mathcal{M}}(\widehat{l}_n) - \widehat{\tau}_{A_\mathcal{M}}(1) = \sum_{j=1}^{\widehat{l}_n - 1} l(\widehat{\mathcal{B}}_j)$ and $f(\widehat{\mathcal{B}}_j) = \sum_{i=1+\widehat{\tau}_A(j)}^{\widehat{\tau}_A(j+1)} f(X_i)$. By convention, $\widehat{\mu}_n(f)$ (respectively $\widehat{\sigma}_n^2(f)$,

$\widehat{n}_{A_{\mathcal{M}}}$) equals to 0, when $\widehat{l}_n \leqslant 1$. 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 section 3), 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

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

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

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

with sampling distribution

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

4.4.1 Further assumptions and preliminary results

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

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

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

$\mathcal{H}_1(k, \nu)$: 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 4.3 *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 $n \to \infty$,

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

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

Remark 4.4 *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, 91a, 91b), Rosenblatt (1970), Birgé (1983), Doukhan & Ghindès (1983), Prakasa Rao (1983), Athreya & Atuncar (1998) or Clémençon (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 4.2 *Suppose that conditions (i)-(ii) are fulfilled by the split chain* $X^{\mathcal{M}}$ *for the atom* $A_{\mathcal{M}}$*. 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* $n \to \infty$

$$\widehat{\sigma}_n^2(f) \to \sigma_f^2 \text{ in } P_\nu\text{- probability,}$$
$$\widehat{n}_{A_{\mathcal{M}}}^{1/2}\,\frac{\widehat{\mu}_n(f)-\mu(f)}{\widehat{\sigma}_n(f)} \to \mathcal{N}(0,1) \text{ in distribution under } P_\nu.$$

Remark 4.5 *We recall that conditions (i)-(v) for the split chain may be more easily checked in practice by using test functions methods (refer to 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). 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* $s=\rho$.

4.4.2 Main asymptotic theorem

The bootstrap counterparts of the statistics introduced above are then defined as follows. Let $\mathcal{B}_1^*, ..., \mathcal{B}_{l_n^*-1}^*$ 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

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

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

$$H^U_{ARBB}(x) = P^*(n_{A_{\mathcal{M}}}^{*1/2} \widehat{\sigma}_n^{-1}(f) \, (\mu_n^*(f) - \widehat{\mu}_n(f)) \leq x \mid X^{(n+1)}).$$

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

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

and the studentized ARBB distribution estimate

$$H^S_{ARBB}(x) = P^*(t^*_{A_{\mathcal{M}},n} \leq x \mid X^{(n+1)}).$$

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 4.3 *Under the hypotheses of Theorem 4.2, we have the following convergences in distribution under* P_ν

$$\Delta_n^U = \sup_{x \in \mathbb{R}} |H^U_{ARBB}(x) - H^U_\nu(x)| \to 0, \ \ as \ n \to \infty$$
$$\Delta_n^S = \sup_{x \in \mathbb{R}} |H^S_{ARBB}(x) - H^S_\nu(x)| \to 0, \ \ as \ n \to \infty.$$

Remark 4.6 *In consideration of technical difficulties, we confined here the study of the properties of the ARBB distribution estimate to establishing its asymptotic validity. In particular, second order properties of the ARBB 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 (2003a) 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 $\tau_A(1), ..., \tau_A(l_n)$ up to time n, and 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. This suggests that higher order properties of the ARBB should be studied with other techniques, which will be the subject of further investigation. In spite of this we nevertheless expect that the ARBB properties are comparable to the RBB properties.

5 Illustrative examples

We now give two examples, only with a view to illustrate the scope of applications of our methodology. Simulations showing the performance of the RBB for specific regenerative queuing models are exhibited in Bertail & Clémençon (2003c). 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. 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)).

5.1 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 $X_1, ..., X_n$, the amounts of input $S_1, ..., S_n$ are independent from each other and independent from the inter-arrival times $\Delta T_1, ..., \Delta T_n$ and that the distribution of S_i is given by $K(X_i, .)$, for $0 \leqslant i \leqslant n$. Under the further assumption that $(\Delta T_n)_{n\geqslant 1}$ 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

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

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 $K(x, .)$ has infinite tail for all $x \geqslant 0$. 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 $V : \mathbb{R}_+ \to [0, \infty]$ such that $V(0) < \infty$ and for all $x \geqslant 0$:

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

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 3.1), on which RBB statistics may be based.

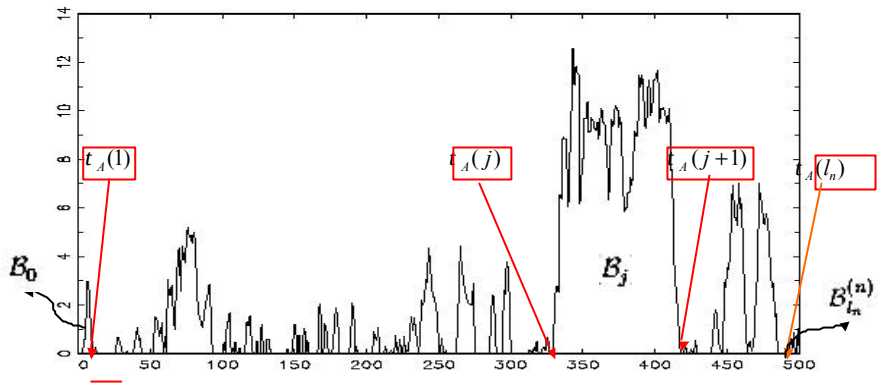


Figure 1: Dividing the trajectory of the storage process into data blocks corresponding to the regeneration times $\tau_A(j)$

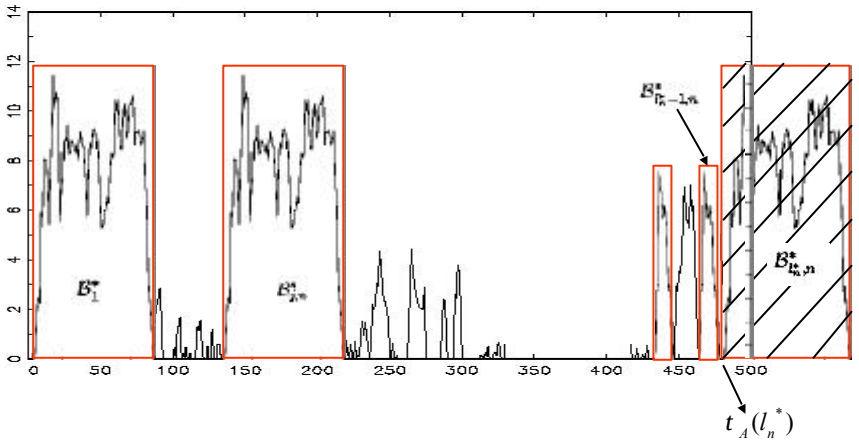


Figure 2: Reconstruction of a storage process data series using the RBB resampling procedure

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 3.2)) and to the gaussian approximation.

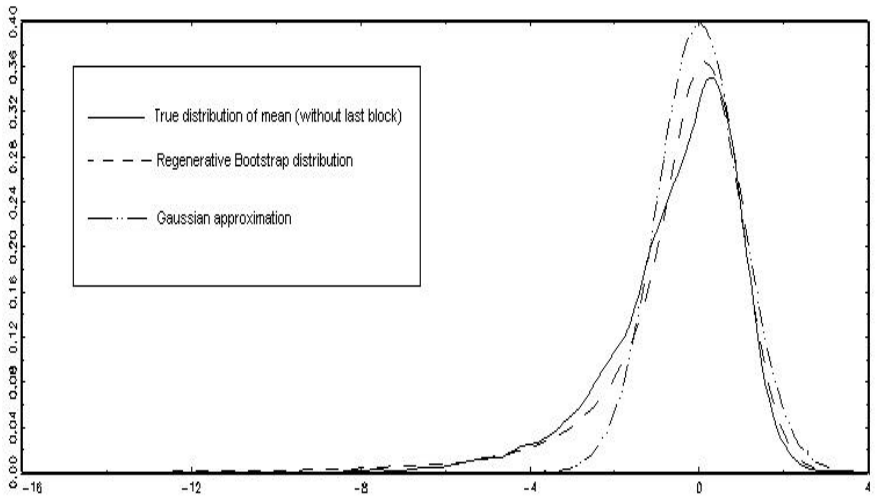


Figure 3: Comparison of true, RBB and gaussian distributions for $n = 200$.

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		∞
γ%	TD	RBB	TD	RBB	ASY	γ%	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
2	-6.179	-5.734	-4.607	-4.695	-2.054	91	1.078	1.085	1.083	1.095	1.341
3	-5.302	-5.014	-4.170	-4.165	-1.881	92	1.125	1.145	1.122	1.150	1.405
4	-4.816	-4.473	-3.708	-3.757	-1.751	93	1.168	1.207	1.177	1.209	1.476
5	-4.374	-4.134	-3.430	-3.477	-1.645	94	1.220	1.276	1.236	1.277	1.555
6	-4.086	-3.853	-3.153	-3.243	-1.555	95	1.287	1.360	1.299	1.356	1.645
7	-3.795	-3.607	-2.966	-3.045	-1.476	96	1.366	1.453	1.380	1.442	1.751
8	-3.576	-3.374	-2.771	-2.866	-1.405	97	1.433	1.568	1.479	1.549	1.881
9	-3.370	-3.157	-2.606	-2.709	-1.341	98	1.540	1.722	1.646	1.685	2.054
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 tails of the true distribution (TD), RBB and gaussian distributions.

5.2 Example 2 : nonlinear AR models

Consider now the general heteroskedastic autoregressive model

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

where $m : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}_+^*$ 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, $k \leqslant n$ 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,

$$\inf_{(x,y)\in V_{x_0}^2} p(x, y) \geqslant \delta(2\varepsilon)^{-1}. \tag{4}$$

Any compact interval $V_{x_0}(\varepsilon)$ is thus a small set for the chain X, which satisfies the minorization condition $\mathcal{M}(1, V_{x_0}(\varepsilon), \delta, \mathcal{U}_{V_{x_0}(\varepsilon)})$, where $\mathcal{U}_{V_{x_0}(\varepsilon)}$ denotes the uniform distribution on $V_{x_0}(\varepsilon)$ (see definition 3.2). Hence, in the case when one knows x_0, ε and δ such that (2) holds (this simply amounts to know 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.

We point out that 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 little prior information about the structure of this chain (*i.e.* about functions m, σ and g) is available, an empirical method for choosing 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 $X^{(n+1)}$, that is

$$N_n(\varepsilon) = E(\sum_{i=1}^{n} I\{X_i \in V_{x_0}(\varepsilon),\ Y_i = 1\} \mid X^{(n+1)}) \tag{5}$$
$$= \frac{\delta(\varepsilon)}{2\varepsilon} \sum_{i=1}^{n} I\{(X_i, X_{i+1}) \in V_{x_0}(\varepsilon)^2\} \frac{1}{p(X_i, X_{i+1})}.$$

Since the transition density p and its minimum over $V_{x_0}(\varepsilon)^2$ are unknown, a practical criterion $\widehat{N}_n(\varepsilon)$ to optimize could be obtained by replacing p by an estimate p_n and $\delta(\varepsilon)/2\varepsilon$ by a sharp lower bound $\widehat{\delta}_n(\varepsilon)/2\varepsilon$ for p_n over $V_{x_0}(\varepsilon)^2$ in expression (5). Properties of such a choice (which gives remarkable results in practice for variance estimation and bootstrap approximation according to our experience) will be studied in more details in a further work. 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.

Simulation results Here are empirical evidences for two specific models.

The $AR(1)$ model :

$$X_{i+1} = \alpha X_i + \varepsilon_{i+1},\ i \in \mathbf{N},$$

with $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\alpha = 0.95$, $X_0 = 0$ and for a trajectory of length $n = 200$.

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

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

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

A much more detailed simulation study as well as applications to different types of 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 classical kernel estimator $\widehat{p}_n$ of Nadaraya-Watson 's type (see Doukhan & Ghindès (1983) for instance). The small set is selected by maximizing the empirical criterion $\widehat{N}_n(\varepsilon)$ 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 $\widehat{N}_n(\varepsilon)$ 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 $\widehat{\delta} = 0.110$ (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.

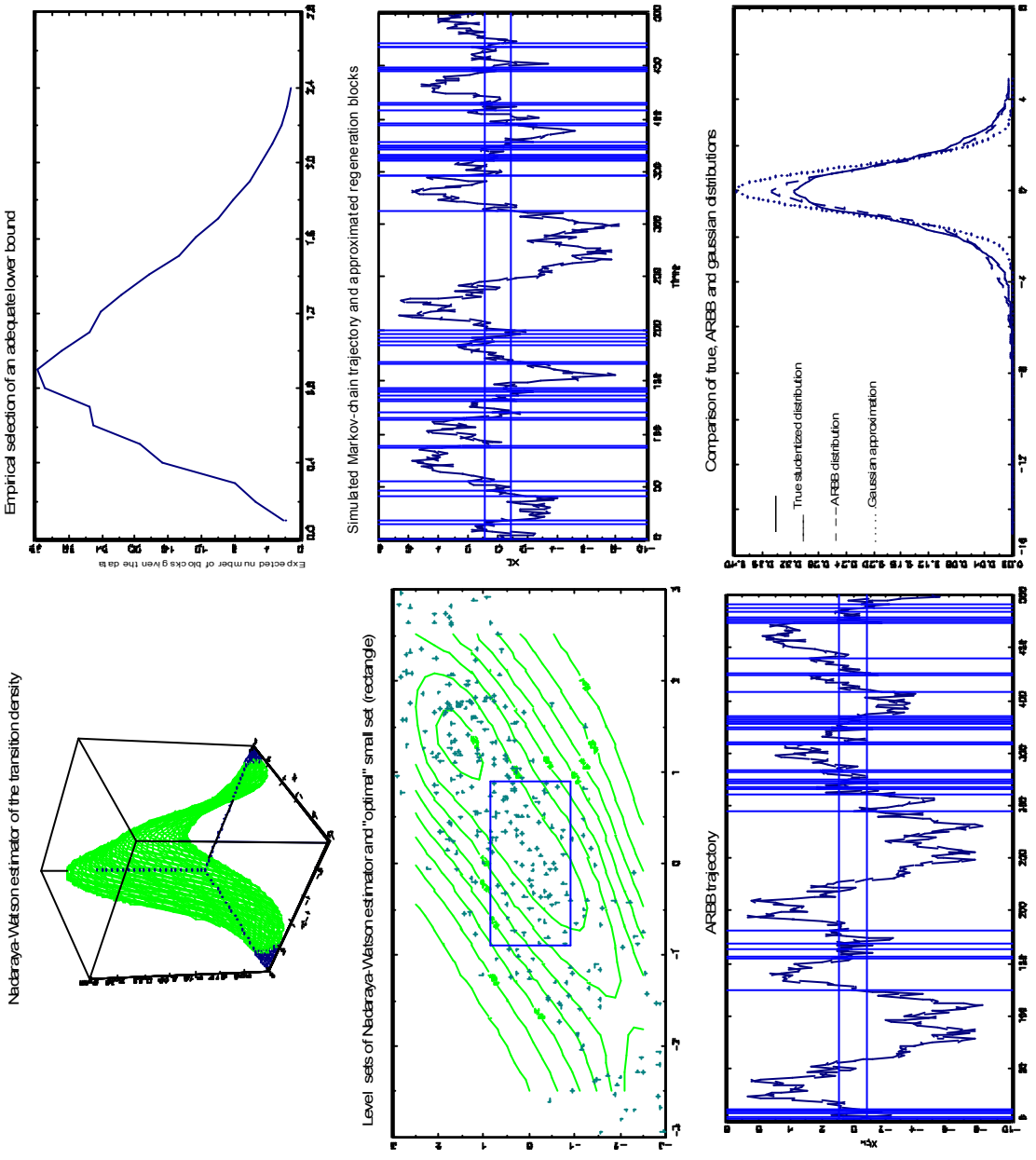
	AR		AR-ARCH		
γ	TD	ARBB	TD	ARBB	ASY
1	-3.639	-3.754	-2.532	-2.683	-2.326
2.5	-2.772	-2.818	-2.025	-2.146	-1.960
5	-2.346	-2.388	-1.793	-1.866	-1.645
10	-1.741	-1.797	-1.429	-1.452	-1.282

	AR		AR-ARCH		
γ	TD	RBB	TD	ARBB	ASY
90	1.683	1.583	1.362	1.416	1.282
95	2.160	1.934	1.732	1.826	1.645
97.5	2.731	2.533	2.004	2.143	1.960
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

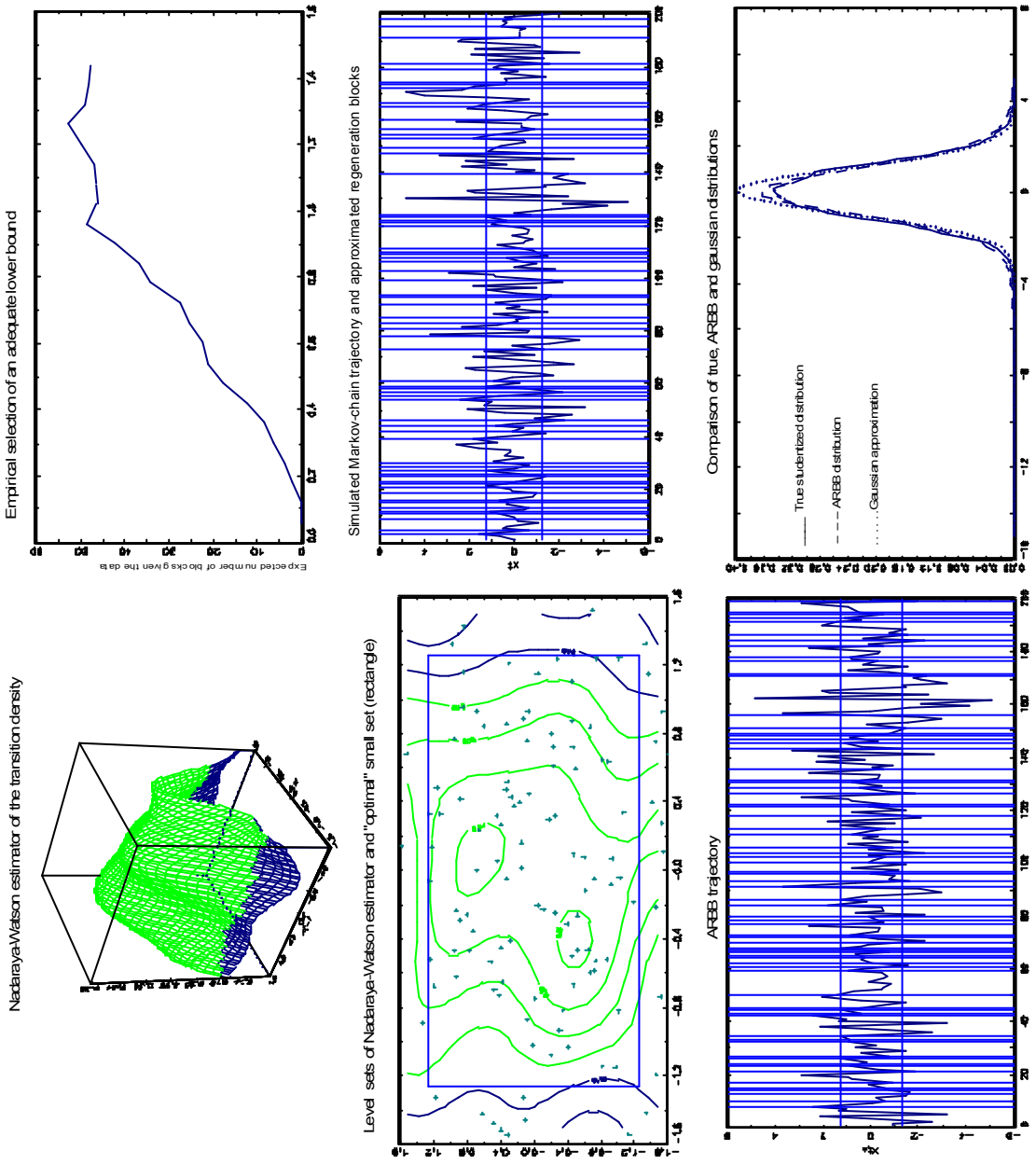
These pictures speak volumes : for both models the true distribution is accurately approximated. But note in particular the difference in the size of the "optimal" small set and in the number of pseudo-regenerations between these models. We point out that, though remarkable when compared to the gaussian 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
$\alpha = 0.2,\ \beta = 0.7,\ n = 200, X_0 = 0$

6 Proofs

6.1 Proof of Theorem 3.1.

In the following, C and K are constants which may be not necessarily the same at each appearance. We denote by $E^*(. \mid X^{(n)})$ 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 $X^{(n)}$). The proof 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 & Clémençon (2003a) (cf Theorem 5.1 therein) respectively are fulfilled for the RBB reconstructed series. 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

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

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

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

and the term at the right hand side almost surely converges to 0 as $n \to \infty$ by virtue of the SLLN ($\mu_n(f) \to \mu(f)$, while $(l_n-1)^{-1}\sum_{j=1}^{l_n-1} l(\mathcal{B}_j) \to E_A(\tau_A)$ as $n \to \infty$). Moreover the SLLN also yields that $|C_n(\mu(f)) - C(t)| \to 0$ *a.s.* as $n \to \infty$. Thus $C_n(t,\mu_n(f))$ converges to $C(t)$ uniformly over any compact set. Since $\overline{lim}_{t\to\infty}C(t) < 1$, 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

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

In particular,

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

as $n \to \infty$.

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

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

Thus for n large enough,

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

It follows from Bertail & Clémençon (2003a) that, as $n \to \infty$,

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

with

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

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

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

with

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

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 $n \to \infty$,

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

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 $n \to \infty$,

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

with

$$\beta = cov(l(\mathcal{B}_j),\ f(\mathcal{B}_j) - \mu(f)l(\mathcal{B}_j)),$$

provided that condition (ii) is fulfilled with $s = 6 + \varepsilon$. Note in addition that, as $n \to \infty$,

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

The proof is then finished by observing that the E.E. of the true distributions (see expressions (8) and (9) in Bertail & Clémençon (2003a)) and the one of the RBB distribution match up to $O_{P_\nu}(n^{-1})$ in the unstandardised case and $O_{P_\nu}(n^{-1}\log(n))$ in the standardized case, as $n \to \infty$.

6.2 Proof of Theorem 4.1

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

$$Y_{\tau_S(k)} \sim Ber\left(\delta\phi\left(X_{\tau_S(k)+1}\right)/p(X_{\tau_S(k)}, X_{\tau_S(k)+1})\right),$$
$$\widehat{Y}_{\tau_S(k)} \sim Ber\left(\delta\phi\left(X_{\tau_S(k)+1}\right)/p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})\right),$$

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

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

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

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

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

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

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

Observe that, we almost surely have

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

Consequently, we may write

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

Hence, under (i), we have

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

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

6.3 Proof of Theorem 4.2

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 $\mathcal{B}_1, ..., \mathcal{B}_{l_n}$ dividing the segment $X^{(n+1)} = (X_1, ..., X_{n+1})$ 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 $\widehat{\mathcal{B}}_1, ..., \widehat{\mathcal{B}}_{\widehat{l}_n}$ 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 6.2. We shall use the notations $l_n = \sum_{i=1}^{n} I\{X_i \in S, Y_i = 1\}$, $\tau_{A_{\mathcal{M}}} = \tau_{A_{\mathcal{M}}}(1) = \inf\{n \geqslant 1, (X_n, Y_n) \in A_{\mathcal{M}}\}$, $\tau_{A_{\mathcal{M}}}(j+1) = \inf\{n > \tau_{A_{\mathcal{M}}}(j), (X_n, Y_n) \in A_{\mathcal{M}}\}$, $l(\mathcal{B}_j) = \tau_{A_{\mathcal{M}}}(j) - \tau_{A_{\mathcal{M}}}(j)$ for $j \geqslant 1$. Set $n_{A_{\mathcal{M}}} = \tau_{A_{\mathcal{M}}}(l_n) - \tau_{A_{\mathcal{M}}}(1)$ and let $\mu_n(f) = n_{A_{\mathcal{M}}}^{-1} \sum_{j=1}^{l_n} f(\mathcal{B}_j)$ and $\sigma_n^2(f) = n_{A_{\mathcal{M}}}^{-1} \sum_{j=1}^{l_n} \{f(\mathcal{B}_j) - \mu_n(f) l(\mathcal{B}_j)\}^2$ be the respective counterparts of $\widehat{\mu}_n(f)$ and $\widehat{\sigma}_n^2(f)$ 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, $\widehat{\tau}_1 = \widehat{\tau}_{A_{\mathcal{M}}}$) the (random) time corresponding to the first visit to $S \times \{1\}$ of $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ (resp., of $(X_i, \widehat{Y}_i)_{1 \leqslant i \leqslant n}$) as well as the time $\tau_2 = \tau_{A_{\mathcal{M}}}(l_n)$ (resp., $\widehat{\tau}_2 = \widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)$) corresponding to the last visit.

Lemma 6.1 *Let $\gamma \geqslant 2$. Under $\mathcal{H}_2(2\gamma)$ and $\mathcal{H}_3$, there exists a constant C such that for $i \in \{1, 2\}$,*

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

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 $\delta\phi\left(X_{\tau_S(k)+1}\right) / p(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ and $\delta\phi\left(X_{\tau_S(k)+1}\right) / p_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$ are bounded from below by $q = \delta \inf_{x \in S} \phi(x)/R$. Given the joint distribution of the $(Y_i, \widehat{Y}_i)$'s (refer to 6.2 above) and in particular that

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

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

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

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

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

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 $g : (E, \mathcal{E}) \to \mathbf{R}$ 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 $l_n \leqslant 1$ (resp., when $\widehat{l}_n \leqslant 1$). The following lemma provides an asymptotic bound for

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

with $M = 1, 2$.

Lemma 6.2 *Let $\gamma \geqslant 4$. Under $\mathcal{H}_2(\gamma)$ and $\mathcal{H}_3$, we have as $n \to \infty$*

$$D_n^{(1)}(g) = O_{P_\nu}(n^{-1}\alpha_n^{1/2}), \tag{6}$$

$$D_n^{(2)}(g) = O_{P_\nu}(1), \tag{7}$$

as $n \to \infty$.

Proof. Bound (6) immediately follows from lemma 6.1.
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

$$t_1 = \inf\left\{\tau_S(k);\ \max(\widehat{\tau}_1, \tau_1) \leqslant \tau_S(k) \leqslant \min(\widehat{\tau}_2, \tau_2),\ Y_{\tau_S(k)} \neq \widehat{Y}_{\tau_S(k)}\right\},$$
$$t_{j+1} = \inf\left\{\tau_S(k);\ t_j < \tau_S(k) \leqslant \min(\widehat{\tau}_2, \tau_2),\ Y_{\tau_S(k)} \neq \widehat{Y}_{\tau_S(k)}\right\}$$

with $j = 1, ..., N_n - 1$. And for $1 \leqslant j \leqslant N_n$, 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

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

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 $E_\nu((t_j^{(2)} - t_j)^r) \leqslant c_r(q) \sup_x E_x(\tau_S^r)$ for any $j \geqslant 1$, $r \leqslant \gamma$. By Cauchy-Schwarz 's inequality we have for any deterministic sequence of positive integers m_n,

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

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 6.2, 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

$$\begin{aligned} P(N_n > m_n) &\leqslant e^{-m_n} E(e^{N_n}) \\ &\leqslant e^{-m_n}(1 + q_n e)^n. \end{aligned} \tag{8}$$

Now by choosing $m_n \uparrow \infty$ such that $m_n/n \to 0$ and $n\alpha_n/m_n \to 0$ as $n \to \infty$, we deduce from (8) that $n^{-1} E_\nu(\sum_{j=1}^{N_n}(t_j^{(2)} - t_j)^2) = O(1)$, as $n \to \infty$. By an analogous argument, one shows that $n^{-1} E_\nu(\sum_{j=1}^{N_n}(t_j - t_j^{(1)})^2) = O(1)$, as $n \to \infty$. ■

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 6.3 *Under the assumptions of Theorem 4.2, we have the following convergences in P_ν-probability as $n \to \infty$,*

$$n^{1/2}(\widehat{\mu}_n(f) - \mu_n(f)) \to 0, \tag{9}$$

$$\left|\widehat{l}_n/n - l_n/n\right| \to 0. \tag{10}$$

Moreover we have for $k = 1, 2$

$$n^{-1}\sum_{j=1}^{l_n - 1} l(\mathcal{B}_j)^k - n^{-1}\sum_{j=1}^{\widehat{l}_n - 1} l(\widehat{\mathcal{B}}_j)^k \;\to\; 0, \tag{11}$$

$$n^{-1}\sum_{j=1}^{l_n - 1} f(\mathcal{B}_j)^k - n^{-1}\sum_{j=1}^{\widehat{l}_n - 1} f(\widehat{\mathcal{B}}_j)^k \;\to\; 0, \tag{12}$$

$$n^{-1}\sum_{j=1}^{l_n - 1} l(\mathcal{B}_j) f(\mathcal{B}_j) - n^{-1}\sum_{j=1}^{\widehat{l}_n - 1} l(\widehat{\mathcal{B}}_j) f(\widehat{\mathcal{B}}_j) \;\to\; 0. \tag{13}$$

Proof. From (6) in lemma 6.2 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 $n \to \infty$. Given that $n_{A_{\mathcal{M}}}/n \to 1$ P_ν a.s. as $n \to \infty$, this combined to (6) again with $g = f$ obviously yield (9).

Now observe that

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

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 $k = 1, ..., L_n$ and $P(\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)} \mid X^{(n+1)}) \leqslant (\delta \inf_{x \in S} \phi(x))^{-1} \sup_{(x,y) \in S^2} |p_n(x,y) - p(x,y)|$, this entails

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

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

Now, (11) (respectively, (12)) straightforwardly results from lemma 6.2 with $g \equiv 1$ (resp., with $g = f$).

And (13) 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 6.2 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 $\widehat{\sigma}_n^2(f) - \sigma_n^2(f) \to 0$ in P_ν-*pr.*, as $n \to \infty$. Hence, given that $\sigma_n^2(f) - \sigma_f^2$ in P_ν- *pr.*, as $n \to \infty$ (see the preliminary remarks in §3.2.1) the consistency of $\widehat{\sigma}_n^2(f)$ is established.

Finally, combining this to (9) and the CLT for the sample mean $\mu_n(f)$ relating to the atomic split chain (cf §3.2.1) proves that $\widehat{n}_{A_{\mathcal{M}}}^{1/2} \widehat{\sigma}_n(f)^{-1} (\widehat{\mu}_n(f) - \mu(f)) \to \mathcal{N}(0,1)$ in P_ν- distribution, as $n \to \infty$.

6.4 Proof of Theorem 4.3

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

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

as $n \to \infty$.

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 (see for instance 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 independent segments $\widehat{\mathcal{B}}_j^*$, $j \geqslant 1$. By the L.L.N., we have $(l_n-1)^{-1}\sum_{j=1}^{l_n-1} l(\mathcal{B}_j) \to E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})$, P_ν a.s. . Lemma 6.3 thus entails that, as $n \to \infty$,

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

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

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

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

$$\frac{l(\widehat{\mathcal{B}}_j^*)}{n} \xrightarrow{P^*} 0 \text{ and } \frac{n_{A_{\mathcal{M}}}^*}{n} \xrightarrow{P^*} 1 \text{ in } P_\nu\text{- probability along the sample.} \tag{15}$$

Now note that by definition of l_n^* we have

$$\frac{l_n^*}{\sum_{j=1}^{l_n^*} l(\widehat{\mathcal{B}}_j^*)} \leq \frac{l_n^*}{n} \leq \frac{l_n^*}{\sum_{j=1}^{l_n^*-1} l(\widehat{\mathcal{B}}_j^*)}. \tag{16}$$

Combining the L.L.N. to lemma 6.3 we have as $n \to \infty$,

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

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

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

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

$$P^*(n_1 \leq l_n^* - 1 \leq n_2 \mid X^{(n+1)}) \geq 1 - \varepsilon \text{ in } P_\nu\text{- probability} \qquad (17)$$

Combining (15), (17) and Markov's inequality implies that for any $\eta > 0$,

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

for n large enough. Since this is true for any $\eta,\ \varepsilon > 0$, it follows (using again (15)) that

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

along the sample in P_ν- probability, as $n \to \infty$.

Now it is sufficient to apply the classical bootstrap CLT (see Bickel & Freedmann (1981)) to the i.i.d. r.v.'s $\{ f(\mathcal{B}_j^*) - \widehat{\mu}_n(f) l(\mathcal{B}_j^*) \}_{j \geqslant 1.}$ These r.v.'s are centered with variance

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

which converges to $E_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}}) \sigma_f^2$ in P_ν- probability under the hypotheses of Theorem 4.3 (cf Theorem 4.2 and lemma 6.3).

The studentized case We essentially have to prove that, as $n \to \infty$,

$$\sigma_n^*(f) - \widehat{\sigma}_n(f) \xrightarrow{P^*} 0 \text{ in } P_\nu \text{ - } pr. \text{ along the sample.}$$

With arguments similar to those used in the unstudentized case, it is easy to see that

$$\sigma_n^*(f)^2 = \frac{1}{n} \sum_{j=1}^{1+\left\lfloor nE_{A_{\mathcal{M}}}(\tau_A)^{-1} \right\rfloor} \{f(\mathcal{B}_j^*) - \widehat{\mu}_n(f) l(\mathcal{B}_j^*)\}^2 + o_{P^*}(1)$$

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

References

[1] Asmussen, S. (1987). *Applied Probabilities and Queues.* Wiley.

[2] Athreya, K.B., Atuncar, G.S. (1998). Kernel estimation for real-valued Markov chains. *Sankhya*, **60**, series A, No 1, 1-17.

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

[4] Athreya, K.B., Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.,* **245**, 493-501.

[5] 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.

[6] Bertail, P., Clémençon, S. (2003a). Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. *Submitted.*

[7] Bertail, P., Clémençon, S. (2003b). Note on the regeneration-based bootstrap for atomic Markov chains. *Submitted.*

[8] Bertail, P. , Clémençon, S. (2003c). Bootstrapping queuing system, with applications to work modulated server. *Submitted.*

[9] Bickel, P.J., Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196-1217.

[10] Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahr. verw. Gebiete,* **65**, 181-237.

[11] Bolthausen, E. (1980). The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains. *Z. Wahr. Verw. Gebiete*, **54**, 59-73.

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

[13] Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.,* **16**, 1709-1722.

[14] Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincar. Statist. Probab.,* **XIX**, 281-296.

[15] Brockwell, P.J., Resnick, S.J., Tweedie, R.L. (1982). Storage processes with general release rules and additive inputs. *Adv. Appl. Probab.,* **14**, 392-433.

[16] Browne, S., Sigman, K. (1992). Work-modulated queues with applications to storage processes. *J. Appl. Probab.,* **29**, 699-712.

[17] Bühlmann, P. (1997). Sieve Bootstrap for time series. *Bernoulli,* **3**, 123-148.

[18] Bühlmann, P. (2002). Bootstrap for time series. *Statistical Science*, **17**, 52-72.

[19] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, **14**, 1171-1179.

[20] Clémençon, S. (2000). Adaptive estimation of the transition density of a regular Markov chain. *Math. Meth. Stat.,* **9**, No. 4, 323-357.

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

[22] Datta, S., McCormick W.P. (1993). Regeneration-based bootstrap for Markov chains. *Canadian J. Statist.,* **21**, No.2, 181-193.

[23] Datta, S., McCormick, W. P. (1995). Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *Journal of Multivariate Analysis*, **52**, 83-106.

[24] Doukhan, P., Ghindès, M. (1983). Estimation de la transition de probabilité d'une chaîne de Markov Doeblin récurrente. *Stochastic Process. Appl.,* **15**, 271-293.

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

[26] Feller, W. (1968). *An Introduction to Probability Theory and its Applications: vol. I.* John Wiley & Sons, NY, 2nd edition.

[27] Feller, W. (1971). *An Introduction to Probability Theory and its Applications: vol. II.* John Wiley & Sons, NY, 3rd edition

[28] Götze, F., Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zeit. Wahrschein. verw. Geb.*, **64**, 211-239.

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

[30] Hall, P. (1985). Resampling a coverage pattern. *Stoch. Process. Applic.*, **20**, 231-246.

[31] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer.

[32] Harrison, J.M., Resnick, S.J. (1976). The stationary distribution and first exit probabilities of a storage process with general release rule. *Math. Oper. Res.,* **1**, 347-358.

[33] Horowitz, J. (2002) Bootstrap methods for Markov processes. To appear in *Econometrica.*

[34] Jain, J., Jamison, B. (1967). Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.,* **8**, 19-40.

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

[36] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.,* **17**, 1217-1241.

[37] Lahiri, S.N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and non stationary data. In *Exploring The Limits of the Bootstrap.* Ed. Le Page, R., and Billard, L., John Wiley, NY.

[38] Liu R., Singh K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring The Limits of The Bootstrap.* Ed. Le Page R and Billard, L., John Wiley, NY.

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

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

[41] Malinovskii, V. K. (1989). Limit theorems for Harris Markov chains II. *Theory Prob. Appl.*, **34**, 252-265.

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

[43] Nummelin, E. (1978). A splitting technique for Harris recurrent chains. *Z. Wahrsch. Verw. Gebiete,* **43**, 309-318.

[44] Nummelin, E. (1984). *General irreducible Markov chains and non negative operators.* Cambridge University Press, Cambridge.

[45] Paparoditis, E. and Politis, D.N. (2002). The local bootstrap for Markov processes. *J. Statist. Plan. Infer.*, **108**, 301–328.

[46] Politis, D.N. , Romano, J.P. (1992). A General Resampling Scheme for Triangular Arrays of alpha-mixing Random Variables with Application to the Problem of Spectral Density Estimation, *Ann. Statist.*, **20**, 1985-2007.

[47] Politis, D.N., Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.,* **22**, 2031-2050.

[48] Politis, D.N., Romano, J.P., Wolf, T. (2000). *Subsampling.* Springer Series in Statistics, Springer, N.Y.

[49] Politis, D.N. (2003). The impact of bootstrap methods on time series analysis. To appear in *Statistical Science*.

[50] Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation.* Academic Press, NY.

[51] Rachev, S. T., Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. II: Applications.* Springer.

[52] Rajarshi, M.B. (1990). Bootstrap in Markov sequences based on estimates of transition density. *Ann. Inst. Stat. Math.*, **42**, 253-268.

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

[54] Roberts, G.O., Rosenthal, J.S. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. *Electr. Journ. Prob.*, No 9, 1-21.

[55] Rosenblatt, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference,* Ed. M. Puri, 199-210.

[56] Roussas, G. (1969). Nonparametric Estimation in Markov Processes. *Ann. Inst. Stat. Math.*, 73-87.

[57] Roussas, G. (1991a). Estimation of transition distribution function and its quantiles in Markov Processes. In *Nonparametric Functional Estimation and Related Topics*, Ed. G. Roussas, 443-462.

[58] Roussas, G. (1991b). Recursive estimation of the transition distribution function of a Markov Process. *Stat. Probab. Letters*, **11**, 435-447.

[59] Shao, J., Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, NY.

[60] Smith, W. L. (1955). Regenerative stochastic processes. *Proc. Royal Stat. Soc., A,* **232**, 6-31.

[61] Thorisson, H. (2000). *Coupling, Stationarity and Regeneration.* Springer.