Edgeworth expansions of suitably normalized sample mean statistics for atomic Markov chains

Patrice Bertail
Laboratoire de Statistiques CREST

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

Abstract

This paper is devoted to the problem of estimating functionals of type $\mu(f) = \int f d\mu$ from observations drawn from a positive recurrent atomic Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with stationary distribution μ. The properties of different estimators are studied. Beyond an accurate estimation of their bias, the estimation of their asymptotic variance is considered. We also show that the results of Malinovskii (1987) on the validity of the formal Edgeworth expansion for sample mean statistics of type $T_n = n^{-1} \sum_{i=1}^{n} f(X_i)$ extend to their studentized versions, normalized by the asymptotic variance estimates we consider.

Résumé

Cet article est consacré au problème de l'estimation d'une fonctionnelle linéaire $\mu(f) = \int f d\mu$ à partir de l'observation d'une chaîne de Markov récurrente positive $X = (X_n)_{n \in \mathbb{N}}$ possédant un atome accessible et de distribution stationnaire μ. Les propriétés de plusieurs estimateurs sont étudiées. Au delà d'une estimation précise de leurs biais respectifs, nous nous intéressons également à l'estimation de la variance asymptotique de ces estimateurs. Nous montrons aussi que les résultats de Malinovskii (1987) concernant le développement d'Edgeworth de l'estimateur $T_n = n^{-1} \sum_{i=1}^{n} f(X_i)$ s'étendent à la version studentisée, lorsqu'il est normalisé par l'estimateur de la variance asymptotique que nous proposons.

1 Introduction

In Malinovskii (1987) the validity of the Edgeworth expansion has been established for a sample mean statistic $T_n = n^{-1}\sum_{i=1}^{n} f(X_i)$ of a Harris recurrent Markov chain X under very general conditions. The main limitation for exploiting these asymptotic results is of practical nature. As a matter of fact, a practical use of these results, for constructing asymptotic confidence intervals for instance, requires the knowledge of the asymptotic variance, which is used to standardize the sample mean. Therefore, the asymptotic variance is generally unknown in practice and must be estimated. In the setting of Markov chains with a known accessible atom (which includes the whole case of Markov chains with a countable state space, as well as numerous specific chains widely used in queuing/storage models) we study in the present paper a specific estimator of the asymptotic variance and show the validity of the Edgeworth expansion for studentized sample mean statistics, when normalized by this estimator. The construction of the estimator relies on a practical use of the so-called *regenerative method*, which consists, in the case when the chain possesses an accessible atom, in dividing the trajectory of the chain into i.i.d. blocks of observations (namely, *regeneration cycles*) corresponding to the successive visits to the atom. As in Malinovskii (1987), the proof of the asymptotic results is also based on the regenerative technique. Beyond the legitimate investigation of the normal approximation for studentized statistics, it is noteworthy that the arguments put forward in this paper are crucial to show the gain in accuracy provided by specific regeneration-based block bootstrap methods for Markov chains (see Datta & McCormick (1993a) and Bertail & Clémençon (2003a, b)).

This paper is organized as follows. In section 2, notations, as well as the assumptions needed in the next sections, are set up. In section 3, we first consider the problem of estimating functionals of type $\mu(f) = \int f d\mu$ from a realization $X_1, ..., X_n$ of an atomic Markov chain X with stationary probability measure μ. In the nonstationary case, we give an accurate estimation for the bias of estimators constructed by suitable truncations of the sample mean statistic $\widehat{\mu}_n(f) = n^{-1}\sum_{i=1}^{n} f(X_i)$. In the case when the chain possesses a known Harris recurrent atom, an estimate of the asymptotic variance of the sample mean statistic is exhibited in section 4, and an asymptotic bound of its bias is also given. In section 5 a specific way of studentization of the sample mean statistic based on this estimate (which we call *regeneration-based studentization*) is considered. The validity of the Edgeworth expansion is shown for this studentized version of the sample mean. Proofs are given in section 6.

2 Assumptions and notation

Throughout this paper, we consider a time-homogeneous Harris recurrent Markov chain $X = (X_n)_{n\in\aleph}$ valued in a countably generated measurable space $(E, \mathcal{E})$ with transition probability $\Pi(x, dy)$ and stationary distribution $\mu(dy)$ (refer to Revuz (1984) or Chung (1967) for the basic concepts of the Markov chain theory). For any probability distribution ν on $(E, \mathcal{E})$ (respectively, for any $x \in E$) we denote by P_ν (resp., P_x) the probability on the underlying space such that $X_0 \sim \nu$, (resp., $X_0 = x$) and by $E_\nu(.)$ (resp., $E_x(.)$) the P_ν-expectation (resp., the P_x-expectation).

For any subset $C \in \mathcal{E}$, we denote the successive *return times* to C by

$$\tau_C = \tau_C(1) = \inf\{n \geqslant 1,\ X_n \in C\},$$
$$\tau_C(j+1) = \inf\{n \geqslant 1 + \tau_C(j),\ X_n \in C\},\ \text{for } j \geqslant 0.$$

The initial distribution of the chain will be denoted by ν and $1_{\mathcal{A}}$ will denote the indicator function of the event $\mathcal{A}$.

In the present paper we assume that the chain X possesses a known accessible atom A, that is to say a subset $A \in \mathcal{E}$ such that for all x, y in A, $\Pi(x, .) = \Pi(y, .)$ and $\mu(A) > 0$. We denote by P_A (respectively, $E_A(.)$) the probability on the underlying space such that $X_0 \in A$ (resp., the P_A-expectation). In this setting, the stationary distribution μ may be represented as an occupation measure. By virtue of Kac's theorem (see Theorem 10.2.2 in Meyn & Tweedie (1996)), we have:

$$\mu(B) = E_A(\tau_A)^{-1} E_A(\sum_{i=1}^{\tau_A} 1_{\{X_i \in B\}}),\ \text{for any } B \in \mathcal{E}.$$

The main step in the application of the regenerative method for investigating the asymptotic properties of such an atomic chain consists in dividing the sample paths of the chain into "blocks" corresponding to consecutive visits to the atom:

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

The strong Markov property implies that the blocks $\mathcal{B}_j$ are i.i.d. random variables valued in the torus $T = \cup_{n=1}^{\infty} E^n$.

Beyond the case of a Markov chain with a countable state space, for which any recurrent state is an atom, it is noteworthy that many specific atomic Markov chains are widely used in the applications, especially in the area of operations research for modeling storage and queuing systems (refer to Asmussen (1987) for an exhaustive overview). We give below an example of such a Markov chain, which is a refinement of the classical *GI/G/1* queuing model (see Browne & Sigman (1992)).

Example 2.1 *(Work-modulated single server queue) Consider a general single server queuing model, evolving through the random arrival customers and the service times they bring: there is one server and customers are served in order of arrival. Denote by* $(T_n)_{n\in\aleph}$ *the sequence of arrival times of customers into the service operation (by convention the first customer arrives at time* $T_0 = 0$*) and by* $(\tau_n)_{n\in\aleph}$ *the sequence of end of service times. Hence the* n^{th} *customer arrives at time* T_n *and leaves at time* τ_n*). If* W_n *denotes the time he has to wait before he begins being served, we have* $W_0 = 0$ *and*

$$W_{n+1} = (W_n + \Delta\tau_n - \Delta T_{n+1})_+,$$

for all $n \in \aleph$*, with* $(x)_+ = \max(x, 0)$*,* $\Delta\tau_n = \tau_n - \tau_{n-1}$ *and* $\Delta T_n = T_n - T_{n-1}$*. Let* $K(w, dx)$ *be a transition probability kernel on* $\Re_+$*. Assume that, conditionally to* $W_1, ..., W_n$*, the service times* $\Delta\tau_1, ..., \Delta\tau_n$ *are independent from each other and independent from the interarrival times* $\Delta T_1, ..., \Delta T_n$ *and the distribution of* $\Delta\tau_i$ *is given by* $K(W_i, .)$ *for* $1 \leqslant i \leqslant n$*. Then, assuming further that* $(\Delta T_n)_{n\in\aleph}$ *is an i.i.d. sequence with common distribution* G*, independent from* $W = (W_n)_{n\in\aleph}$*, the waiting time process* W *is a Markov chain with transition probability* Π *given by*

$$\Pi(W_n,\ \{0\}) = \Gamma(W_n,\ [W_n, \infty[),$$
$$\Pi(W_n,\ ]w,\ \infty[) = \Gamma(W_n,\ ] - \infty,\ W_n - w[),$$

for any $w > 0$*, where* $\Gamma = G * \breve{K}$ *is the convolution product between* G *and the transition kernel* $\breve{K}$ *image of* K *by the mapping* $x \longmapsto -x$*. The study of the stochastic stability is made easy when the atom* $\{0\}$ *is accessible. One shows that* W *is* δ_0*-irreducible as soon as* $K(w, .)$ *has infinite tail for all* $w \geqslant 0$*. In this case, the chain is positive recurrent if and only if there exist a test function* $V : \Re_+ \to [0, \infty]$ *such that* $V(0) < \infty$ *and* $b > 0$ *such that*

$$\int \Pi(x, dw)V(w) - V(x) \leqslant -1 + b1_{\{x=0\}},$$

for any $x \geqslant 0$ (refer to Meyn & Tweedie (1992) for further detail).

3 On estimating the mean

Let $X^{(n)} = (X_1, ..., X_n)$ be a realization of length n of the chain X. We consider the problem of estimating a functional of type $\mu(f) = \int f(x)\mu(dx)$, where f is a μ-integrable real valued function defined on the state space $(E, \mathcal{E})$ (note that $\mu(f) = \mu(A)E_A(\sum_{i=1}^{\tau_A} f(X_i))$, cf section 2). A simple and natural estimator of $\mu(f)$ is the empirical estimator $\widehat{\mu}_n(f) = n^{-1}\sum_{i=1}^{n} f(X_i)$. By virtue of the LLN for additive functionals of a positive recurrent Markov chain (refer to Theorem 17.1.7 in Meyn & Tweedie (1996) for instance), this estimator is strongly consistent as soon as the initial distribution ν fulfills the regularity condition

$$P_\nu(\tau_A < \infty) = 1.$$

Remark 3.1 *By the representation of the stationary distribution μ using the atom, one may show that in the stationary case, this condition is always fulfilled since $P_\mu(\tau_A = k) = \mu(A)P_A(\tau_A \geqslant k)$.*

Whereas the estimator $\widehat{\mu}_n(f)$ is zero-bias when the chain is stationary, its bias is significant in all other cases. In Malinovskii (1985) (see also Theorem 3 in Malinovskii (1987)) an accurate evaluation of the first order term in the bias of the sample mean $\widehat{\mu}_n(f)$ is given, when the starting distribution is not the stationary one.

Proposition 3.1 *Let $f : (E, \mathcal{E}) \to \Re$ be a measurable function and ν be a probability distribution on $(E, \mathcal{E})$. Let us suppose that the following "block" moment conditions are satisfied*

$$E_A((\sum_{i=1}^{\tau_A} |f(X_i)|)^4) < \infty, \; E_A(\tau_A^4) < \infty,$$
$$E_\nu((\sum_{i=1}^{\tau_A} |f(X_i)|^2)) < \infty, \; E_\nu(\tau_A^2) < \infty,$$

as well as the Cramer condition $\overline{\lim}_{t\to\infty}|E_A(\exp(it\sum_{i=1}^{\tau_A} f(X_i)))| < 1$. *Define*

$$\alpha = E_A(\tau_A),\ \beta = E_A(\tau_A \sum_{i=1}^{\tau_A}\{f(X_i) - \mu(f)\}),$$
$$\varphi_\nu = E_\nu(\sum_{i=1}^{\tau_A}\{f(X_i) - \mu(f)\}),\ \gamma = \alpha^{-1}E_A(\sum_{i=1}^{\tau_A}(\tau_A - i)\{f(X_i) - \mu(f)\}).$$

Then, we have as $n \to \infty$

$$E_\nu(\widehat{\mu}_n(f)) = \mu(f) + (\varphi_\nu + \gamma - \beta/\alpha)n^{-1} + O(n^{-3/2}). \tag{1}$$

Define also the sample mean based on the observations (eventually) collected after the first regeneration time only by

$$\widetilde{\mu}_n(f) = (n - \tau_A)^{-1} \sum_{i=1+\tau_A}^{n} f(X_i)$$

with the convention $\widetilde{\mu}_n(f) = 0$, *when* $\tau_A > n$, *as well as the sample mean based on the observations collected between the first and last regeneration times before* n *by*

$$\overline{\mu}_n(f) = (\tau_A(l_n) - \tau_A)^{-1} \sum_{i=1+\tau_A}^{\tau_A(l_n)} f(X_i)$$

with $l_n = \sum_{i=1}^{n} 1_A(X_i)$ *and the convention* $\overline{\mu}_n(f) = 0$, *when* $l_n \leqslant 1$. *We have, as* $n \to \infty$

$$E_\nu(\widetilde{\mu}_n(f)) = \mu(f) + (\gamma - \beta/\alpha)n^{-1} + O(n^{-3/2}), \tag{2}$$
$$E_\nu(\overline{\mu}_n(f)) = \mu(f) - (\beta/\alpha)n^{-1} + O(n^{-3/2}). \tag{3}$$

Remark 3.2 *We recall that "block" moment conditions may be classically replaced by drift criteria of Lyapounov's type, which often appear as more tractable in practice. One may refer to chapter 11 in Meyn & Tweedie (1996) for further details about such conditions as well as many examples.*

This result points out that, by using the data collected from to the first visit to the atom A only, one eliminates the only quantity depending on the initial distribution ν in the first order term of the bias (more precisely, the term φ_ν is induced by the component $\sum_{i=1}^{\tau_A} f(X_i)$ of the sum, while γ is induced by $\sum_{i=1+\tau_A(l_n)}^{n} f(X_i)$). This observation is crucial, when the matter is to approximate the sampling distribution of such statistics by using Bootstrap procedures in a nonstationary setting. Given the impossibility to approximate the distribution of the "first block sum" $\sum_{i=1}^{\tau_A} f(X_i)$ from one single realization of the chain starting from ν, it is thus better to use the estimators $\widetilde{\mu}_n(f)$ or $\overline{\mu}_n(f)$ than $\widehat{\mu}_n(f)$ in practice: for these estimators, it is actually possible to implement specific Bootstrap methodologies, in order to construct second order correct confidence intervals for instance (see Bertail & Clémençon (2003a, b)). We also emphasize that other consistent estimates may be considered, such as

$$\mu_n^*(f) = n^{-1} \sum_{i=1+\tau_A}^{n} f(X_i),$$

with the usual convention regarding empty summation. But unfortunately, as an elementary calculation shows, the latter estimator does not keep the property regarding to the first order term in the bias mentioned above in the nonstationary case. The proof of (2) and (3) goes exactly along the same lines as the proof of (1) in Malinovskii (1985) and is thus omitted.

4 Estimation of the asymptotic variance of the sample mean statistic

Beyond strong consistency, sample mean statistics may be shown to be asymptotically normal in some cases, since it is proved that the CLT holds, under specific moment conditions, for additive functionals of type $\sum f(X_i)$.

Condition 4.1 *(CLT Moment condition for f and ν) The Markov chain X is such that*

$$E_A(\tau_A^2) < \infty, \; E_\nu(\tau_A) < \infty$$

and

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

Remark 4.1 *Note that these conditions do not depend on the accessible atom chosen.*

We have the following result (see Theorem 17.2.2 in Meyn & Tweedie (1996) for instance).

Theorem 4.2 *If the Markov chain X fulfills the CLT Moment Condition for f and ν , then we have the convergence in distribution under P_ν :*

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

with a normalizing constant

$$\sigma^2(f)=\mu(A)\,E_A((\sum_{i=1}^{\tau_A}\{f(X_i)-\mu(f)\})^2),$$

for $\mu_n(f)$ being any of the three estimates $\widehat{\mu}_n(f)$, $\widetilde{\mu}_n(f)$ or $\overline{\mu}_n(f)$.

Remark 4.2 *It is noteworthy that the asymptotic variance $\sigma^2(f)$ differs from the variance of $f(X_i)$ under the stationary distribution (except in the i.i.d. case, which corresponds to the case when the whole state space is an atom), that is equal to $var_\mu(f)=\mu(A)\,E_A(\sum_{i=1}^{\tau_A}\{f(X_i)-\mu(f)\}^2)$.*

We now address the problem of estimating the asymptotic variance from the observations $X_1,...,X_n$. Let us consider the number of visits to the atom A between time 0 and time n, $l_n=\sum_{i=0}^{n}1_A(X_i)$, and form the l_n-1 blocks

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

when $l_n>1$. We set for $1\leqslant j\leqslant l_n-1$, $f(\mathcal{B}_j)=\sum_{i=1+\tau_A(j)}^{\tau_A(j+1)}f(X_i)$. From the expression of the asymptotic variance

$$\sigma^2(f)=\mu(A)\,E_A((\sum_{i=1}^{\tau_A}f(X_i)-\mu(f)\tau_A)^2),$$

we propose the following estimators of $\sigma^2(f)$, adopting the usual convention regarding to empty summation,

$$\sigma_n^2(f)=n^{-1}\sum_{j=1}^{l_n-1}(f(\mathcal{B}_j)-\overline{\mu}_n(f)s_j)^2, \tag{4}$$

where $s_1 = \tau_A(2) - \tau_A(1), ..., s_{l_n - 1} = \tau_A(l_n) - \tau_A(l_n - 1)$ denote the lengths of the blocks dividing the trajectory. Observe that this estimator is independent from the observations collected before the first visit to A and after the last one before time n. Whereas it is all the same from the estimation point of view, whether $\overline{\mu}_n$ is replaced by $\widehat{\mu}_n$ or $\widetilde{\mu}_n$ in (4) and the blocks sums $f(\mathcal{B}_0) = \sum_{i=1}^{\tau_A} f(X_i)$ and $f(\mathcal{B}_{n,l_n}) = \sum_{i=1+\tau_A(l_n)}^{n} f(X_i)$ are used in the computation of the estimate or not, it will make much easier the calculation in the forthcoming Edgeworth expansion.

Recall that $l_n \to \infty$ a.s. and $l_n/n \to \mu(A)$ a.s. as $n \to \infty$. Hence, when the CLT Moment Condition is fulfilled, a straightforward application of the LLN shows that this estimator is strongly consistent under P_ν.

Proposition 4.3 *(Strong consistency) If X fulfills the CLT Moment Condition for f and ν, then we have as $n \to \infty$, $\sigma_n^2(f) \to \sigma^2(f)$, P_ν a.s.*

Remark 4.3 *In the case of a general irreducible chain X with a transition kernel $\Pi(x, dy)$ satisfying a minorization condition*

$$\forall x \in S,\ \Pi(x, dy) \geqslant \delta\psi(dy),$$

for an accessible measurable set S, a probability measure ψ and $\delta \in]0,1[$ (note that such a minorization condition always holds for Π or an iterate when the chain is irreducible), an atomic extension (X, Y) of the chain may be explicitly constructed by the Nummelin splitting technique (see Nummelin (1984)) from the parameters (S, δ, ψ) and the transition probability Π. In Bertail & Clémençon (2003b), a full methodology based on the simulation of a sequence $(X_1, Y_1^*), ..., (X_n, Y_n^*)$ with a distribution approximating in some sense the one of the regenerative extension (X, Y) from the parameters (S, δ, Φ), the original observation segment $X_1, ..., X_n$ and an estimate of the transition kernel Π based on the latter, has been developed. This allows to extend a specific Bootstrap procedure (namely the Regenerative Block-Bootstrap, see Bertail & Clémençon (2003b)) for Markov chains with a known atom to the case of irreducible chains. It is likely that such a methodology could be applied successfully to the problem of asymptotic variance estimation, so as to *extend the statistical procedure described above to the much more general case of positive recurrent Markov chains. This goes beyond the scope of the present paper, but will be the subject of further research.*

The result below gives an order of magnitude of the bias of this estimator. The Cramer conditions appearing (which will not be assumed later) are maybe not necessary but make the proof easier.

Proposition 4.4 *Let $f : (E, \mathcal{E}) \to \Re$ be a measurable function and ν a probability distribution on $(E, \mathcal{E})$. In addition to conditions of Proposition 5, assume that the Markov chain X fulfills the "block" moment conditions*

$$E_A((\sum_{i=1}^{\tau_A} |f(X_i)|)^{6+\varepsilon}) < \infty, \ E_A(\tau_A^{6+\varepsilon}) < \infty$$

for some $\varepsilon > 0$, as well as the Cramer conditions

$$\overline{\lim_{t\to\infty}} \left| E_A \exp\left\{ it(\sum_{i=1}^{\tau_A}(f(X_i) - \mu(f))^2 \right\} \right| < 1,$$
$$\overline{\lim_{t\to\infty}} \left| E_A(\exp\left\{ it\tau_A \sum_{i=1}^{\tau_A}(f(X_i) - \mu(f)) \right\} \right| < 1.$$

Then, we have

$$E_\nu(\sigma_n^2(f)) = \sigma^2(f) + O(n^{-1}), \ \text{as } n \to \infty.$$

Remark 4.4 *We mention that a precise study of $n(E_\nu(\sigma_n^2(f)) - \sigma^2(f))$, as $n \to \infty$, could be carried out, if one first establishes a non uniform limit theorem for U-statistics of m-lattice i.i.d. random vectors, similar to the result established in Dubinskaite (1982) for sample mean statistics of m-lattice i.i.d. random vectors (extensively used in Malinovskii (1985) and in our proof to derive the expansion* (1)*). This will be dealt with in further investigation.*

We emphasize that in a non i.i.d. setting, it is generally difficult to construct an accurate (positive) estimator of the asymptotic variance. When no structural assumption, except stationarity and square integrability, is made on the underlying process X, a possible method, currently used in practice, is based on so-called *blocking techniques*. Indeed under some appropriate mixing conditions (which ensure that the following series converge), it can be shown that the variance of $n^{-1/2}\widehat{\mu}_n(f)$ may be written $var(n^{-1/2}\widehat{\mu}_n(f)) = \Gamma(0) + 2\sum_{t=1}^{n}(1 - t/n)\Gamma(t)$ and converges to $\sigma^2(f) = \sum_{t=\infty}^{\infty}\Gamma(t) = 2\pi g(0)$, where $g(w) = (2\pi)^{-1}\sum_{t=-\infty}^{\infty}\Gamma(t)\cos(wt)$ and $(\Gamma(t))_{t\geqslant 0}$ denote respectively the spectral density and the autocovariance sequence of the discrete-time stationary process X. Most of the estimators of $\sigma^2(f)$ that have been proposed in the literature (such as the Bartlett spectral density estimator, the moving-block jackknife/subsampling variance estimator, the overlapping or non-overlapping batch means estimator) may be seen as variants of the basic *moving-block bootstrap estimator* (see Künsch (1989))

$$\hat{\sigma}^2_{M,n} = \frac{M}{Q}\sum_{i=1}^{Q}(\overline{\mu}_{i,M,L} - \mu_n(f))^2, \tag{5}$$

where $\overline{\mu}_{i,M,L} = M^{-1}\sum_{t=L(i-1)+1}^{L(i-1)+M} f(X_t)$ is the mean of f on the i-th data block $(X_{L(i-1)+1}, \ldots, X_{L(i-1)+M})$. Here, the size M of the blocks and the amount L of 'lag' or overlap between each block are deterministic (eventually depending on n) and $Q = [\frac{n-M}{L}] + 1$, denoting by $[\cdot]$ the integer part, is the number of blocks that may be constructed from the sample $X_1, ..., X_n$. In the case when $L = M$, there is no overlap between block i and block $i+1$ (as the original solution considered by Carlstein (1985)), whereas the case $L = 1$ corresponds to maximum overlap (see Politis & Romano (1995)). Under suitable regularity conditions (mixing and moments conditions), it can be shown that if $M \to \infty$ with $M/n \to 0$ and $L/M \to a \in [0, 1]$ as $n \to \infty$, then we have

$$E(\hat{\sigma}^2_{M,n}) - \sigma^2(f) = O(1/M) + O(\sqrt{M/n}), \tag{6}$$

$$Var(\hat{\sigma}^2_{M,n}) = 2c\frac{M}{n}\sigma^4(f) + o(M/n), \tag{7}$$

as $n \to \infty$, where c is a constant depending on a, taking its smallest value (namely $c = 2/3$) for $a = 0$. This result shows that the bias of such estimators may be very large. Indeed, by optimizing in M we find the optimal choice $M = n^{1/3}$, for which we have $E(\hat{\sigma}^2_{M,n}) - \sigma^2(f) = O(n^{-1/3})$. Various extrapolation and jackknife techniques or kernel smoothing methods have been suggested to get rid of this large bias (refer to Politis & Romano (1995), Götze & Künsch (1996) and Bertail & Politis (2001)). The latter somehow amount to make use of Rosenblatt smoothing kernels of order higher than two (taking some negative values) for estimating the spectral density at 0.

However, the main drawback in using these estimators is that they take negative values for some n, and lead consequently to face problems, when dealing with studentized statistics.

In our specific Markovian framework, the estimate $\sigma_n^2(f)$ is much more natural and allows to avoid these problems. This is particularly important when the matter is to establish Edgeworth expansions at orders higher than two in such a non i.i.d. setting. As a matter of fact, the bias of the variance may completely cancel the accuracy provided by higher order Edgeworth expansions (but also the one of its Bootstrap approximation) in the studentized case, given its explicit role in such expansions (see Götze & Künsch (1996)). The purpose of the next section is to show that for the particular class of positive recurrent Markov chains with an atom, we can get an Edgeworth expansion with a rate $O_P(log(n)n^{-1})$, close to the optimal rate $O_P(n^{-1})$ that can be obtained in the i.i.d. case, under rather weak assumptions (including nonstationary situations).

5 Edgeworth expansion for the studentized sample mean statistic

According to Proposition 4.3, under the assumption that the CLT Moment Condition is fulfilled, the sample mean statistic $\widehat{\mu}_n(f)$ (respectively $\widetilde{\mu}_n(f)$, $\overline{\mu}_n(f)$), when renormalized by the sequence $\sigma_n^2(f)$, is thus asymptotically pivotal. Now we show that it admits an Edgeworth expansion. The main difficulty in establishing such an expansion arises from the random character of the number of blocks, namely, $l_n - 1$ (note that conditioning on l_n is useless, since, conditionally to l_n, the $f(\mathcal{B}_j)$'s, $1 \leqslant j \leqslant l_n - 1$, are obviously not i.i.d.). Thus, we can not directly apply the results on studentized Edgeworth expansions (see Hall (1987)).

To derive an Edgeworth expansion for the studentized sample mean, we will assume that specific "block" moments and Cramer conditions hold. These hypotheses are stated below in the same spirit as in Malinovskii (1987).

Assume that the chain X fulfills the following conditions.

(i) (*Cramer condition*)

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

(ii) (*Non degenerate asymptotic variance*)

$$\sigma^2(f) > 0.$$

(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*)

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

Note that, in the case when f is bounded, the "block" moment conditions (iii)-(iv) may be obviously replaced by some regularity conditions involving τ_A only (see Clémençon (2001)). Besides, links between conditions of type (iii) above and conditions on the rate of decay of strong mixing coefficients of a noncyclic chain have been studied in Bolthausen (1982).

(vi) (*Boundedness of the density of the block sums*) The density of the $\sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$'s is bounded.

This last condition is mainly technical but is clearly satisfied in many practical situations. In what follows, $\mu_n(f)$ denotes indifferently any of the estimates $\widehat{\mu}_n(f)$, $\widetilde{\mu}_n(f)$ or $\overline{\mu}_n(f)$. We define the standardized sample mean

$$t_n = n^{1/2}\sigma^{-1}(f)(\mu_n(f) - \mu(f)),$$

the studentized sample mean

$$\widetilde{t}_n = n^{1/2}\sigma_n^{-1}(f)(\mu_n(f) - \mu(f)),$$

and the renormalized asymptotic bias

$$b = \lim_{n\to\infty} n\,\sigma^{-1}(f)E_\nu(\mu_n(f)) - \mu(f))$$

which is given in Proposition 3.1, depending on whether $\mu_n(f)$ is equal to $\widehat{\mu}_n(f)$, $\widetilde{\mu}_n(f)$ or $\overline{\mu}_n(f)$. The expansions for these different estimators only differ from one another in the bias term. We are now ready to state our main result.

Theorem 5.1 *Under assumptions (i)-(vi) with* $s = 4$, *the following Edgeworth expansion is valid uniformly over* R,

$$\Delta_n = \sup_{x \in \mathsf{R}} |P_\nu (t_n \leq x) - E_n^{(2)}(x)| = O(n^{-1}) \text{ as } n \to \infty$$

with

$$E_n^{(2)}(x) = \Phi(x) - n^{-1/2}\frac{k_3(f)}{6}(x^2 - 1)\phi(x) - n^{-1/2}b\phi(x), \tag{8}$$

and

$$k_3(f) = \alpha^{-1}(M_{3,A} - \frac{3\beta}{\sigma(f)}),$$

where

$$M_{3,A} = E_A \left\{ \sum_{i=1}^{\tau_A} (f(X_i) - \mu(f)) \right\}^3 / \sigma(f)^3,$$

$\Phi(x)$ *denotes the distribution function of the standard normal distribution and* $\phi(x) = d\Phi(x)/dx$. *A similar result holds for the studentized statistic under (i)-(vi) with* $s = 8 + \varepsilon$, *for some* $\varepsilon > 0$

$$\Delta_n^S = \sup_{x \in \mathsf{R}} |P_\nu(n^{1/2}\sigma_n^{-1}(f)(\mu_n(f) - \mu(f)) \leq x) - F_n^{(2)}(x)| = O(n^{-1}\log(n)), \tag{9}$$

as $n \to \infty$, *with*

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

Remark 5.1 *Note that in the i.i.d. case we may choose* $A = E$ *(so that* $\tau_A = 1$, $\alpha = 1$*) and we have then* $b = 0$. *Hence, the Edgeworth expansion of the studentized sample mean reduces in that case to the well known form* $\Phi(x) + n^{-1/2}\frac{1}{6}k_3(2x^2 + 1)\phi(x)$ *with* $k_3 = E_\mu(\{f(X_i) - \mu(f)\}^3)/\sigma(f)^3$, *given in Hall (1987). Besides, under the hypothesis that the following series converge,* *we have (see Theorem 6 in Malinovskii (1987))*

$$\begin{aligned}\sigma(f)^{-3}k_3(f) &= E_\mu(\widetilde{f}^3(X_i)) \\ &+ 3\sum_{i=1}^{\infty}\{E_\mu\left(\widetilde{f}^2(X_1)\widetilde{f}(X_{i+1})\right) + E_\mu\left(\widetilde{f}(X_1)\widetilde{f}^2(X_{i+1})\right)\} \\ &+ 6\sum_{i=1,j=1}^{\infty} E_\mu\left(\widetilde{f}(X_1)\widetilde{f}(X_{i+1})\widetilde{f}(X_{i+j+1})\right),\end{aligned} \tag{10}$$

where $\widetilde{f} = f - \mu(f)$.

Remark 5.2 *When formulated in terms of decay of strong mixing coefficients, our conditions are weaker than the usual ones, which assume an exponential rate for the decay (see for instance Nagaev (1961), Götze & Hipp (1983), Datta & McCormick (1993b)): condition (iv) with* $s = 8 + \varepsilon$ *is typically fulfilled in the bounded case as soon as the strong mixing coefficients sequence decreases at a polynomial rate* $n^{-\rho}$ *for some* $\rho > 7 + \varepsilon$*. However, the condition* $s = 8 + \varepsilon$ *is clearly not optimal (see Hall(1987) for optimal results in the i.i.d. case) and is technically required because we proceed in the proof by conditioning firstly on the variance estimate: it seems reasonable to expect that the result actually holds when condition (iv) is satisfied for some* $s > 4$*, as in the i.i.d. case if we also assume* $E_A(\tau_A^s(\sum_{i=1}^{\tau_A} f(X_i))^s) < \infty$*. Finally, note that, for the Cramer condition (i) to hold, it is sufficient to prove that at least one term in the sum has an absolutely continuous part. Of course condition (i) is more general and may hold even in the discrete case, when* $\sum_{i=1}^{\tau_A} f(X_i)$ *is non-lattice.*

The writing of the terms involved in the Edgeworth expansions using the atom A allows to deduce easily empirical counterparts, which is not the case when they are expressed by using infinite sums (10). We set

$$\widehat{M}_{3,n} = n^{-1}\sum_{j=1}^{l_n-1}\{f(\mathcal{B}_j) - \overline{\mu}_n(f)s_j\}^3/\sigma_n(f)^3,$$

$$\widehat{\beta}_n = n^{-1} \sum_{j=1}^{l_n - 1} s_j \{ f(\mathcal{B}_j) - \overline{\mu}_n(f) s_j \} / \sigma_n(f)$$

and consider the empirical estimator of the skewness defined by

$$\widehat{k}_{3,n} = \widehat{M}_{3,n} - 3\widehat{\beta}_n.$$

A straightforward application of the SLLN (see Theorem 17.1.7 in Meyn & Tweedie (1996)) shows that it is strongly consistent.

Proposition 5.2 *Under the assumptions that the initial distribution fulfills the regularity condition $P_\nu(\tau_A < \infty) = 1$ and that condition (iii) is satisfied with $s = 3$, we have as $n \to \infty$*

$$\widehat{k}_{3,n} \longrightarrow k_3(f), \; P_\nu \; a.s. \; .$$

Following the work of Abramovitz & Singh (1985), it may be easily shown that, under further moment assumptions, the Edgeworth expansion may be inverted to yield better confidence intervals for the sample mean statistic. These results also pave the way for studying the second order validity of the regeneration-based Bootstrap procedure proposed in Datta & McCormick (1993a) (see Bertail & Clémençon (2003a)) for atomic chains, as well as variants for general Harris recurrent Markov chains (refer to Bertail & Clémençon (2003b)).

6 Proofs

6.1 Proof of Proposition 4.4

Set $\widetilde{f} = f - \mu(f)$ and consider the variances $\sigma_\tau^2 = E_A((\tau_A - \alpha)^2)$, $\Sigma^2(f) = E_A(((\sum_{i=1}^{\tau_A} \widetilde{f}(X_i))^2 - \alpha\sigma^2(f))^2)$ and $\Gamma^2(f) = E_A((\tau_A \sum_{i=1}^{\tau_A} \widetilde{f}(X_i) - \beta)^2)$ (recall the notations $\alpha = E_A(\tau_A)$ and $\beta = E_A(\tau_A \sum_{i=1}^{\tau_A} \widetilde{f}(X_i))$ introduced in Proposition 5). Decompose $n(\sigma_n^2(f) - \sigma^2(f))$ into six terms as follows

$$n(\sigma_n^2(f) - \sigma^2(f)) = \sum_{i=1}^{6} D_i,$$

with

$$\begin{aligned}
D_1 &= \sum_{j=1}^{l_n-1}\{(\widetilde{f}(\mathcal{B}_j))^2-\alpha\sigma^2(f)\},\\
D_2 &= \alpha\sigma^2(f)(-1+\sum_{i=1}^{n}\{1_A(X_i)-\mu(A)\}),\\
D_3 &= (\mu(f)-\overline{\mu}_n(f))^2\sum_{j=1}^{l_n-1}s_j^2,\\
D_4 &= 2(\mu(f)-\overline{\mu}_n(f))\sum_{j=1}^{l_n-1}\{s_j\widetilde{f}(\mathcal{B}_j)-\beta\},\\
D_5 &= 2\beta(\mu(f)-\overline{\mu}_n(f))(l_n-\alpha^{-1}n),\\
D_6 &= 2\beta(\mu(f)-\overline{\mu}_n(f))(\alpha^{-1}n-1).
\end{aligned}$$

• The proof that $E_\nu(D_1)=O(1)$ as $n\to\infty$ straightforwardly results from the argument given in the proof of Theorem 1 in Malinovskii (1985), based on a non uniform limit theorem established in Dubinskaite (1982, 1984) (see Lemma 6.5 below), which must be applied in our case to the i.i.d. sequence of 1-lattice two dimensional random vectors $(\Sigma(f)^{-1}((\widetilde{f}(\mathcal{B}_j))^2-\alpha\sigma^2(f)),\ \sigma_\tau^{-1}(s_j-\alpha))_{j\geqslant1}$. Details are thus omitted.

• The application of bound (1) in Proposition 3 to the indicator function 1_A (respectively to f) particularly entails that $E_\nu(D_1)=O(1)$ (respectively, $E_\nu(D_6)=O(1)$) as $n\to\infty$.

• By using Cauchy-Schwarz's inequality, we have

$$E_\nu(D_3)^2\leqslant E_\nu((\mu(f)-\bar{\mu}_n(f))^4)E_\nu((\sum_{j=1}^{l_n-1}s_j^2)^2).$$

Therefore, under our "block" moment conditions, we have according to Theorem 4 in Malinovskii (1987), $E_\nu((\mu(f)-\bar{\mu}_n(f))^4)=O(n^{-2})$. Besides, by simply using the fact that l_n is bounded by n and that $(s_j^2)_{j\geqslant1}$ is an i.i.d. sequence by virtue of the strong Markov property, we derive that

$$E_\nu((\sum_{j=1}^{l_n-1}s_j^2)^2)\leqslant E_A((\tau_A^2-E_A(\tau_A))^2)n+(E_A(\tau_A^2))^2n^2.$$

Combining these two bounds, we obtain that $E_\nu(D_3)=O(1)$ as $n\to\infty$.

• Apply Cauchy-Schwarz's inequality to get

$$E_\nu(D_4)^2\leqslant E_\nu((\mu(f)-\overline{\mu}_n(f))^2)E_\nu((\sum_{j=1}^{l_n-1}\{s_j\widetilde{f}(\mathcal{B}_j)-\beta\})^2).$$

From Theorem 2 in Malinovskii (1985) (see also Theorem 3 in Malinovskii (1987)), we have $E_\nu((\mu(f)-\overline{\mu}_n(f))^2) = O(n^{-1})$ as $n \to \infty$. Moreover, the argument proving this result may also be used to show that $E_\nu((\sum_{j=1}^{l_n-1}\{s_j\widetilde{f}(\mathcal{B}_j)-\beta\})^2) = O(n)$, as $n \to \infty$, by considering the i.i.d. sequence of 1-lattice two dimensional random vectors $(\Gamma(f)^{-1}(s_j\widetilde{f}(\mathcal{B}_j)-\beta),\ \sigma_\tau^{-1}(s_j-\alpha))_{j\geqslant 1}$. Hence, we have $E_\nu(D_4) = O(1)$ as $n \to \infty$.

• Finally, the bound $E_\nu(D_4) = O(1)$ as $n \to \infty$ may be deduced exactly the same way, using first Cauchy-Schwarz's inequality and then applying twice Theorem 2 in Malinovskii (1985), to the function f on the one hand and to the indicator function 1_A on the other hand.

6.2 Proof of the main theorem

• In the following we only consider the case $\mu_n(f) = \widehat{\mu}_n(f)$. The cases $\mu_n(f) = \widetilde{\mu}_n(f)$ and $\mu_n(f) = \overline{\mu}_n(f)$ differ in the treatment of the bias only and may be derived in a similar fashion. The first Edgeworth expansion and control of Δ_n follows immediately from Malinovskii (1987)'s Theorem 1 and its simplified form given in Theorem 5 except that one should read $-\frac{E_\nu(\Sigma_{f,n})}{(E_\pi(\Sigma_{f,n}^2))^{1/2}}$ instead of $\frac{E_\nu(\Sigma_{f,n})}{(E_\pi(\Sigma_{f,n}^2))^{1/2}}$ of course in his result (notice that this term corresponds to the bias, and vanishes in the stationary case).

• To make the reading of the proof much more easy and emphasize the dependence of the statistics considered on the i.i.d. regeneration blocks we introduce the following notations. We denote by $l(\mathcal{B}_j) = s_j = \tau_A(j+1)-\tau_A(j)$ the length of block $\mathcal{B}_j$, $j \geqslant 1$, of which the mean is $E_A(\tau_A) = \alpha$ and the variance is $E_A((\tau_A-\alpha)^2) = \sigma_\tau^2$. We also denote by $l(\mathcal{B}_0) = \tau_A$ and $l(\mathcal{B}_n^{(n)}) = n - \tau_A(l_n)$ the lengths of the first and last (nonregenerative) blocks. Consider the following decomposition

$$\begin{aligned} &n(\mu_n(f)-\mu(f)) - \varphi_\nu - \gamma \\ &= F(\mathcal{B}_0) + \sum_{j=1}^{l_n-1} F(\mathcal{B}_j) + F(\mathcal{B}_{l_n}^{(n)}) \end{aligned}$$

with for $i \geqslant 1$,

$$\begin{aligned} F(\mathcal{B}_j) &= \sum_{\tau_A(j)+1}^{\tau_A(j+1)} \{f(X_i)-\mu(f)\} \\ &= f(\mathcal{B}_j) - l(\mathcal{B}_j)\,\mu(f) \end{aligned}$$

and

$$\begin{aligned} F(\mathcal{B}_0) &= f(\mathcal{B}_0) - l(\mathcal{B}_0)\mu(f) - \varphi_\nu, \\ F(\mathcal{B}_{l_n}^{(n)}) &= f(\mathcal{B}_{l_n}^{(n)}) - l(\mathcal{B}_{l_n}^{(n)})\mu(f) - \gamma. \end{aligned}$$

By the strong Markov properties the $F(\mathcal{B}_j)'s$, $j \geq 1$ are i.i.d. r.v.'s with mean zero and variance $\sigma_F^2 \underset{def}{=} \alpha\, \sigma^2(f)$. Notice also that by construction (see Proposition 3.1), we have

$$E_\nu(F(\mathcal{B}_0)) = 0$$

and

$$E_\nu(F(\mathcal{B}_{l_n}^{(n)})) = O(n^{-1/2}) \text{ as } n \to \infty.$$

We also recall that with these notations, for $j \geqslant 1$,

$$\beta = cov(l(\mathcal{B}_j), F(\mathcal{B}_j)).$$

The matter is here to extend Malinovskii (1987)'s results to derive an Edgeworth expansion for $\widetilde{t}_n$. In the following we shall derive such an expansion up to $O(n^{-1}\log(n))$.

6.2.1 Preliminary lemmas

The following classical lemma (see Chibisov (1972)) will be used extensively in the proof.

Lemma 6.1 *Assume that W_n admits an Edgeworth expansion on the normal distribution up to $O(n^{-1}l(n))$, for some function $l(n)$ such that $l(n)/n \to 0$ as $n \to \infty$. Assume that R_n is such that $P(n|R_n| > \eta l(n))$ is either $O(n^{-1}l(n))$ or $O(n^{-1})$ as $n \to \infty$ for some constant $\eta > 0$, then $W_n + R_n$ and $W_n/(1+R_n)^{1/2}$ (when defined) have the same Edgeworth expansion as W_n up to $O(n^{-1}l(n))$.*

In the following we will typically choose $l(n)$ to be n^ε, $1 > \varepsilon \geq 0$ or $l(n) = log(n)$ or $log(n)^{1/2}$. In the same spirit, we will also use the following inequalities and estimates.

Lemma 6.2 *Suppose that the following "block" moment condition is fulfilled*

$$E_A(|\sum_{i=1}^{\tau_A} \tilde{f}(X_i)|^2) < \infty,$$

then there exists some constants c_0 and c_1 such that we have for all n,

$$P_\nu(n^{-1} \mid \sum_{j=1}^{l_n-1} F(\mathcal{B}_j) \mid \geqslant x) \leqslant c_0\{\exp(-\frac{nx^2}{c_1+yx})$$

$$+nP_A(\mid \sum_{i=1}^{\tau_A} \tilde{f}(X_i) \mid \geqslant y) + P_\nu(\tau_A > n/2) + P_A(\tau_A > n/2)\}.$$

In particular under the condition (iii) and (iv) with $s = 8+\varepsilon$, $\varepsilon > 0$, there exists some constant $\eta > 0$ such that, as $n \to \infty$,

$$P_\nu(n^{-1/2} \mid \sum_{j=1}^{l_n-1} F(\mathcal{B}_j) \mid \geqslant \eta \log(n)^{1/2}) = O(n^{-1})$$

and

$$P_\nu(n^{1/2} \mid \frac{1}{n}\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)l(\mathcal{B}_j) - \alpha^{-1}\beta) \mid \geqslant \eta \log(n)^{1/2}) = O(n^{-1}).$$

Proof. The first inequality may be derived from the argument of Theorem 15 in Clémençon (2001) based on the Fuk & Nagaev's inequality for sums of unbounded r.v.'s (see also Theorem 6.1 in Rio (2000) for an argument based on block mixing techniques). In particular, for $x = \eta \log(n)^{1/2} n^{-1/2}$, $y = \log(n)^{-1/2} n^{1/2}$, if we choose $\eta > 0$ such that $\eta^2 \geq c_1 + \eta$, applying Chebyshev's inequality to the last three terms in the right hand side of the inequality yields

$$\begin{aligned}
&P_\nu(n^{-1/2} \mid \sum_{j=1}^{l_n-1} \tilde{f}(\mathcal{B}_j) \mid \geqslant \eta \log(n)^{1/2}) \\
&\qquad \leqslant c_0\{\exp(-\frac{\eta^2\log(n)}{c_1+\eta}) + \frac{(\log n)^{2+s/2}}{n^{1+s/2}} E_A(\mid \sum_{i=1}^{\tau_A} \tilde{f}(X_i)|^{4+\varepsilon/2}) \\
&\qquad + 2n^{-1}E_\nu(\tau_A) + 2n^{-1}E_A(\tau_A)\} \\
&\qquad \leq C_1 n^{-1}.
\end{aligned}$$

The second bound may be established similarly, using Cauchy-Schwarz inequality. ■

The lemma below shows how the estimated variance may be linearized with a controlled remainder.

Lemma 6.3 *Under the hypotheses of Theorem 5.1 we have*

$$\sigma_n^2(f) = n^{-1}\sum_{j=1}^{l_n-1} g(\mathcal{B}_j) + r_n \tag{11}$$

with, for $j \geqslant 1$,

$$g(\mathcal{B}_j) = F(\mathcal{B}_j)^2 - 2\alpha^{-1}\beta F(\mathcal{B}_j)$$

and for some $\eta_1 > 0$,

$$P(nr_n > \eta_1 \log(n)) = O(n^{-1}),\ \textit{as}\ n \to \infty.$$

Proof. We have

$$\begin{aligned}\sigma_n^2(f) &= n^{-1}\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)^2 - 2(\overline{\mu}_n(f) - \mu(f))n^{-1}\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)l(\mathcal{B}_j) \\ &\quad + (\overline{\mu}_n(f) - \mu(f))^2 n^{-1}\sum_{j=1}^{l_n-1} l(\mathcal{B}_j)^2 \\ &= n^{-1}\sum_{j=1}^{l_n-1} g(\mathcal{B}_j) + r_n\end{aligned}$$

with $r_n = r_{1,n} + r_{2,n} + r_{3,n}$

$$r_{1,n} = -2((\overline{\mu}_n(f) - \mu(f))(n^{-1}\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)l(\mathcal{B}_j) - \alpha^{-1}\beta)$$

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

$$r_{3,n} = 2(1 - (1 - l(\mathcal{B}_0)/n - l(\mathcal{B}_{l_n}^{(n)})/n)^{-1})n^{-1}\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)\alpha^{-1}\beta$$

The control of the remainder follows from Lemma 6.2, we have for $\eta_1 > 0$ large enough

$$\begin{aligned}&P_\nu(n|r_{1,n}| > \eta_1 \log(n)) \leq \\ &P_\nu(n^{1/2}|\overline{\mu}_n(f) - \mu(f)| \geq 2\eta_1^{1/2}\log(n)^{1/2}) + \\ &\qquad P_\nu(n^{-1/2}|\sum_{j=1}^{l_n-1} F(\mathcal{B}_j)l(\mathcal{B}_j) - \alpha^{-1}\beta| \geq \eta_1^{1/2}\log(n)^{1/2}) \\ &\qquad = O(n^{-1}),\end{aligned}$$

as $n \to \infty$. We also have, using the same arguments, that

$$P_v(n|r_{2,n}| > \eta \log(n)) = O(n^{-1}),\ \text{as}\ n \to \infty.$$

Finally, since $E_\nu(l(\mathcal{B}_0)^2) = E_\nu(\tau_A^2) < \infty$ and $E(l(\mathcal{B}_{l_n}^{(n)})^2) \leqslant E_A(\tau_A^2) < \infty$, we have by virtue of Markov inequality

$$P_v(l(\mathcal{B}_0) > n^{1/2}) = O(n^{-1}),$$
$$P_\nu(l(\mathcal{B}_{l_n}^{(n)}) > n^{1/2}) = O(n^{-1}),$$

as $n \to \infty$. Besides on the event $\{l(\mathcal{B}_0) \leq n^{1/2}\} \cap l(\mathcal{B}_{l_n}^{(n)}) \leq n^{1/2}\}$, we have for $n \geq 4$, $|1 - (1 - l(\mathcal{B}_0)/n - l(\mathcal{B}_{l_n}^{(n)})/n)^{-1})| \leq 4n^{-1/2}$. Thus, for $\eta_2 > 0$,

$$\begin{aligned} P_v(nr_{3,n} > \eta_2 \log(n)) &= P_\nu(|n^{-1/2} \sum_{j=1}^{l_n - 1} F(\mathcal{B}_j)|\alpha^{-1}\beta > \eta_2 \log(n)) + O(n^{-1}) \\ &= O(n^{-1}) \end{aligned}$$

as $n \to \infty$, by applying Lemma (6.2). ■

The following lemma implies that we may restrict the study of the standardized sums to values of l_n in an interval $I_n(\varepsilon) = [n\alpha^{-1} - n^{1/2+\delta},\ n\alpha^{-1} + n^{1/2+\delta}] \cap [1, n]$. It derives from the same argument as Lemma 6.2, applied to the indicator function 1_A of A.

Lemma 6.4 *Let $X = (X_n)_{n\in\aleph}$ be a Markov chain with an atom A. Suppose that X is positive recurrent with stationary distribution μ. Let $l_n = \sum_{i=1}^n I\{X_i \in A\}$ be the number of visits of X to A between time 1 and time n. Assume further that there exists $p \geqslant 2$ such that $E_A(\tau_A^p) < \infty$, and that there exists $q \geqslant 1$ such that the initial distribution ν satisfies $E_\nu(\tau_A^q) < \infty$. Then as $n \to \infty$, we have*

$$P_\nu\left(n^{1/2}\left|l_n/n - \mu(A)\right| \geqslant n^\delta\right) = O\left(n^{-1}\right),$$

for all δ such that $\delta > (2/p - 1/2)_+$ and $\delta \geqslant (1/q - 1/2)_+$.

The following lemma (which is a non-uniform version of Malinovskii (1987), see Lemma 1 p. 283) is a consequence of Dubinskaite (1984)'s Theorem 2 and its corollaries 8 and 9. To state the result, we use the usual notations for characteristic functions and Edgeworth expansion in the multidimensional case (see section 7 of Battacharya & Rao (1975)). Let $\phi_{0,W}$ be the density of the normal density with mean 0 and variance W. Its Fourier transform is given by

$$\widehat{\phi_{0,W}}(t) = \exp\left(-\frac{1}{2}(t'Wt)\right)$$

For some square integrable r.v. ξ taking its values in R^p with covariance matrix W , the polynomial associated with the cumulants $\{\chi_\theta\}$ of order $\theta = (\theta_i)_{1\leqslant i\leqslant p} \in \mathsf{N}^p$ such that $|\theta| = \sum_{i=1}^{p} |\theta_i|$ is less than 3 is denoted by

$$\widetilde{P}_1(it, \{\chi_\theta\}_{|\theta|\leq 3}) = \frac{i^3}{6} E\left((t'\xi)^3\right)$$

and let

$$P_1(-\phi_{0,W},\ \{\chi_\theta\}_{|\theta|\leq 3})(t) = -\sum_{|\theta|\leqslant 3} \frac{\chi_\theta}{\theta_1!...\theta_p!} \phi_{0,W}^{(\theta)}(t)$$

be the corresponding transformation, the explicit form of which is given p. 55 of Battacharya & Rao (1975).

Lemma 6.5 Edgeworth Expansion for 1-lattice distribution [*Dubinskaite (1982, 1984)*]: *Let $\xi = (\xi_1, \xi_2, \xi_3)$ a centered random vector such that ξ_3 is lattice with minimal span $H > 0$ and is valued in $\{kH+\alpha\}_{k\in\mathsf{Z}}$. Suppose that ξ_1 satisfies the Cramer condition and that ξ_2 has a bounded density. Assume further that the covariance matrix $W = var(\xi_1, \xi_2, \xi_3)$ is non singular and that $E|\xi_i|^4 < \infty,\ 1 \leqslant i \leqslant 3$. Then, for an i.i.d. sequence $(\xi_{1,i}, \xi_{2,i}, \xi_{3,i})_{i\geqslant 1}$ drawn from ξ, we have up to a constant $C > 0$*

$$\begin{aligned} &|\frac{\sqrt{m}}{H} P(\sqrt{m}\sum_{i=1}^{m} \xi_{1,i} \leq x, \sum_{i=1}^{m} \xi_{3,i} = kH + \alpha m \mid \sqrt{m}\sum_{i=1}^{m} \xi_{2,i} = z) p_{f_m}(z) \\ &-E_{W,m}^{(2)}(x, z, \frac{kH+\alpha m}{\sqrt{m}})| \leq C.m^{-1}\left(1 + |\frac{kH+\alpha m}{\sqrt{m}}| + |z|\right)^{-4}, \end{aligned}$$

with

$$E_{W,n}^{(2)}(x, z, \frac{kH+\alpha m}{\sqrt{m}}) = \int_{-\infty}^{x} DE_{W,n}^{(2)}(y, z, \frac{kH+\alpha m}{\sqrt{m}}) dy,$$

where

$$\begin{aligned} DE_{W,n}^{(2)}(y, z, \frac{kH+\alpha m}{\sqrt{m}}) &= \phi_{0,W}(y, z, \frac{kH+\alpha m}{\sqrt{m}}) \\ &\quad + \frac{1}{\sqrt{m}} P_1(-\phi_{0,W},\ \{\chi_v\})(y, z, \frac{kH+\alpha m}{\sqrt{m}}) \end{aligned}$$

and p_{f_m} denotes the density of $f_m = \sqrt{m}\sum_{i=1}^{m} \xi_{2,i}$.

Proof. Given that $s = 4$ is even, it is legitimate to choose $s^* = s$ and $r = 0$ in Dubinskaite (1984)'s Theorem 2. Condition P_{l-m} of this theorem holds because of the boundedness condition on the density of ξ_2. Since W is assumed to be nonsingular, the smallest eigenvalue of W is strictly positive. The function $L_{s,n}$ may be thus bounded by $C \sum E(\xi_{1,i}^4)$, so that all the terms of the bound may be swallowed into the constant C (depending on the underlying probability). Note that Lemma 1 in Malinovskii (1987) is the uniform version (over x) of this lemma, with the choice $s_1 = 3$ and $\delta = 1$. ■

The following lemma is interesting for other calculations of the same type and will allow us to control the terms in the sums appearing in the Edgeworth expansion.

Lemma 6.6 *Let $a_{n,m} = (n - \alpha m)/(\sigma_\tau \sqrt{m})$ and $DE_m(y, \lambda) = \phi_V(y, \lambda) + m^{-1/2} P(y, \lambda)\phi_V(x, \lambda)$, where $P(y, \lambda)$ is a polynomial in $y \in \mathrm{R}^p$ and $\lambda \in \mathrm{R}$, and V a nonsingular covariance matrix, then there exists some nonnegative constant K and a polynomial $Q(.)$ independent from n with a fixed degree such that*

$$\left\{ \sum_{m=1}^{n} \frac{\alpha}{\sigma_\tau \sqrt{m}} DE_m(y, a_{n,m}) - \int_{-\infty}^{\infty} \phi_V(y, \lambda) d\lambda - \frac{\sqrt{\alpha}}{\sqrt{n}} \int_{-\infty}^{\infty} P(y, \lambda)\phi_V(y, \lambda) d\lambda \right.$$
$$\left. + \frac{1}{2} \frac{\sigma_\tau}{\alpha^{1/2}\sqrt{n}} \int_{-\infty}^{\infty} \lambda \phi_V(y, \lambda) d\lambda) \right\} \leq Q(y) \exp(-K||y||^2)\, n^{-1}.$$

Moreover, for some nonnegative constant K and a polynomial $R(.)$, we have that

$$|\sum_{m=1}^{n} \frac{1}{m^{3/2}} P(y, a_{n,m}) \phi_V(y, a_{n,m})| \leq R(y) \exp(-K||y||^2)\, n^{-1}. \qquad (12)$$

Proof. The proof follows from the argument given in Malinovskii (1985) (see his equations (10) to (15))). By Taylor expansion, for any function F with continuous derivatives $\frac{\partial^{(i)} F(y,\lambda)}{\partial \lambda^{(i)}}$ with respect to λ, we have that

$$\left| \int_{a_{n,n}}^{a_{n,1}} F(y, \lambda) d\lambda - \sum_{m=1}^{n} (a_{n,m} - a_{n,m+1}) F(y, a_{n,m}) - \right. \qquad (13)$$
$$\left. \sum_{m=1}^{n-1} \frac{1}{(2)!} \left. \frac{\partial^{(1)} F(y, \lambda)}{\partial \lambda} \right|_{\lambda = a_{n,m}} (a_{n,m} - a_{n,m+1})^2 \right| \qquad (14)$$
$$\leq \sum_{m=1}^{n-1} \frac{1}{3!} |a_{n,m} - a_{n,m+1}|^2 \sup_{\lambda \in [a_{n,m}, a_{n,m+1}]} \frac{\partial^{(2)} F(y, \lambda)}{\partial \lambda^{(2)}}.$$

Noticing that

$$a_{n,m} - a_{n,m+1} = \frac{\alpha}{\sigma_\tau \sqrt{m}} + a_{n,m+1}((1 + m^{-1})^{1/2} - 1), \qquad (15)$$

use first the Taylor expansion (13) with $F(y, \lambda) = \phi_V(y, \lambda)$. For these functions, we obviously have for some non negative constants K, k and some polynomial $P^i(y, \lambda)$ of degree less than i

$$\sup_{\lambda \in [a_{n,m}, a_{n,m+1}]} \frac{\partial^{(i)} F(y, \lambda)}{\partial \lambda^{(i)}} \leq C\, P^i(y, a_{n,m}) \exp(-K||y||^2) \exp(-k a_{n,m}^2) \qquad (16)$$

In the following $P_i, i = 1, 2, ...$ is a sequence of polynomials in y of finite degree (typically lower than 8) and K_i, $i = 1, 2, ...$ some non negative constants. Proceeding as Malinovskii (1985, 1987) (see (13)), it is then easy to see that

$$\sum_{m=1}^{n} |a_{n,m} - a_{n,m+1}|^2 \sup_{\lambda \in [a_{n,m}, a_{n,m+1}]} \frac{\partial^{(2)} F(y, a_{n,m})}{\partial \lambda^{(2)}} \leq n^{-1} P_1(y) \exp(-K_1||y||^2).$$

Using successively (13) with $F(y, \lambda) = \phi_V(y, \lambda)$ and $\frac{\partial \phi_V(y,\lambda)}{\partial \lambda}$, we get

$$\begin{aligned} &| \int_{a_{n,n}}^{a_{n,1}} \phi_V(y, \lambda) d\lambda - \sum_{m=1}^{n} \frac{\alpha}{\sigma_\tau \sqrt{m}} \phi_V(y, a_{n,m}) \\ &- \frac{1}{2} \frac{\alpha^{3/2}}{\sigma_\tau} \frac{1}{\sqrt{n}} \int_{a_{n,n}}^{a_{n,1}} \frac{\partial \phi_V(y, \lambda)}{\partial \lambda} d\lambda - \frac{1}{2} \frac{\sigma_\tau}{\alpha^{1/2}} \frac{1}{\sqrt{n}} \int_{a_{n,n}}^{a_{n,1}} \lambda \phi_V(y, \lambda) d\lambda| \\ &\leq n^{-1} P_2(y) \exp(-K_2||y||^2). \end{aligned}$$

It follows that

$$\begin{aligned} &| \sum_{m=1}^{n} \frac{\alpha}{\sigma_\tau \sqrt{m}} \phi_V(y, a_{n,m}) - \int_{a_{n,n}}^{a_{n,1}} \phi_V(y, \lambda) d\lambda \\ &+ \frac{1}{2} \frac{\sigma_\tau}{\alpha^{1/2}} \frac{1}{\sqrt{n}} \int_{a_{n,n}}^{a_{n,1}} \lambda \phi_V(y, \lambda) d\lambda| \leq n^{-1} P_3(y) \exp(-K_3||y||^2). \end{aligned}$$

Now following Malinovskii (1985) (equations (12) and (13) using (15), we have

$$\begin{aligned} &| \sum_{m=1}^{n} \frac{\alpha}{m \sigma_\tau} P(y, a_{n,m}) \phi_V(y, a_{n,m}) - \sum \frac{1}{\sqrt{m}} (a_{n,m} - a_{n,m+1}) P(y, a_{n,m}) \phi_V(y, a_{n,m})| \\ &\leq C_1 n^{-1} P_4(y) \exp(-K_4||y||^2), \end{aligned}$$

as well as

$$
\begin{aligned}
&|\sum_{m=1}^{n} \frac{1}{\sqrt{m}}(a_{n,m} - a_{n,m+1})P(y, a_{n,m})\phi_V(y, a_{n,m}) \\
&- \frac{\alpha^{1/2}}{\sqrt{n}} \sum_{m=1}^{n} P(y, a_{n,m})\phi_V(y, a_{n,m})(a_{n,m} - a_{n,m+1})| \\
&\leq C_2 n^{-1/2} \sum_{m=1}^{n} |(\frac{n}{\alpha m})^{1/2} - 1|a_{m,n} P(y, a_{n,m}) \exp(-K_5 a_{n,m}) \exp(-K_5||y||^2) \\
&\leq n^{-1} P_6(y) \exp(-K_6||y||^2).
\end{aligned}
$$

And using (13) with $F(y, \lambda) = P(y, \lambda)\phi_V(y, \lambda)$, we have

$$
\begin{aligned}
&|\frac{\alpha^{1/2}}{\sqrt{n}} \sum_{m=1}^{n} P(y, a_{n,m})\phi_V(y, a_{n,m})(a_{n,m} - a_{n,m+1}) - \int_{a_{n,n}}^{a_{n,1}} \frac{\alpha^{1/2}}{\sqrt{n}} P(y, \lambda)\phi_V(y, \lambda) d\lambda| \\
&\leq n^{-1} P_7(y) \exp(-K_7||y||^2).
\end{aligned}
$$

The proof follows by combining these three inequalities and by observing that for $\alpha > 1$ the remainder in the integrals $\int_{-\infty}^{a_{n,n}}$ and $\int_{a_{n,1}}^{\infty}$ may be bounded by Cn^{-1} for some constant $C > 0$. The proof of (12) is similar. ■

6.2.2 Edgeworth expansion of the standardized sum

The main problem for obtaining the Edgeworth expansion is to control the first and last blocks, which are not regenerative blocks, on the one hand and the randomness of the number of blocks on the other hand. We use the same techniques as the ones required to establish similar results in Bolthausen (1980) and in Malinovskii (1987, 1989). Once some necessary basic tools developed, we only give here the main ideas of the proof. We proceed in five steps, as follows: reduce the original problem to a simplified version (step 1), partition the probability space according to the number of regenerative blocks and the length of the first and last blocks (step 2), derive an Edgeworth expansion for each element induced by the partition (step 3), then sum up all the expansions and approximates the sums involved by Riemann integrals (step 4) and finally compute explicitly the main term of the expansion (step 5).

Step 1 : reduction to a simplified statistic Lemma 6.1 and Lemma 6.3 imply that establishing the Edgeworth expansion of the original standardized statistic reduces, up to $O(n^{-1}\log(n))$, to obtain the Edgeworth expansion of

$$P_\nu\left(\frac{\sum_{j=0}^{l_n} F(\mathcal{B}_j)}{\left(\sum_{j=1}^{l_n-1} g(\mathcal{B}_j)\right)^{1/2}} \le x - \frac{\phi_\nu}{\sigma(f)} n^{-1/2} - \frac{\gamma}{\sigma(f)} n^{-1/2}\right)$$

We thus focus on the Edgeworth expansion of

$$L_n = \frac{\sum_{j=0}^{l_n} F(\mathcal{B}_j)}{\left(\sum_{j=1}^{l_n-1} g(\mathcal{B}_j)\right)^{1/2}}$$

Combining Lemmas 6.1 and 6.4 with $p = 4$ and $\delta = \varepsilon$ yields that

$$P_\nu(L_n \le x) = P_\nu(L_n \le x,\ l_n \in I_n(\varepsilon)) + O(n^{-1}), \text{ as } n \to \infty.$$

where $I_n(\varepsilon) = [n\alpha^{-1} - n^{1/2+\delta},\ n\alpha^{-1} + n^{1/2+\delta}] \cap [1, n]$.

Step 2 : partitioning Consider the partition of the underlying probability space into the following disjoint measurable subsets

$$U_r = \{\tau_A(1) = r,\ \tau_A(2) - \tau_A(1) > n - r\},$$

$$\begin{aligned} U_{r,l,m} &= \{\tau_A(1) = r,\ \tau_A(m) = n - l\ ,\ \tau_A(m+1) > n\} \\ &= \{\tau_A(1) = r,\ \sum_{j=2}^{m} \tau_A(j) - \tau_A(j-1) = n - r - l,\ \tau_A(m+1) > n\}. \end{aligned}$$

Now define for $j \geqslant 1$,

$$\sigma_G^2 = E((g(\mathcal{B}_j) - E(g(\mathcal{B}_j))^2)$$

and write

$$T_m(u, v) = u + v + m^{-1/2} \sum_{j=1}^{m} F(\mathcal{B}_j)/\sigma_F,$$

$$S_m^2 = \sigma_F^2 \left(1 + \sigma_G/\sigma_F^2 G_m\right)^{1/2},$$

with

$$G_m = m^{-1/2} \sum_{j=1}^{m} (g(\mathcal{B}_j) - \sigma_F^2)/\sigma_G.$$

We have as $n \to \infty$

$$P_\nu(L_n \le x, l_n \in I_n(\varepsilon)) = I + II \ + O(n^{-1}),$$

with

$$I = \sum_r P_\nu(\{L_n \le x\} \cap U_r) \le P_\nu(\tau_A(2) > n,\ L_n \le x) = O(n^{-1}),$$

$$II = \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \sum_{m+1 \in I_n(\varepsilon)} \int\int P\left(\frac{T_m(u,v)}{S_m} \le x\ ,\ \sum_{j=1}^{m} l(\mathcal{B}_j) = n - r - l\right)$$
$$P_\nu(L^r \in du, \tau_A(1) = r) P_A(L^l \in dv,\ \tau_A(m+1) > l),$$

where L^r (resp.L^l) is the distribution under P_ν of $m^{-1/2}F(\mathcal{B}_0)/\sigma_F$ when $\tau_A(1) = r$ and L^l the distribution of $m^{-1/2}F(\mathcal{B}_{m+1})/\sigma_F$ when the length of $\mathcal{B}_{m+1}$ is l. To simplify the notations we set

$$P_r(du) = P_\nu(L^r \in du, \tau_A(1) = r)$$

and

$$P_{m,l}(dv) = P_A(L^l \in dv,\ \tau_A(m+1) > l).$$

Notice that by Lemmas 6.4 and 6.1 we may indifferently put $\sum_{m+1 \in I_n(\varepsilon)}$ or $\sum_{m=0}^{n-1}$ in II up to $O(n^{-1})$.

Step 3 : Edgeworth expansion for 1-lattice distribution Thus we essentially have to show that

$$III = P\left(\frac{T_m(u,v)}{S_m} \le x\ ,\ \sum_{j=1}^{m-1} l(\mathcal{B}_j) = n - r - l\right)$$

admits an Edgeworth expansion with a remainder such that the sums and integrals in II are of order $O(n^{-1})$. The second component may be written as a lattice sum

$$L_m \underset{def.}{\equiv} m^{-1/2} \sum_{j=1}^{m} (l(\mathcal{B}_j) - \alpha)/\sigma_\tau = a_{n,l,m,r}$$

where

$$a_{n,l,m,r} = (n - r - l - \alpha m)/(\sigma_\tau \sqrt{m}).$$

Conditioning on G_m we get

$$III = P_m(x, a_{n,l,m,r}) = \int p_{G_m}(z) \times$$
$$P(\frac{m^{-1/2}\sum_{j=1}^m F(\mathcal{B}_j)}{\sigma_F} \leq x(z,u,v,m), L_m = a_{n,l,m,r} | \; G_m = z)dz$$

with
$$x(z,u,v,m) = x(1+m^{-1/2}\sigma_G/\sigma_F^2 z)^{1/2} - u - v,$$
and denoting by $p_{G_m}(z)$ the density of G_m. Notice that one cannot condition first on the quadratic term and then directly apply Theorem 2 in Malinovskii (1987) because of the form of the variance (which is a sum of functions of the blocks and not of the original data) and the non uniformity of the bound in y (see his last expression on p. 273). The Edgeworth expansion of the expression under the integral in III may be deduced using Lemma 6.5. For this, consider $(\xi_{1,j}, \, \xi_{2,j}, \, \xi_{3,j})_{j\geqslant 1}$ with $\xi_{1,j} = F(\mathcal{B}_j)/\sigma_F$, $\xi_{2,j} = (l(\mathcal{B}_j) - \alpha)/\sigma_\tau$ (which is lattice with span $H = \sigma_\tau^{-1}$) and $\xi_{3,j} = (g(\mathcal{B}_j) - \sigma_F^2)/\sigma_G$, that is by construction of the blocks an i.i.d. sequence. Note that the condition $E|\xi_{3,j}|^4 < \infty, \; 1 \leqslant i \leqslant 3$ reduces to condition (iii) with $s = 8$. From (6.5) we get that

$$\sup_x |P_m(x, a_{n,l,m,r}) - \frac{\sigma_\tau^{-1}}{\sqrt{m}} \int_{-\infty}^{\infty} E_{W,m}^{(2)}(x(z,u,v,m), z, a_{n,l,m,r})dz| \qquad (17)$$
$$\leq Cm^{-3/2} \left(1 + |a_{n,l,m,r}|^3\right)^{-1},$$

where $W = (W_{i,j})_{1\leq i,j\leq 3}$ is a symmetric (3,3) matrix with

$$W_{11} = W_{22} = W_{33} = 1,$$
$$W_{1,2} = \sigma_G^{-1}\sigma_F^{-1} cov(F(\mathcal{B}_j), g(\mathcal{B}_j)) = \sigma_F^2 M_{3,A}/\sigma_G - 2\alpha^{-1}\beta\sigma_F\sigma_G^{-1},$$
$$W_{1,3} = \sigma_F^{-1}\sigma_\tau^{-1} cov(F(\mathcal{B}_j), l(\mathcal{B}_j)) = \sigma_F^{-1}\sigma_\tau^{-1}\beta,$$
$$W_{2,3} = \sigma_G^{-1}\sigma_\tau^{-1} cov(g(\mathcal{B}_j), l(\mathcal{B}_j)).$$

The last inequality in (17) straightforwardly results from the bound
$$\int 1/(1+|\lambda|+|x|)^4 dx \leq C\frac{1}{(1+|\lambda|)^3}.$$

Step 4 : control of the sums of the expansions and their remainders

To prove that the remainder in the expansion of II is of order $O(n^{-1})$, we use the same arguments as the ones used to prove 3.5 in Bolthausen (1980). As a matter of result, we have in our case

$$m^{-3/2} \left(1 + |a_{n,l,m,r}|^3\right)^{-1} \leq C \left\{ \begin{array}{ccc} m^{-3/2} & & |n - \alpha m| \leq 2\sqrt{n} \\ (n-\alpha m)^{-3} & \text{if} & \alpha m > n + 2\sqrt{n} \\ (n - 2\sqrt{n} - \alpha m)^{-2} & & \alpha m < n + 2\sqrt{n} \end{array} \right.$$

so that by straightforward decomposition using the fact that $\sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \int\int P_r(du) P_{m,l}(dv) \le C$ we have

$$\int\int \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \sum_{m+1 \in I_n(\varepsilon)} m^{-3/2} \left(1 + |a_{n,l,m,r}|^3\right)^{-1} P_r(du) P_{m,l}(dv)$$
$$= O(n^{-1}),$$

as $n \to \infty$. The matter is now to show that the main part has the form indicated in (9), that is

$$IV = \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \sum_{m+1 \in I_n(\varepsilon)} \frac{1}{\sigma_\tau \sqrt{m}} \int\int\int_{-\infty}^{\infty} E_{W,m}^{(2)}(x(z,u,v,m), z, a_{n,l,m,r})\ P_r(du)$$
$$\times P_{m,l}(dv) dz$$
$$= F_n^{(2)}(x) + O(n^{-1}).$$

We may rewrite this expression the following way

$$\int_{-\infty}^{\infty} \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \sum_{m=1}^{n} \frac{1}{\sigma_\tau \sqrt{m}} \int\int E_{W,m}^{(2)}(x(z,u,v,m), z, a_{n,l,m,r}) P_r(du) P_{m,l}(dv) dz.$$

A Taylor expansion of $E_{W,m}^{(2)}(x(z,u,v,m), z, a_{n,l,m,r})$ at

$$x(z,m) := x(1 + z m^{-1/2} \sigma_G / \sigma_F^2)^{1/2}$$

yields for some $x^* \in [x(z,u,v,m),\ x(z,m)]$

$$E_{W,m}^{(2)}(x(z,u,v,m), z, a_{n,l,m,r})$$
$$= E_{W,m}^{(2)}(x(z,m), z, a_{n,l,m,r}) + (u+v) D E_{W,m}^{(2)}(x(z,m), z, a_{n,l,m,r})$$
$$+ 2^{-1}(u+v)^2 \partial D E_{W,m}^{(2)}(x^*, z, a_{n,l,m,r}) / \partial x.$$

Using the same arguments as in Malinovskii (1985, 1987) (see (4) and proof of Theorem 2 with $s = 4$), it is cumbersome but rather straightforward (using as in Bolthausen (1980) the fact that for some non negative constants k_1, k_2 and k_3 , $\phi_{0,W}(x,z,\lambda) \le \exp(-k_1 x^2) \exp(-k_2 z^2) \exp(-k_3 \lambda^2)$ and bounds of type (16) combined with lemma 6.6 (see (12)) to show that, for either

$$v(z,u,v,m) = (u+v)^2 \frac{\partial D E_{W,m}^{(2)}}{\partial x}(x^*, z, a_{n,l,m,r})$$

or else

$$v(z,u,v,m) = m^{-1/2} P_1(-\phi_{0,W}, \{\chi_\theta\}_{|\theta| \leqslant 3})(x(z,m), z, a_{n,l,m,r})(u+v),$$

we have

$$\sum_{l=1}^{\sqrt{n}}\sum_{r=1}^{\sqrt{n}}\sum_{m+1\in I_n(\varepsilon)} m^{-1/2}\int_{-\infty}^{\infty}\int\int v(z,u,v,m)P_r(du)P_{m,l}(dv)dz = O(n^{-1}),$$

as $n \to \infty$. This is easier in our situation, since we have already recentered the original statistic, so that

$$\sum_{l=1}^{\sqrt{n}}\sum_{r=1}^{\sqrt{n}}\left(\int\int (u+v)P_r(du)P_{m,l}(dv)\right) = O(n^{-1}) \tag{18}$$

and

$$\sum_{l=1}^{n}\sum_{r=1}^{n}\left(\int\int (u^2+v^2)P_r(du)P_{m,l}(dv)dz\right) \leq Cm^{-1}, \tag{19}$$

given the assumed moment conditions for τ_A and $f(\mathcal{B}_0)$ under P_ν. We thus get

$$IV = \sum_{l=1}^{\sqrt{n}}\sum_{r=1}^{\sqrt{n}}\sum_{m+1\in I_n(\varepsilon)}\frac{1}{\sigma_\tau\sqrt{m}}\int_{-\infty}^{\infty}E_{W,m}^{(2)}(x(z,m),z,a_{n,l,m,r})dz$$
$$\times\int\int P_r(du)P_{m,l}(dv) + \sum_{l=1}^{\sqrt{n}}\sum_{r=1}^{\sqrt{n}}\sum_{m+1\in I_n(\varepsilon)}\frac{1}{\sigma_\tau\sqrt{m}}\int_{-\infty}^{\infty}\phi_{0,W}(x(z,m),z,a_{n,l,m,r})$$
$$\times\left(\int\int_{-\infty}^{\infty}(u+v)P_r(du)P_{m,l}(dv)\right)dz + O(n^{-1}).$$

Now use exactly the same arguments as in Malinovskii (1987) p. 279-280 (or Malinovskii (1985), p. 331), that is to say, develop

$$F_{W,m}^{(2)}(x,a_{n,l,m,r}) = \int_{-\infty}^{\infty}E_{W,m}^{(2)}(x(z,m),z,a_{n,l,m,r})dz$$

at the point $a_{n,m} = (n-\alpha m)/(\sigma_\tau\sqrt{m})$ to get that

$$IV = V + VI\ + O_P(n^{-1}),$$

with

$$V = \sum_{m=1}^{n} \frac{1}{\sigma_\tau \sqrt{m}} F^{(2)}_{W,m}(x, a_{n,m}) \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} P_\nu(\tau_A(1) = r) P_A(\tau_A(m+1) > l)$$
$$+ O_P(n^{-1})$$
$$= \sum_{m=1}^{n} \frac{\alpha}{\sigma_\tau \sqrt{m}} F^{(2)}_{W,m}(x, a_{n,m}) + O_P(n^{-1}),$$
$$VI = \sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \sum_{m \in I_n(\varepsilon)+1} \int_{-\infty}^{\infty} \frac{1}{\sigma_\tau \sqrt{m}} \int_{-\infty}^{x} \phi_{0,W}(y(z,m), z, a_{n,l,m,r}) \times$$
$$\int \left(\int \int (u+v) P_r(du) P_{m,l}(dv) \right) dydz + O(n^{-1}).$$

But we have

$$\sum_{l=1}^{\sqrt{n}} \sum_{r=1}^{\sqrt{n}} \int \int \; P_r(du) P_{m,l}(dv) = \sum_{r=1}^{\sqrt{n}} P_\mu(\tau_A = r) \sum_{l=1}^{\sqrt{n}} P_A(\tau_A > l)$$
$$= \alpha \; + O_P(n^{-1}).$$

Now the main difference with the calculations in Malinovskii (1987) lies in the last term VI, which is is simply the second term in his expression $A_{1,0}$ (see also the term A_1 p. 329 in Malinovskii (1985)). Once again we use the fact that the original statistic is correctly recentered (see (18, 19)) and Lemma 6.6 to get

$$VI = O(n^{-1}), \text{ as } n \to \infty.$$

It should be noticed that in opposition to Malinovskii (1987)'s term $A_{1,0}$, which is the equivalent of VII in our expansion, VII does not contribute to the expansion because of the recentering and the fact that we standardized by $\sqrt{m}$ instead of $\sqrt{n/\alpha}$ after having conditioned on the variance.

Step 5 : explicit computation of the main part The proof is finished by observing that a straightforward Taylor expansion at x and a repeated use of Lemma 6.6 yield

$$V = \int_{-\infty}^{\infty} \int_{-\infty}^{x} \int_{-\infty}^{\infty} \left\{ \phi_W(y,z,\lambda) + n^{-1/2} \alpha^{1/2} P_1(-\phi_{0,W}, \; \{\chi_v\})(y,z,\lambda)) \right\} d\lambda dy dz$$
$$+ n^{-1/2} \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha^{1/2} \sigma_G / \sigma_F^2 \; xz \phi_W(x,z,\lambda) dz d\lambda$$
$$- \frac{1}{2} n^{-1/2} \sigma_\tau \alpha^{-1/2} \int_{-\infty}^{x} \int_{-\infty}^{\infty} \int_{\infty}^{-\infty} \lambda \phi_W(y,z,\lambda) dz d\lambda dy$$
$$- \sigma(f)^{-1} (\phi_\nu + \gamma) n^{-1/2} \phi(x) + O(n^{-1}).$$

The control of the remainder is uniform over x because of the exponential bounds given in Lemma 6.6. Furthermore, some easy gaussian algebra yields

$$\int_{-\infty}^{x}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xz\phi_W(x,z,\lambda)dzd\lambda dy = x^2\phi(x)W_{1,2},$$
$$\int_{-\infty}^{x}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda\phi_W(x,z,\lambda)dzd\lambda = -W_{1,3}\phi(x).$$

Combining all the terms, the final expansion becomes

$$\begin{aligned}
&\Phi(x) - n^{-1/2}\alpha^{1/2}\frac{1}{6}M_{3,A}(x^2-1)\phi(x)\\
&+ n^{-1/2}\alpha^{1/2}\frac{1}{2}M_{3,A}x^2\phi(x) - n^{-1/2}\alpha^{-1/2}\beta/\sigma_F\, x^2\phi(x)\\
&+ \frac{1}{2}n^{-1/2}\alpha^{-1/2}\beta/\sigma_F\phi(x) - \sigma(f)^{-1}(\phi_\nu+\gamma)n^{-1/2}\phi(x)\\
&- n^{-1/2}\alpha^{-1/2}\beta/\sigma_F\phi(x) + n^{-1/2}\alpha^{-1/2}\beta/\sigma_F\phi(x)\\
&= \Phi(x) + n^{-1/2}\frac{1}{6}\alpha^{1/2}2(M_{3,A} - 3\alpha^{-1}\beta/\sigma_F)(x^2-1)\phi(x)\\
&+ \frac{1}{2}n^{-1/2}(M_{3,A} - 3\alpha^{-1/2}\beta/\sigma_F)\phi(x)\\
&- \alpha^{1/2}(\phi_\nu+\gamma)\phi(x)n^{-1/2} + n^{-1/2}\alpha^{-1/2}\beta\sigma_F^{-1}\phi(x)
\end{aligned}$$

and the result follows by recalling that

$$\sigma_F^2 = \alpha\sigma(f)^2$$

and using Proposition 3.1 for the form of the bias.

References

[1] Abramovitch, L., Singh, K (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* , **13**, pp 116-132.

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

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

[4] Bertail, P., Clémençon, S. (2003b). Approximate regenerative block-bootstrap for Harris recurrent Markov chains. *Submitted for publication.*

[5] Bolthausen, E. (1980).The Berry Essen Theorem for functionals of discrete Markov chains, *Z. Wahrsch. Verw. Gebiete*, **54**, pp. 59-73.

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

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

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

[9] Chung, , K.L. (1967). *Markov chains with stationary transition probabilities.* Springer-Verlag, Berlin, 2nd edition.

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

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

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

[13] Dubinskaite, I. (1982). Limit Theorems in R^k I. *Lithuanian Math. J.,* **22**, pp 129-140.

[14] Dubinskaite, I. (1984). Limit Theorems in R^k II. *Lithuanian Math. J.,* **24**, pp 256-265.

[15] Götze, F, and Hipp, C. (1983), Asymptotic expansions for sums of weakly dependent Random vectors. *Z. Wahr. verw. Gebiete*, **64**, pp 211-239.

[16] Hall, P. (1987). Edgeworth expansion for student's t statistic under minimal moment conditions. *Ann. Statist., 15,* pp. *920-931*

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

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

[19] Nagaev, S.V. (1961). More exact statements of limit theorems for homogeneous Markov chains. *Theory Probab. Appl.,* **6**, pp 62-81.

[20] Meyn, S.P., Tweedie, R.L., (1992). Stability of Markovian processes I: Discrete time chains. *Adv. Appl. Probab.,* **24**, pp 542-574.

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

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

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