On polynomial mixing and convergence rate for stochastic difference and differential equations

A.Yu.Veretennikov *
Institute of Information Transmission Problems
19 Bolshoy Karetnii, 101447 Moscow, Russia
email: ayu@sci.lpi.ac.ru

1991 Mathematics Subject Classification. 60F15, 60J05, 60J60.
Keywords. Mixing, recurrence, Markov process, SDE, polynomial convergence.

*This work was made when the author visited Weierstrass-Institut fuer Angewandte Analysis und Stochastik, Berlin; the work was supported by DFG grant 436 RUS 17/133/97 and by INTAS grants 93-1585 and 94-4194

Abstract

Polynomial bounds for β-mixing and for the rate of convergence to the invariant measure are established for discrete time Markov processes and solutions of SDEs under weak stability assumptions.

1 Introduction

We establish polynomial bounds for certain mixing coefficients (β-mixing and weakear ones) and convergence rate to the invariant measure for two clasees of Markov processes with discrete and continuous time in R^d. The first one is the class of processes which satisfy the equation and conditions

$$X_{n+1} = X_n + f(X_n) + W_{n+1},$$
$$(W_n) - \text{ i.}i.d.,\ EW_n = 0,\ E|W_n|^{m_0} < \infty,\ m_0 > 0, \tag{1}$$

and there exist such $r, C_0, M_0 > 0$ that

$$|x + f(x)| \le |x|(1 - r|x|^{-2}),\ \ |x| \ge M_0,$$
$$|x + f(x)| \le C_0,\ \ |x| \le M_0. \tag{2}$$

The second one consists of the processes in R^d which satisfy the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \tag{3}$$

where (W_t) is a d-dimensional Wiener process, Borel vector-function b is locally bounded and there exist such $r > 0,\ M_0 > 0$ that

$$(b(x), x) \le -r, \quad |x| \ge M_0. \tag{4}$$

We assume that there exists a homogeneous Markov solution of equation (3) which is unique in law. Some additional assumptions will be given later. For both equations we are interested in the ergodic case when there exists a unique invariant probability measure. Assumptions of theorems 1 and 5 imply this property.

Let us denote by X_t^x the solution of (1) or (3) with the initial data $x \in R^d$, $\mu_t^x = \mathcal{L}(X_t^x)$ - its marginal distribution, μ – the invariant probability measure. Remind the definition of β-mixing coefficient:

$$\beta^x(t) = \sup_{s\ge 0} E_x var_{F^X_{\ge t+s}}(P(B|F^X_{\le s}) - P(B)),$$

where F_I^X is a sigma-field generated by the values $\{X_s,\ s \in I\}$ and E_x means expectation for the process with the initial data x. Also, denote by $\beta(t)$ the average value $\int \beta^x(t)\mu(dx)$, i.e. the β-mixing coefficient for the stationary version of the process.

Our main goal is the estimates

$$var(\mu^x(t) - \mu) \le C(1 + |x|^m)(1 + t)^{-(k+1)} \tag{5}$$

with some $m > 0$, $k > 0$,

$$\beta^x(t) \le C(1 + |x|^m)(1 + t)^{-(k+1)}, \tag{6}$$

and a similar bound for $\beta(t)$,

$$\beta(t) \le C(1 + t)^{-(k+1)}. \tag{7}$$

The bounds (5)–(7) are useful in many applications if one can control all constants. Such bounds were established in Veretennikov (1997) for non-degenerate stochastic equation (3) under condition (4) if the constant r is large enough ($r > 3/2$ in the case $d = 1,\ \sigma \equiv 1$). In this paper we propose a different method to get similar estimates which does not use the nondegeneracy. Notice, however, that it requires more restrictive assumptions. It is interesting that previously we got two estimates (5) and (6) more or less simultaneously while now we first establish (5) and then deduce (6) as a corollary under more strong assumptions.

In the discrete time case the polynomial estimates of convergence and β-mixing were obtained earlier under more restrictive assumptions of the type

$$|x + f(x)| \le |x|(1 - |x|^{-1-\alpha}), \quad \alpha < 1,$$

see Tuominen, Tweedie (1994), Ango Nze (1994).

Remark. In Veretennikov (1997) this condition was shown with a wrong degree. It should be read as here.

A problem of polynomial estimates for certain hitting times like (8) below was considered and used in Lamperti (1963), Aspandiiarov, Iasnogorodski (1994), Aspandiiarov, Iasnogorodski, Menshikov (1996) and for martingales and SDEs in Menshikov, Williams (1996). Notice that (8) is not sufficient for the convergence and mixing without additional moment bounds for the process. We will get all bounds from (2) or (4). Some part of the calculus below resembles theorem 1 from Aspandiiarov, Iasnogorodski, Menshikov (1996). However, the setting and the main goal are different. In particular, we do not have exactly the inequality in the assumption of the theorem cited above.

We exploit a version of the approach which was used in Veretennikov (1987), Veretennikov (1997) and Gulinsky, Veretennikov (1993), chapter 5. It is based on the following bounds. Let $B_R = \{x \in R^d : \ |x| \le R\}$, $\tau = \tau_r = \inf(t \ge 1 : \ |X_t| \le R)$. The first bound we need is

$$E_x \tau^{(k+1)} \le C(1 + |x|^m) \tag{8}$$

with some $R > 0$, $k, m > 0$. In general, functions different from polynomials could also be used in the right hand side of (8). Other bounds we need are

$$\sup_{t \ge 0} E_x |X_t|^m 1(t \le \tau) \le C(1 + |x|^m) \tag{9}$$

together with

$$\int |x|^{m'} \mu(dx) < \infty \tag{10}$$

In previous works we established a slightly different estimate (cf. Veretennikov (1997))

$$\sup_t E_x |X_t|^m \le C(1 + |x|^m) \tag{11}$$

(after time change in some cases). One may notice that (11) implies (10) with $m' = m$.

2 Discrete time case: main results

Consider the process $(X_t, \ t = 0, 1, \ldots)$ which satisfies (1) and condition (2). Let $B \subset R^d$ and $\tau_1^B := \inf(t \ge 0 : \ X_t \in B)$ and $\tau_{n+1}^B := \inf(t \ge \tau_n^B + 1 : \ X_t \in B)$. If B is fixed, we will omit the index in τ. Denote also $\hat\tau = \hat\tau^B = \inf(t \ge 1 : \ X_t \in B)$. Define the "*process on B*", $X_t^B := X_{\tau_t^B}$. Denote by $P^B(n, x, dx')$ the n-step transition probability of (X_t^B).

We say that the *local Doeblin condition* holds true for the process (X_t) if for any $R \ge 0$ large enough the process on $B = B_R := (y \in R^d : \ |y| \le R)$ satisfies the following: there exists such integer $n_0 = n_0(R) > 0$ that

$$(D_l) \qquad \inf_{x,x'} \int \min\left\{\frac{P^B(n_0, x, dy)}{P^B(n_0, x', dy)}, 1\right\} P^B(n_0, x', dy) =: q(R, n_0) > 0,$$

where $P(dy)/P'(dy)$ means the derivative of the absolute continuous part of P w.r.t. P'. The singular part may also exist. The assumption (D_l) requires, in particular, that the singular part is not close to 1.

We assume that (X_t) satisfies the local Doeblin condition (D_l). Of course, this implies the irreducibility (see Meyn, Tweedie (1993)). Denote $r_m := (m-1)E|W_1|^2/2$.

Theorem 1 *Assume that the process (X_t) satisfies (1), (2) and the local Doeblin condition (D_l). If $m_0 \ge 2$, $r > r_m$ with $2 \le m \le m_0$ then the (unique) invariant measure μ exist and (9) is satisfied with this value m. If $m_0 > 2$, $2 < m \le m_0$, $r > r_2$ and $k < (m-2)/2$ then (8) holds true and (10) is satisfied with $m' = m-2$. If $m_0 > 4$, $2 < m \le m_0 - 2$ and $k < (m-2)/2$ then (5) holds true.*

Let $\mu^m(dy) := (1+|y|^m)\mu(dy)$, $\mu_t^{x,m}(dy) := (1+|y|^m)\mu_t^x(dy)$. For the sake of simplicity, we expose next results of this section under an assumption that r is large enough. In fact, one can formulate more precise conditions like $r > r'$ for some r' if needed. However, the formulations then become unreasonably complicated.

Theorem 2 *Let $m_0 > 4$ and (2) be satisfied. Then for any $2 < m < m_1 - 2 \le m_0 - 2$, $k < (m-2)/2$ there exist $C,\ r_0 > 0$ such that $r \ge r_0$ implies*

$$var(\mu_t^{x,m} - \mu^m) \le C(1+|x|^{m_1})(1+t)^{-(k+1)}. \qquad (12)$$

This estimate is a generalization of (5) under a more restrictive assumtion on the values r and m_0. It turns out that (12) may be rather helpful in applications. We now expose two straightforward corollaries of this bound.

Theorem 3 *Let $m_0 > 4$ and (2) be satisfied. Then for any $2 < m < m_1 - 2 \le m_0 - 2$ there exist $C,\ r_0 > 0$ such that $r \ge r_0$ implies*

$$\sup_t E_x|X_t|^m \le C(1+|x|^{m_1}). \qquad (13)$$

Moreover, for any $2 < m \le m_0 - 2$, $k < (m-2)/2$ there exist such $r,\ C > 0$ that inequalities (6), (7) hold true.

Of course, one can consider $m \le 2$ using Hölder's inequality. We mentioned above that the "usual" order is to get (13) and then to show (5). Here the order is inverse and assumptions for (13) are more severe.

The following corollary is a rather partial result of the CLT type. However, it can be helpful, in particular, in certain problems of a non-parametric estimation for a nonlinear autoregression models in statistics. Also notice that this is an example of the direct use of mixing bounds established for a certain class of processes. Usual situation is that one either assumes such bounds or it is easier to proceed differently, avoiding the use of mixing.

Let $\xi_k = h(X_k)$ where the Borel function h is s.t. $|h(x)| \le C_h(1+|x|^2)$, $S_n^h := \sum_{i=0}^{n-1}(\xi_k - \bar{h})$ where $\bar{h} = \int h(x)\mu(dx)$; the last value is finite.

Theorem 4 *Let $m_0 > 4$ and r is large enough. Then S_n satisfies the CLT, i.e.*

$$n^{-1/2} S_n \Rightarrow \mathcal{N}(0, s^2), \tag{14}$$

where the variance $s^2 = \sum_{i=0}^{\infty} cov(\xi_0, \xi_i)$ is non-negative and finite; "cov" means the covariance calculated in the stationary regime. Moreover, for all functions h and f satisfying given constraints with fixed constants C_h, M_0, C_0, $C_W := E|W_1|^{m_0}$, for $r \ge r(C_h, M_0, C_0, C_W)$, the variance s^2 is uniformly bounded, and the sequence $(n^{-1/2}S_n)$ is uniformly tight.

Remark. Apparently, one can establish such a tightness without the CLT, i.e. perhaps, not using the assumtion (D_l) nor irreducibility.

3 Continuous time case: main results

There is no "natural" m_0 for the equation (3), or in some sense we can let $m_0 = +\infty$. Hence, a stability assumption only deals with the value r from (4) and the coefficients of the equation, as in Veretennikov (1997).

Assume that the coefficient b is locally bounded, the matrix $d \times d$ function σ is continuous and function $a = \sigma\sigma^*$ is uniformly bounded. Condition (4) is always assumed which suffices to guarantee the non-exposion case, see below. Repeating notations from Veretennikov (1997), let

$$\lambda_+ = \sup_{x \ne 0} (a(x)\frac{x}{|x|}, \frac{x}{|x|}), \quad \Lambda = \sup_x \frac{Tr\ a(x)}{d}$$

and define a new constant

$$r(m) = (d\Lambda + (m-2)\lambda_+)/2, \quad m \ge 2.$$

The difference from Veretennikov (1997) is that σ may degenerate. However, we assume that there exists a solution of equation (3) which is unique in law, at least, locally. The global uniqueness will then follow from the non-explosion. Moreover, we always consider the case of a unique ergodic class. Note that $\lambda_+ = 0$ does not mean $\sigma \equiv 0$ if $d > 1$. We also assume that for "the process inside any B_R" the local Doeblin condition (D_l) is satisfied. One can provide this assumption for the case of SDEs either by a nondegeneracy of the diffusion coefficient or a Hörmander type conditions.

Moreover, we assume that for any $R > 0$,

$$(T) \qquad \sup_{|x|=R} E_x T < \infty,$$

where $T := \inf(t \ge 0 : \ |X_t| \ge R+1)$. This property can also be provided by a nondegeneracy or a Hörmander type condition.

Theorem 5 *Let assumptions (4), (D_l) and (T) be satisfied. If $r > r(2)$ then there exists a (unique) invariant measure μ and (9) holds with $m = 2$. If $2 < m$, $r > r(m)$ then (9) is satisfied with this value m and the estimate (10) holds true with $m' = m - 2$. If $r > r(m)$, $k \in (0, (m-2)/2)$ then the process satisfies (8). If $m_0 > 4$, $2 < m \le m_0$, $r > r(m)$, $k < (m-2)/2$ then inequality (5) holds true.*

Theorem 6 *Let (4), (D_l) and (T) be satisfied. Then for any $m,\ m'\ , k > 0$, $m_1 > m+2$, there exist such $r_0,\ C > 0$ that if $r \ge r_0$ then estimates (8)–(10), (12), (13) and (6), (7) hold true.*

Remark. In Veretennikov (1997) the formula displayed above (28) should be read as

$$(a_{ij}p)_{x_i x_j} - (b_i p)_{x_i} = 0, \ (\tilde{a}_{ij}\tilde{p})_{x_i x_j} - (\tilde{b}_i \tilde{p})_{x_i} = 0.$$

Also, it should be noted that these two equalities are understood in the weak sense. We argued in Veretennikov (1997) that two probability densities which satisfy one of those equations, say, the first one, necessarily coincide. This is still true for weak solutions, so that one need not assume any additional regularity conditions on coefficients. Indeed, even in the weak form, each equation just means that the invariance of the measure for a certain diffusion process. Such invariant measure is unique because of the nondegeneracy assumption.

4 Discrete time case: preliminary results and proofs

We will establish four lemmas. Then the proof of the theorem will follow from lemma 4 by the considerations in section 4 of Veretennikov (1997). Though in that paper ergodic diffusion processes were considered, the calculus of section 4 is valid for any Markov process satisfying certain estimates. These estimates will be established, in fact, in lemma 4 below. For the sake of simplicity of exposition, we assume condition (D_l) to be satisfied with $n_0 = 1$. Changes needed for $n_0 > 1$ are obvious.

Lemma 1 *Let (1) and (2) be satisfied with $m_0 \geq 2$ and $r > r_2$. Then for R large enough there exists such $C > 0$ that*

$$E_x \tau \leq C(1 + |x|^2) \tag{15}$$

and (9) is satisfied with $m = 2$. If, moreover, $m_0 > 2$, $2 \leq m \leq m_0$ and $r > r_m$ then inequality (9) holds true.

In particular, the process is positive recurrent w.r.t. B_R, at least, if R is large enough. We denote by F_n the sigma-field generated by X_n.

Proof. Let us consider the value $E_x |X_{n+1}|^2 1(n < \tau)$ (we omit the index in $\tau = \tau_R$. We have,

$$\begin{aligned} &E_x |X_{n+1}|^2 1(n < \tau) \\ &= E_x |X_n + f(X_n) + W_{n+1}| 1(n < \tau) \\ &\leq E_x 1(n < \tau) E[|X_n + f(X_n) + W_{n+1}|^2 \, |X_n] \end{aligned}$$

Denote

$$\zeta = (X_n + f(X_n)), \quad \Delta = W_{n+1}.$$

Then

$$|\zeta+\Delta|^2 = |\zeta|^2 + 2(\sum_{k=1}^{d}\zeta^k\Delta^k) + (\sum_{k=1}^{d}(\Delta^k)^2).$$

So one has,

$$\begin{array}{c} 1(n<\tau)E[|X_n+f(X_n)+W_{n+1}|^2\, X_n] \\ \le 1(n<\tau)\left(|\zeta|^2 + 2(\sum_{k=1}^{d}\zeta^k E\Delta^k) + (\sum_{k=1}^{d}E(W_{n+1}^k)^2)\right) \\ \le 1(n<\tau)\left\{(1-r|X_n|^{-2})^2 + |X_n|^{-2}2r_2\right\} \\ \le 1(n<\tau)\left\{1-2r'|X_n|^{-2}+2r_2|X_n|^{-2}\right\}, \end{array}$$

where the new constant r' in the last inequality is $r' < r$ and it may be chosen arbitrary close to r if we take R large enough, so that, in particular, $r' > r_2$. Then we get with some $c_2 > 0$,

$$E_x|X_{n+1}|^2 1(n<\tau) \le E_x|X_n|^2 1(n<\tau)(1-c_2|X_n|^{-2})).$$

This implies (9) with $m=2$ by induction. Now we choose $R > c_2^{1/2}$ so that with some new constant $c>0$ one has,

$$E_x|X_{n+1}|^2 1(n+1<\tau) \le E_x|X_n|^2 1(n<\tau) - cE_x 1(n<\tau). \qquad (16)$$

This implies,

$$c^{-1}E_x\tau \le 1+|x|^2.$$

Consider the case $m_0 > 2$, $r > r_m$. Let $2 < m \le m_0$. By Taylor's formula,

$$\begin{array}{c} |\zeta+\Delta|^m = |\zeta|^m + m|\zeta|^{m-2}(\zeta,\Delta) \\ +(m/2)|\Delta|^2|\zeta+s\Delta|^{m-2} + (m/2)(m-2)|(\zeta+s\Delta,\Delta)|^2|\zeta+s\Delta|^{m-4} \\ \le |\zeta|^m + m|\zeta|^{m-2}(\zeta,\Delta) \\ +(m/2)[1+(m-2)]|\Delta|^2(1+\varepsilon)|X_n+f(X_n)|^{m-2} + C_{\varepsilon,m}|\Delta|^m. \end{array}$$

Here $s\in[0,1]$; the constant $\varepsilon>0$ may be taken arbitrary small. Since $E\Delta=0$, we get for R large enough,

$$\begin{array}{c} E[1(n<\tau)|\zeta+\Delta|^m|\zeta] \le 1(n<\tau)|X_n|^m(1-r|X_n|^{-2})^m \\ +m(1+\varepsilon)r_2 1(n<\tau)|X_n|^{m-2} + C_{\varepsilon,m}1(n<\tau)E|\Delta|^m. \end{array}$$

Let $r > r' > r_m$. Then there exists $R > 0$ s.t.

$$E_x\{1(n < \tau)|X_{n+1}|^m|F_n\} \le 1(n < \tau)|X_n|^m(1 - r'm|X_n|^{-2})$$
$$+m(1+\varepsilon)r_2 1(n < \tau)|X_n|^{m-2} + C_{\varepsilon,m}1(n < \tau)E|\Delta|^m.$$

This implies, again for R large enough,

$$E_x 1(n+1 \le \tau)|X_{n+1}|^m + C^{-1}E_x 1(n < \tau)|X_n|^{m-2}$$
$$\le E_x 1(n < \tau)|X_n|^m \le E_x 1(n \le \tau)|X_n|^m. \tag{17}$$

By induction, this gives (9). It follows from Hölder's inequality that the same holds true with any $m \le m_0$. Lemma is proved.

Notice that to get (16) we used, in fact, $|X_n| > R$ rather than $n < \tau$.

By virtue of the Harris theorem (see Meyn, Tweedie (1993), theorem 10.2.1) positive recurrent process (X_t) possesses a unique invariant probability measure μ.

Lemma 2 *Let $m_0 \ge 2$, $r > r_{m_0}$. Then*

$$\int |x|^{m_0-2}\mu(dx) < \infty. \tag{18}$$

Proof. It suffices to show (18) for $m_0 > 2$. By induction, we get from (17), for $2 \le m \le m_0$, $|x| > R$,

$$E_x \sum_{k=0}^{\tau-1} |X_k|^{m-2} \le C|x|^m. \tag{19}$$

Now we can use a Harris representation for the invariant measure. Consider the "process X on $B = B_R$". Because of the local Doeblin assumption for X, this process possesses an invariant measure μ^B. Now, the invariant measure of (X_t) is equal, with some constant $c(B)$, to

$$\mu(A) = c(B)^{-1}\int_B \mu^B(dx)E_x \sum_{k=1}^{\hat\tau} 1(\xi_k \in B) \tag{20}$$

(see Meyn, Tweedie (1993), proposition 10.4.8). The constant $c(B)$ is equal to $\int_B \mu^B(dx)E_x\hat\tau$ which is finite and not less than 1. It follows from (20) that for any positive function h,

$$\int h(x)\mu(dx) = c(B)^{-1}\int_B E_x \sum_{k=1}^{\hat\tau} h(X_k). \qquad (21)$$

Since $E_x|X_1|^m \le C(1+|x|^m)$ for $m \le m_0$, we estimate,

$$\int |x|^{m-2}\mu(dx) = c(B)^{-1}\int_B \mu^B(dx)E_x\sum_{k=1}^{\hat\tau^z}|X_k|^{m-2}$$
$$\le C\int_B \mu^B(dx)CE_x(1+|X_1|^m) \le C\int_B \mu^B(dx)C(1+|x|^m) < \infty$$

which gives (18). Lemma is proved.

Lemma 3 *Let $m_0 > 2$, $r > r_m$, $2 < m \le m_0$, $0 < k < (m-2)/2$. Then for R large enough,*

$$E_x\tau^{k+1} \le C_{k,m}(1+|x|^m).$$

Notice that if $r > r_2$ then also $r > r_m$ for some $r > 2$.

Proof. We use inequality (16), this time multiplying it by $(n+2)^k$. Using the identity $n+2 = (n+1)(1+(n+1)^{-1})$, one estimates,

$$(n+2)^k E_x|X_{n+1}|^2 1(n<\tau) \le (n+2)^k E_x|X_n|^2 1(n<\tau)(1-c_2|X_n|^{-2})$$
$$\le (n+1)^k E_x|X_n|^2 1(n<\tau)(1-c_2|X_n|^{-2})(1+1/(n+1))^k$$
$$\le (n+1)^k E_x|X_n|^2 1(n<\tau)\,(1-c_2|X_n|^{-2})(1+k/(n+1)+C_k/(n+1)^2)$$
$$\le (n+1)^k E_x|X_n|^2 1(n<\tau)\,(1-c_2|X_n|^{-2}+C'_k/(n+1)).$$

Let us use the identity $1 = 1(|X_n|^2 \le c(n+1)) + 1(|X_n|^2 > c(n+1))$ choosing here $c < c_2/C'_k$. Then $(C'k(n+1)^{-1} - c_2|X_n|^{-2})\,1(|X_n|^2 \le c(n+1)) \le -C|X_n|^{-2}1(|X_n|^2 \le c(n+1))$. So we estimate,

$$(n+2)^k E_x|X_{n+1}|^2 1(n<\tau) - (n+1)^k E_x|X_n|^2 1(n<\tau)$$
$$\le (n+1)^k E_x|X_n|^2 1(n<\tau)(-c_2|X_n|^{-2} + C'_k/(n+1))1(|X_n|^2 \le c(n+1))$$
$$+(n+1)^k E_x|X_n|^2 1(n<\tau)(-c_2|X_n|^{-2} + C'_k/(n+1))1(|X_n|^2 > c(n+1))$$
$$\le -C(n+1)^k E_x 1(n<\tau)1(|X_n|^2 \le c(n+1))$$
$$+(n+1)^k E_x|X_n|^2 1(n<\tau)(-c_2|X_n|^{-2} + C'k/(n+1))1(|X_n|^2 > c(n+1))$$
$$\le C(n+1)^k E_x|X_n|^2 1(n<\tau)(n+1)^{-1}1(|X_n|^2 > c(n+1))$$
$$-C(n+1)^k E_x 1(n<\tau).$$

Note that $|X_n|^{-2}1(n < \tau) \le R^{-2}1(n < \tau)$. We take R so large that $C|X_n|^{-2}1(n<\tau) \le 1(n<\tau)$. Then one obtains,

$$\begin{aligned}
&C^{-1}(n+1)^k E_x 1(n<\tau) \\
&\le \Big\{(n+1)^k E_x|X_n|^2 1(n<\tau) - (n+2)^k E_x|X_{n+1}|^2 1(n+1<\tau)\Big\} \\
&\quad + C(n+1)^k E_x|X_n|^2(n+1)^{-1}1(n<\tau)1(|X_n|^2 > c(n+1)).
\end{aligned}$$

We estimate,

$$\begin{aligned}
&(n+1)^k E_x|X_n|^2(n+1)^{-1}1(n<\tau)1(|X_n|^2 > c(n+1)) \\
&\le E_x|X_n|^m 1(n<\tau)(n+1)^{k-1-(m_1-2)/2}.
\end{aligned} \tag{22}$$

We can take any $k \in (0,(m-2)/2)$. For such values k we have,

$$\sum_{n=0}^{\infty} C(n+1)^{k-1-(m_1-2)/2}(1+|x|^m) \le C(1+|x|^m).$$

Hence,

$$E_x\sum_{n=0}^{\tau-1}(n+1)^k = \sum_{n=0}^{\infty}(n+1)^k E_x 1(n<\tau) \le C(1+|x|^m).$$

Therefore,

$$E_x\tau^{k+1} \le C(1+|x|^m).$$

Lemma is proved.

Notice that the same calculus could be made for $E_x|X_{n+1}^{m'}1(n<\tau)$ with any $2<m'<m$.

Now let us consider the direct product of two identical probability spaces where two (independent) copies of our Markov process (X_t) and (X'_t) with initial data $X_0 = x_0$, $X'_0 = x'_0$ are defined. We will not change notations for probability and expactation. Let $R_1 > R$, $\gamma \equiv \gamma_{R_1} = \inf(t \ge 0 : \max(|X_t|,|X'_t|) \le R_1)$, $\gamma(t) = \min(\gamma, t)$.

Lemma 4 *Let $m_0 > 2$, $2 < m \le m_0$, $r > r_m$. Then there exist such R_1, $C > 0$ that*

$$\begin{aligned}
&E_{x,x'}1(n+1) \le \gamma(t))(|X_{n+1}|^m + |X'_{n+1}|^m)) \\
&+C^{-1}E_{x,x'}1(n<\gamma(t))(|X_n|^{m-2} + |X'_n|^{m-2}) \\
&\le E_{x,x'}1(n \le \gamma(t))C(1+|x|^m + |x'|^m).
\end{aligned} \tag{23}$$

Proof. Repeating the calculus of lemma 1 with $m > 2$, we obtain for $|X_n| > R$ and $|X'_n| > R$, respectively,

$$E[|X_{n+1}|^m + C^{-1}|X_n|^{m-2}|X_n] \le |X_n|^m,$$

and

$$E[|X'_{n+1}|^m + C^{-1}|X'_n|^{m-2}|X'_n] \le |X'_n|^m.$$

Moreover, if $|X_n| \le R$ then

$$E[|X_{n+1}|^m + C^{-1}|X_n|^{m-2}|X_n] \le P(R),$$

where $P(R)$ is some polynomial of R. If $n < \gamma(t)$ then $|X_n| \le R$ implies $|X'_n| > R_1$. Let us choose $R_1 > R$ s.t. $(1/2)C^{-1}R_1^{m-2} > P(R)$. Then we get (23) with a new C by induction. Lemma is proved.

Lemma 5 *Then there exist such* R_1, $C > 0$ *that*

$$E_{x,x'}\gamma^{k+1} \le C(1 + |x|^m + |x'|^m). \tag{24}$$

Proof. Similarly to lemma 3, we obtain,

$$\begin{gathered}
(n+2)^k E_x(|X_{n+1}|^m + |X'_{n+1}|^m)1(n < \gamma(t)) \\
-(n+1)^k E_x(|X_n|^m + |X'_n|^m)1(n < \gamma(t)) \\
\le C(n+1)^k E_x|X_n|^m 1(n < \gamma(t))(n+1)^{-1}1(|X_n|^2 > c(n+1)) \\
+C(n+1)^k E_x|X'_n|^m 1(n < \gamma(t))(n+1)^{-1}1(|X'_n|^2 > c(n+1)) \\
-C(n+1)^k E_x(|X_n|^{m-2} + |X'_n|^{m-2})1(n < \gamma(t)).
\end{gathered}$$

Now we take R_1 so large that $(|X_n|^{m-2} + |X'_n|^{m-2}) \ge 1$ if $n < \gamma(t)$. Then wwe get,

$$\begin{gathered}
C^{-1}(n+1)^k E_{x,x'}1(n < \gamma(t)) \\
\le \Big\{(n+1)^k E_{x,x'}(|X_n|^m + |X'_n|^m)1(n < \gamma(t)) \\
-(n+2)^k E_{x,x'}(|X_{n+1}|^m + |X'_{n+1}|^m)1(n+1 < \gamma(t))\Big\} \\
+C(n+1)^k E_{x,x'}(|X_n|^m(n+1)^{-1}1(|X_n|^2 > c(n+1)) \\
+(|X'_n|^m(n+1)^{-1}1(|X'_n|^2 > c(n+1))1(n < \gamma(t)).
\end{gathered}$$

The rest of the proof is similar as for lemma 3 after (22). Lemma 5 is proved.

Proof of theorem 1. Consider the couple of independent copies of our Markov process, (X_t, Y_t) with initial values $X_0 = x \in R^d$, Y_0 distributed with the invariant measure μ^{inv}. Fix $s_0 = 0$ (later we will use $s_0 \geq 0$, hence, we need this notation). Define the sequence of stopping-times $\gamma_1 < \gamma_2 < \dots$ as follows:

$$\gamma_1 = \inf(t \geq s_0 : \ |X_t| \leq R \text{ and } |Y_t| \leq R),$$

for $n \geq 1$

$$T_n = \inf(t \geq \gamma_n : \ |X_t| \geq R+1 \text{ or } \ |Y_t| \geq R+1);$$

$$\gamma_{n+1} = \inf(t \geq T_n : \ |X_t| \leq R \ \text{ and } \ \ |Y_t| \leq R).$$

We have due to lemma 4,

$$E((\gamma_1 - s_0)^{k+1} | \hat{F}_{s_0}) \leq C(1 + |Y_{s_0}|^m + |X_{s_0}|^m).$$

Similarly,

$$E((\gamma_{n+1} - \gamma_n)^{k+1} | \hat{F}_{\gamma_k}) \leq C, \quad n \geq 1.$$

Let $n(t) := \sup(n \geq 0 : \ \gamma_n \leq t)$. By virtue of the last inequality and a strong markovian property of (X_t, Y_t), one gets

$$P(n(t) \to \infty, \ t \to \infty) = 1.$$

Using a coupling method (cf. Nummelin (1984)) it is possible to define a new process $(\tilde{X}_t)$ and a random value $L = L_{s_0} \geq s_0$ on a certain extension of the probability space (Ω, F, P) (we do not change the notation for probability and expectation) which is equivalent in distribution to the process (X_t) and, moreover,

$$P(\tilde{X}_t = X_t, \, t \leq L_{s_0}) = P(\tilde{X}_t = Y_t, \ t \geq L_{s_0}) = 1,$$

and L_{s_0} is a $\hat{F}_t \equiv F_t^{X,Y,\tilde{X}}$-stopping time. Moreover, due to the local Doeblin condition and the markovian propety, there exists $q \in (0,1)$ s.t.

$$\sup_{s_0 \geq 0} P(L_{s_0} > \gamma_n | \hat{F}_{s_0}) \leq q^n \quad \forall n.$$

We omit the index $s_0 = 0$ in the rest of the proof. We have,

$$var(\mu_t^x - \mu) \leq P(L > t) = \textstyle\sum_{n=0}^{\infty} E(I(L > t) I(\gamma_n \leq t < \gamma_{n+1}))$$
$$\leq \textstyle\sum_{n=0}^{\infty} P(L > \gamma_n)^{1/a} P(\gamma_{n+1} > t)^{1/c} \leq \sum_{n=0}^{\infty} q^{n/a} P(\gamma_{n+1} > t)^{1/c}$$

(we used Hölder's inequality with $a^{-1}+c^{-1}=1$, $a>1$, $c>1$). Due to Bienaimé-Chebyshev's inequality, one gets

$$\begin{aligned} P(\gamma_{n+1}>t) &\le t^{-(k+1)}E((\gamma_{n+1})^{k+1} \\ &\le t^{-(k+1)}(n+1)^k \textstyle\sum_{j=0}^n E((\gamma_{j+1}-\gamma_j)^{k+1}) \\ &\le t^{-(k+1)}(n+1)^k(C(1+|X_0|^m+E|Y_0|^m)+Cn). \end{aligned} \tag{25}$$

Therefore, due to lemma 5,

$$P(L>t) \le \sum_{n\ge 0} q^{n/a}[t^{-(k+1)}(n+1)^k(C(1+|x|^m)+Cn)]^{1/c}. \tag{26}$$

For any $\nu>0$ there exist such c close to 1 and $C<\infty$ that

$$P(L>t) \le Ct^{-(k+1-\nu)}(1+|x|^m).$$

This proves theorem 1.

Remark. In fact, we showed, as an auxiliary step, the inequality

$$var(\mu^x(t)-\mu_t^{x'}) \le C(1+|x|^m+|x'|^m)(1+t)^{-(k+1)}$$

under assumptions $m_0>2$, $2<m<m_0$, $k<(m-2)/2$. We need $m_0>4$ when we integrate w.r.t. μ.

Proof of theorem 2. Let $m<m_0-2$. We estimate, using the notations from the previous proof,

$$\begin{aligned} |\mu_t^{x,m}(B)-\mu^m(B)| &= |E_x(1+|X_t|^m)1(X_t\in B)-E(1+|\tilde X_t|^m)1(\tilde X_t\in B)| \\ &\le |E_x(1+|X_t|^m)1(X_t\in B)1(t<L)+E(1+|\tilde X_t|^m)1(\tilde X_t\in B)1(t<L). \end{aligned}$$

Due to the estimate (10),

$$\begin{aligned} & E(1+|\tilde X_t|^m)1(\tilde X_t\in B)1(t<L) \\ & \le \left(E(1+|\tilde X_t|^{m_0-2})\right)^{m/(m_0-2)}(P(t<L))^{1-m/(m_0-2)} \\ & \le C(1+t)^{-(k+1)(1-m/(m_0-2))}, \end{aligned}$$

where k may be taken arbitrary large. It remains to estimate the similar term with X_t. We have,

$$\begin{aligned} & E_x(1+|X_t|^m)1(X_t\in B)1(t<L) \\ & \le E_x(1+|X_t|^m)\textstyle\sum_{i=0}^\infty 1(\gamma_i\le t<\gamma_{i+1}))1(\gamma_i<L)1(L>t). \end{aligned}$$

Due to Hölder's inequality, we estimate with $a, b, c > 1$, $a^{-1}+b^{-1}+c^{-1} = 1$, $a \le m_0/m$,

$$
\begin{aligned}
&E_x(1+|X_t|^m)1(\gamma_i \le t < \gamma_{i+1})1(L > \gamma_i)1(L > t) \\
&\le C\left(E_x(1+|X_t|^{am})1(\gamma_i \le t < \gamma_{i+1})\right)^{1/a} \times \\
&\times P_x(L > \gamma_i)^{1/b} P_x(L > t)^{1/c}.
\end{aligned}
$$

Remind that $P(L > \gamma_i) \le q^i$ with some $q < 1$ and $P_x(L > t) \le C(1 + |x|^m)(1+t)^{-(k+1)}$, where k may be taken arbitrary large. Further, for any $i \ge 1$ one has,

$$
\begin{aligned}
&E_x[(1+|X_t|^{am})1(\gamma_i \le t < \gamma_{i+1})] \\
&\le E_x 1(\gamma_i \le t) E[(1+|X_t|^{am})1(t < \gamma_{i+1})|F_{\gamma_i}].
\end{aligned}
$$

Moreover, for any $|x| \le R$,

$$E_x(1+|X_t|^{am})1(t < \gamma_1) \le C(1+|x|^{am})$$

due to lemma 4. Hence, for any $i \ge 1$ we obtain,

$$E_x(1+|X_t|^{am})1(\gamma_i \le t < \gamma_{i+1}) \le C < \infty.$$

Also, by lemma 4,

$$E_x(1+|X_t|^{am})1(\gamma_0 \le t < \gamma_1) \le C(1+|x|^{am}).$$

Thus,

$$\sum_{i=0}^{\infty} E_x(1+|X_t|^m)1(\gamma_i \le t < \gamma_{i+1})1(L > \gamma_i) \le \sum_{i=0}^{\infty} q^{i/b} C(1+|x|^m) = C(1+|x|^m).$$

Finally, we get,

$$E_x(1+|X_t|^m)1(X_t \in B)1(L > t) \le C(1+|x|^m)(1+t)^{-(k+1)/c}.$$

Since k may be arbitrary, this implies the assertion of theorem 2.

Proof of theorem 3. Due to theorem 2,

$$var\mu_t^{x,m} \le var\mu^m + var(\mu_t^{x,m} - \mu^m)$$

which gives (13).

To get estimates (6) and (7), we repeat the calculus in the proof of theorem 1 with any $s_0 \geq 0$. We have,

$$\beta^x(t) \leq \sup_{s_0} E_x P(L_{s_0} > t + s_0 | F_{s_0}) \leq \sup_{s_0} E_x C(1 + |X_{s_0}|^m + |Y_{s_0}|^m).$$

Now, the estimates (13) and (10)) implies both (6) and (7). Theorem 3 is proved.

Proof of theorem 4. For stationary regime, the weak convergence follows from the CLT due to Ibragimov, Linnik (1971), Indeed, the r.v. ξ_i possess moments of order $2+\delta$ with some $\delta > 0$, and the rate of β-mixing is faster than any polynomial. The covariance $s^2 = \sum_{k=0}^{\infty} cov(\xi_0, \xi_k)$ is finite. Either $s^2 > 0$, or $s^2 = 0$. If $s^2 > 0$ then $var(S_n) = s^2 n(1 + o(1))$. In this case we get the CLT by virtue of theorem 18.4.2 from Ibragimov, Linnik (1971). If $s^2 = 0$ then it follows that $sup_n var(S_n) < \infty$. Then obviously there is a convergence to 0 which is considered as a degenerate Gaussian random value. The uniform tightness follows easily from estimates for the expectation and covariance of $n^{-1/2} S_n$ which both turn out to be uniformly bounded because of the bounds for $\beta(t)$.

For a non-stationary regime we conclude using the convergence of μ_t^x in variation, or directly by the coupling inequality. Theorem 4 is proved.

Remark. Let us say a few words concerning the constants m in the bounds. We use the estimate $c(B) \geq 1$ in the Harris representation (20). For non-degenerate diffusions this does not give an optimal bound (cf. Veretennikov (1997)). On the other hand, this estimate holds true in degenerate cases as well. Perhaps one can improve the estimates using more fine bounds for $\hat{\tau}$.

5 Continuous time case: proofs

As in Veretennikov (1997), let us apply the Itô formula to the process $(1 + t)^k |X_t|^m$ as $t < \tau$, $k > 0$, $m \geq 2$. We have,

$$E(1+\tau)^k |X_\tau|^m - |x|^m$$
$$= E \int_0^t (1+s)^k |X_s|^m [k(1+s)^{-1} + m|X_s|^{-2}(X_s, b(X_s))$$

$$+(1/2)m|X_s|^{-2}\,Tr\,a(X_s) + (1/2)m(m-2)|X_s|^{-4}(a(X_s)X_s, X_s)] \times$$
$$\times\{I(|X_s|^2 \le c(1+s)) + I(|X_s|^2 > c(1+s))\}1(s<\tau)\,ds$$
$$\le E\int_0^t (1+s)^k|X_s|^m[k(1+s)^{-1} - rm|X_s|^{-2} + mr(m)|X_s|^{-2}] \times$$
$$\times 1(s<\tau)1(|X_s|^2 \le c(1+s))\,ds$$
$$+E\int_0^t (1+s)^k|X_s|^m[k(1+s)^{-1} - rm|X_s|^{-2} + mr(m)|X_s|^{-2}] \times$$
$$\times 1(s<\tau)1(|X_s|^2 > c(1+s))\,ds.$$

We choose c s.t. $m(r - r(m)) > c^2 k$. Then we obtain,

$$\left\{E(1+\tau)^{k+1} + E_x(1+\tau)^k|X_\tau|^m\right\} \le C(1+|x|^m)$$
$$+CE\int_0^t (1+s)^k|X_s|^m[(1+s)^{-1} + |X_s|^{-2}/2]I(|X_s|^2 \le c(1+s))1(s<\tau)\,ds$$

This gives the inequality

$$E(1+\tau)^{k+1} + E_x(1+\tau)^k|X_\tau|^m \le C(1+|x|^m) \qquad (27)$$

which is the analogue of the estimates of lemmas 1 and 3 from section 3.

The next step is (10). To show this bound, we use the Harris representation. To this end, some additional construction is needed for SDE case. Let $T_1 = \tau_R$, $T_1' = \inf(t \ge 0 : \ |X_t| \ge R+1$, and by inducton, $T_{n+1} = \inf(t \ge T_n' : \ |X_t| \le R$, $T_{n+1}' = \inf(t \ge T_{n+1} : \ |X_t| \ge R+1$, $n = 1, 2, \ldots$.

Denote $Z_n = X_{T_n'}$. Then (Z_n) is a Markov process which satisfies the Doeblin type condition due to our assumptions. So it has a stationary measure μ' on $B = \{|y| = R+1\}$. Then

$$\mu(A) = c(B)^{-1}\int_B \mu'(dx)E_x\int_0^{T_1'} 1(X_s \in A)\,ds \qquad (28)$$

where $c(B) = \int_B \mu'(dx)E_x T_1'$, see Has'minski (1980). The value $c(B)$ is finite due to assumption (T) and (27). Hence,

$$\int h(y)\mu(dy) = c(B)^{-1}\int_B \mu'(dx)E_x\int_0^{T_1'} h(X_s)\,ds. \qquad (29)$$

We have,

$$E_x\int_0^{T_1'} h(X_s)\,ds = E_x\int_0^{T_1} h(X_s)\,ds + E_x\int_{T_1}^{T_1'} h(X_s)\,ds$$
$$\le C(1+R^m) + C\sup_{|x|+R} E_x T \le C(1+R^m).$$

This shows (10).

Now we can repeat the same considerations which led us to lemma 4. The rest of the proof is identical to that of theorem 1. Theorem 5 is proved.

Notice that, in fact, in the few sentences above we meant that we used the same definition of the sequence (γ_i) as in the previous section, and then assumption (T) is enough for our aims. One could use directly (T_i') instead. In this case (T) should be replaced by $\sup_{|x|=R} E_x T^k < \infty \ \forall k > 0$ which is, formally speaking, more restrictive. However, "usually" one can provide this condition by more or less the same hypotheses like Hörmander ones or nondegeneracy.

Finally, all assertions of theorem 6 follows from the same considerations as in the proofs of theorems 2 and 3.

Acknowledgements The author is thankful to O.V.Lepsky who suggested a question concerning the tightness in the CLT scheme for the nonlinear autoregression process.

6 References

P. Ango Nze (1994). Critéres d'ergodicité de modéles markoviens. Estimation non parametrique sous des hypothéses de dependance. *Thése de Doctorat de Mathématiques Appliquées.* (Univ. Paris IX – Dauphine, Paris).

S. Aspandiiarov, R. Iasnogorodski (1994). General criteria of integrability of functions of passage-times for non-negative stochastic processes and their applications, *preprint.*

S. Aspandiiarov, R. Iasnogorodski, M. Menshikov (1996). Passage-time moments for non-negative stochastic processes and an application to reflected random walks in a quadrant.*Ann. Probab.*__24__(2) 932-960.

O. V. Gulinsky, A. Yu. Veretennikov (1993). Large deviations for discrete-time processes with averaging. VSP, Utrecht, The Netherlands.

R. Z. Has'minski (1980). Stochastic stability of differential equations. Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, and Rockville, Maryland, USA.

I.A.Ibragimov, Yu.V.Linnik (1971). Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen.

J. Lamperti (1963). Criteria for stochastic processes II: passage-time moments.*J.Math.Anal.Appl.*__7__ 127-145.

M. Menshikov, R. J. Williams (1996). Passage-time moments for continuous non-negative stochastic processes and applications.*Adv.Appl.Prob.***28** 747-762.

S. P. Meyn, R. L. Tweedie (1993). Markov chains and stochastic stability. Springer-Verlag, Berlin et al.

E. Nummelin (1984). General irreducible Markov chains and non-negative operators. Cambridge University Press.

P. Tuominen, R. L. Tweedie (1994). Subgeometric rates of convergence of f-ergodic Markov chains.*Adv.Appl.Prob.***26** 775-798.

A. Yu. Veretennikov (1987). Bounds for the mixing rate in the theory of stochastic equations.*Theory Probab.Appl.***32**(2) 273-281.

A. Yu. Veretennikov (1997). On polynomial mixing bounds for stochastic differential equations.*Stoch. Processes Appl.***70** 115-127.