ON EXPONENTIAL BOUNDS FOR MIXING AND
THE RATE OF CONVERGENCE FOR STUDENT PROCESSES

UDC 519.21

N. ABOURASHCHI AND A. YU. VERETENNIKOV

ABSTRACT. Exponential bounds for the \( \beta \)-mixing coefficient are established for the Student diffusion process. The latter is a Markov diffusion process with Student distribution as a stationary measure. The method is based on a direct estimation of moments and on polynomial Lyapunov functions for evaluating exponential functionals of hitting times.

1. INTRODUCTION

The idea of using processes with heavy tails of their marginal distributions is frequently discussed in various applications. Recently it was proposed to consider some particular new classes of diffusions and use them as a base for some stochastic financial applications instead of the Wiener process; see [1].

One of such classes is called the Student diffusion.

The primary reason to replace the Wiener process as a base of stochastic financial theory is the following. It is generally accepted that the heavy-tailed distributions often occur in practice, with their widespread use in communication networks, risky assets and insurance modelling; in contrast, the distributions of the Wiener process do not have heavy tails. Another belief in the mathematical finance world is the so-called “long memory” of the market, which is a property that the classical Black–Merton–Scholes market apparently does not have. There is no strict definitions of short and long memory; however, a folklore opinion is that “short” should somehow relate to the exponential decay of one or another information coefficient, while “long” relates to a slower than exponential decay. So, one possibility is to accept that the exponential mixing corresponds to the short memory, while any slower, e.g., the polynomial mixing rate could be interpreted as the long memory. Of course, this does not pretend to be uniquely defined, because even mixing itself may be understood via different mixing coefficients.

The aim of the present paper is to provide bounds for the mixing rates for Student diffusions. We establish exponential upper bounds for the beta-mixing.

The latter type of mixing could be treated as a short memory. In our further work we will discuss other models with heavy tails and long memory for other classes of processes with the same stationary distribution.

A method for establishing upper bounds for the beta-mixing was proposed in the 1980s for diffusion processes by the second author. For Markov diffusions, it is based on the “local mixing” provided by the Harnack inequality or on another similar tool, and on

2010 Mathematics Subject Classification. Primary 60H10, 60J60.
Key words and phrases. Student diffusion, exponential mixing, heavy tails.

©2010 American Mathematical Society
two bounds, for recurrence and moments, namely

(1) \[ E_x e^{\alpha \tau} \leq h(x), \quad \alpha > 0, \]

and

(2) \[ \sup_{t \geq 0} E_x h(X_t) \leq Ch(x), \]

or, at least,

\[ \sup_{t \geq 0} E_x h(X_t)1(t \leq \tau) \leq Ch(x) \]

for some function \( h \), where

\[ \tau := \inf(t \geq 0 : X_t \in D) \]

with some appropriate set \( D \), usually with the ball \( B_R = \{ x : |x| \leq R \} \).

Several estimates obtained by this method can be found in [6], [7], [9].

One may ask, why do we use the beta-mixing and not some other coefficients? Among other mixing coefficients such as alpha, phi, etc., beta clearly is the most suitable because it is stronger than alpha and yet allows some quantitative evaluation. On the other hand, the stronger coefficients like the \( \phi \)-mixing for diffusions in noncompact state spaces are useless, since normally they do not decay at all due to the noncompactness.

Why not use the covariance or correlation for this purpose? Here the reason is subtler. All members of the family of mixing coefficients, unlike the covariance, possess a very nice feature: any upper bound established for the decay of any coefficient remains valid also for any process that is a function of the “underlying process”. This makes all mixing coefficients a rather universal tool, in particular, very useful for the derivatives market, while the covariance evaluated for a process cannot easily be used to analyze another related process. Eventually, a sufficiently fast decay of alpha (or beta) mixing coefficient, along with the existence of certain moments of the process, provide an easy tool for establishing a central limit theorem, which is a basic result for all statistical applications (cf. [2, Theorem 18.4.1], or Corollary 1 below).

The (nonstationary) \( \beta \)-mixing coefficient is defined as

\[ \beta^2(t) = \sup_{s \geq 0} \var_{s \in \mathcal{F}_{s+s}^X} \left( P(B | \mathcal{F}_{s}^X) - P(B) \right), \]

where \( \mathcal{F}_{s}^X \) is the \( \sigma \)-field generated by the values \( X_s, s \in I \), and \( E_x \) means the expectation for a process with the initial value \( x \). The approach we will use is based on the following bounds, which are versions of inequalities (1) and (2):

(3) \[ E_x \exp(\alpha \tau) \leq C (1 + |x|^{2m}) \]

with some constants \( R > 0, \alpha, \) and \( m > 0 \). Another technical bound, which we will establish below, is the following version of inequality (2):

(4) \[ \sup_{t \geq 0} E_x |X_t|^{2m} \leq C (1 + |x|^{2m}) \]

(formally corresponding to the case of the function \( h(x) = 1 + |x|^{2m} \) in (2)), together with a complementary one,

(5) \[ \int |x|^{2m} \mu_\infty(dx) < \infty, \]
where \( \mu_\infty \) denotes a unique stationary measure of the process. In the sequel, \( \mu^\tau \) will denote a marginal distribution of \( X_t \) given an initial value \( X_0 = x \) for the Markov process \( X \).

In the next Section 2 we define a Student process according to [1] and consider its basic properties known from [1]. Section 3 contains the main results for this process related to the mixing. Proofs are given in Section 4.

The technique of [7] and [6] relates to stochastic differential equations with bounded coefficients and a nondegenerate diffusion, while equation (6) in Section 3 has a linearly growing drift and diffusion. Hence, to apply the technique developed in [7] and [6], we must make sure that unbounded coefficients do not bring any major difficulty. This is the reason why we have to recall all main steps and verify that no new problem arises in the proofs. There is one simplification as compared to [6] and [7], namely, the coupling can be performed here via intersections of trajectories. This is a property of the one-dimensional case.

2. STUDENT DIFFUSION PROCESS

Consider the following stochastic differential equation in \( \mathbb{R}^1 \):

\[
\begin{align*}
\dot{X}_t &= -\theta(X_t - \mu)dt + \sqrt{\nu} \left[ \frac{2\theta \delta^2}{\nu - 1} \left( 1 + \left( \frac{X_t - \mu}{\delta} \right)^2 \right) \right] dB_t, \quad X_0 = x.
\end{align*}
\]

Here

\[
\begin{align*}
\nu > 2, \quad \theta > 0, \quad \delta > 0, \quad \mu \in \mathbb{R},
\end{align*}
\]

and \( B = (B_t, t \geq 0) \) is a standard Brownian motion. Due to the classical Itô theorem, the above stochastic differential equation has a unique strong solution. It is known that this solution possesses the Markov and strong Markov property (cf., e.g., [3]). The Student distribution \( T(\nu, \delta, \mu) \) with the density

\[
\begin{align*}
f(x) &= \frac{c(\nu)}{\delta} \left[ 1 + \left( \frac{x - \mu}{\delta} \right)^2 \right]^{(\nu+1)/2}, \quad x \in \mathbb{R}^1, \quad c(\nu) = \sqrt{\frac{\pi}{\Gamma(\frac{\nu+1}{2})}},
\end{align*}
\]

is stationary for this Markov process. This can easily be checked by verifying the equation \( L^* f = 0 \) for the generator

\[
\begin{align*}
L = -\theta(x - \mu) \frac{d}{dx} + \frac{1}{2} \frac{2\theta \delta^2}{\nu - 1} \left( 1 + \left( \frac{x - \mu}{\delta} \right)^2 \right) \frac{d^2}{dx^2},
\end{align*}
\]

that is,

\[
\begin{align*}
L^* h(x) &= \frac{d}{dx} \left( \theta(x - \mu) h(x) \right) + \frac{\theta \delta^2}{\nu - 1} \left[ 1 + \left( \frac{x - \mu}{\delta} \right)^2 \right] \frac{dh(x)}{dx}.
\end{align*}
\]

In [1]–[3], this density is denoted by \( st_\nu(x) \), for given \( \delta \) and \( \mu \).

Note that the parameter \( \theta \) in equation (6) is not involved in the expression for the stationary density \( st_\nu \), since its change influences the scale of the speed only, and may be removed by a deterministic time change. In particular, if \( X_0 \sim T(\nu, \delta, \mu) \), then \( X \) is stationary. Also, for all \( t, s \geq 0 \),

\[
\begin{align*}
\mathbb{E}(X_{s+t} | X_s = x) &= xe^{-\theta t} + \mu \left( 1 - e^{-\theta t} \right),
\end{align*}
\]

and the autocorrelation function of \( X \) is given by the expression (cf., e.g., [1])

\[
\begin{align*}
r(t) &= \text{Corr}(X_{s+t}, X_s) = e^{-\theta t}, \quad t \geq 0, \quad s \geq 0.
\end{align*}
\]
3. Main results

Theorem 1. For any $0 < m < \nu/2$ and any

$$0 < \alpha < 2m\theta \left(1 - \frac{2m - 1}{\nu - 1}\right),$$

there exists a constant $C > 0$ such that

$$\beta^x(t) \leq C(x) \exp(-\alpha t), \quad C(x) = C \left(1 + \frac{|x - \mu|^{2m}}{\delta}\right).$$

Also, there exists a constant $C > 0$ such that

$$\|\mu^x(t) - \mu\|_{TV} \leq C(x) \exp(-\alpha t), \quad C(x) = C \left(1 + \frac{|x - \mu|^{2m}}{\delta}\right).$$

Corollary 1. If $\nu > 4$, $2 < m < \nu/2$, and $f$ is bounded, then the weak convergence holds for $S_t := \int_0^t f(X_s) \, ds$ in the stationary regime,

$$S_t - tE_{st} f(X_0) \overset{p_{st}}{\rightarrow} Z \sim N(0, \sigma^2),$$

as well as in the nonstationary one; that is, for any initial value $x \in \mathbb{R}$,

$$S_t - tE_{st} f(X_0) \overset{p_x}{\rightarrow} Z \sim N(0, \sigma^2),$$

where $0 \leq \sigma^2 := 2 \int_0^\infty \text{cov}_{st}(f(X_0), f(X_t)) \, dt < \infty$, and all symbols $P_{st}$, $E_{st}$, and $\text{cov}_{st}$ stand for stationary regime.

The assumption that $f$ is bounded can be relaxed.

4. Auxiliary results

First we establish several auxiliary results.

Lemma 1. For any $0 < m < \nu/2$,

$$0 < \alpha < 2m\theta \left(1 - \frac{2m - 1}{\nu - 1}\right),$$

and for sufficiently large $R$, there exists a constant $C > 0$ such that

$$\sup_{t\geq 0} E_x e^{\alpha (t \wedge \tau)} \left|\frac{X_{t \wedge \tau} - \mu}{\delta}\right|^{2m} \leq \left|\frac{x - \mu}{\delta}\right|^{2m}$$

and

$$E_x e^{\alpha \tau} \left|\frac{X_\tau - \mu}{\delta}\right|^{2m} \leq \left|\frac{x - \mu}{\delta}\right|^{2m}.$$
Applying Itô’s formula to the process $f(t, Y_t)$ for $t < \tau$, we have
\[
df(t, Y_t) = \alpha \exp(\alpha t) \psi(Y_t) dt + \exp(\alpha t) \psi'(Y_t) dY_t + \frac{1}{2} \exp(\alpha t) \psi''(Y_t) dt
\]
\[
\equiv \alpha \exp(\alpha t) |Y_t|^{2m} dt + 2m \exp(\alpha t) |Y_t|^{2m-1} \text{sgn}(Y_t) dY_t
\]
\[
+ m(2m - 1) \exp(\alpha t) |Y_t|^{2m-2} (dY_t)^2.
\]
Here $dY_t = C_1 Y_t dt + C_2 \sqrt{1 + Y_t^2} dW_t$, $C_1 = -\theta$, and
\[
C_2 = \sqrt{\frac{2\theta}{\nu - 1}}.
\]
Hence, $(dY_t)^2 = C_2^2 (1 + Y_t^2) dt$. Thus, always with $t < \tau$,
\[
df(t, Y_t) = \exp(\alpha t) \left( \alpha |Y_t|^{2m} dt + (2m|Y_t|^{2m-1} \text{sgn}(Y_t)) \left( C_1 Y_t dt + C_2 \sqrt{1 + Y_t^2} dW_t \right) \right.
\]
\[
\left. + (m(2m - 1)C_1^2) (Y_t^{2m-2}) (1 + Y_t^2) dt \right)
\]
\[
= \exp(\alpha t) \times |Y_t|^{2m} \left( \alpha + 2m C_1 + m(2m - 1)C_2^2 + m(2m - 1)C_2^2 |Y_t|^{-2} + \alpha |Y_t|^{-2m} \right) dt
\]
\[
+ \left( 2m \exp(\alpha t) C_2 |Y_t|^{2m-1} \text{sgn}(Y_t) \sqrt{1 + |Y_t|^2} \right) dW_t.
\]
Recall that
\[
C_1 = -\theta, \quad C_2 = \sqrt{\frac{2\theta}{\nu - 1}},
\]
and $2m < \nu$. We assume that
\[
\alpha + 2m C_1 + m(2m - 1)C_2^2 < 0 \quad \iff \quad \alpha < 2\theta m \left( 1 - \frac{2m - 1}{\nu - 1} \right).
\]
Under these assumptions on $\alpha$ and for sufficiently large $R$, we have
\[
(\alpha + 2m C_1 + m(2m - 1)C_2^2 + m(2m - 1)C_2^2 |Y_t|^{-2} + \alpha |Y_t|^{-2m}) \leq -c < 0.
\]
Therefore,
\[
(15) \quad \mathbb{E}_y f(t \wedge \tau, Y_{t \wedge \tau}) - f(0, y) \leq -c \mathbb{E}_y \int_0^{t \wedge \tau} f(s, Y_s) ds.
\]
In particular, we obtain
\[
\mathbb{E} f(t \wedge \tau, Y_{t \wedge \tau}) - f(0, Y_0) \leq 0
\]
for sufficiently large $R$. This is equivalent to inequality (12). As $t \to \infty$, by virtue of Fatou’s lemma we also have
\[
\mathbb{E} f(\tau, Y_\tau) \leq f(0, Y_0).
\]
This is equivalent to inequality (13). Lemma 1 is proved. \(\square\)

**Remark 1.** A standard procedure of localization explains why the expectation of the stochastic integral vanishes.

**Lemma 2.** For any $0 < m < \nu/2$ and for sufficiently large $R$, there exists a constant $C > 0$ such that
\[
(16) \quad \sup_{t \geq 0} \mathbb{E}_x \left| \frac{X_t - \mu}{\delta} \right|^{2m} \leq C \left( 1 + \left| \frac{x - \mu}{\delta} \right|^{2m} \right).
\]
Proof. In terms of the process \( Y \), we are to show the inequality
\begin{equation}
\sup_{t \geq 0} E_y |Y_t|^{2m} \leq C (1 + |y|^{2m}).
\end{equation}

Let us apply \( f(t, y) \) with \( \alpha = 0 \). Then, as in the proof of Lemma 1
\[
E_y \psi(Y_{t_1}) - E_y \psi(Y_{t_1}) \\
\leq -c E_y \int_{t_1}^{t_2} \psi(Y_s) ds + C_0 E_y \int_{t_1}^{t_2} \psi(Y_s) \mathbb{1}(|Y_s| < R) ds \\
\equiv -c E_y \int_{t_1}^{t_2} \psi(Y_s) ds + E_y \int_{t_1}^{t_2} \psi(Y_s) (C_0 \mathbb{1}(|Y_s| < R) + c \mathbb{1}(|Y_s| < R)) ds \\
\leq -c E_y \int_{t_1}^{t_2} h(Y_s) ds + C \int_{t_1}^{t_2} \psi(Y_s) ds + C (t_2 - t_1)
\]
for all \( 0 \leq t_1 \leq t_2 < \infty \), since \( \inf \psi > 0 \). Therefore, if the derivative of the function
\[
g(t) := E_y \psi(Y_t) \geq 0 \text{ exists},
\]
then it must satisfy the inequality
\begin{equation}
g'(t) \leq -cg(t) + C.
\end{equation}

Clearly, the function \( g \) is well defined and is continuous with respect to \( t \). The fact that it is differentiable follows from the reasoning similar to that in the proof of Lemma 1 (with \( \alpha = 0 \)) if we do not use inequalities but instead write down the Itô formula and compute expectations taking into account the already established fact that the expectation of the stochastic integral vanishes. Finally, any nonnegative locally bounded continuous function which obeys (18), must satisfy the inequality
\[
g(t) \leq g(0) \exp(-ct) + \frac{C}{c} \quad \iff \quad E_y \psi(Y_t) \leq \psi(y) \exp(-ct) + \frac{C}{c}.
\]

Since \( \psi(y) \mathbb{1}(|y| \geq R_1) = |y|^{2m} \mathbb{1}(|y| \geq R) \) and \( \psi \geq 0 \), the latter is even slightly stronger than (17). Lemma 2 is proved.

Lemma 3. For any \( 1 < m < \nu/2 \),
\[
\int |x|^{2m} \mu_\infty(dx) < \infty.
\]

Proof. The proof follows from the explicit representation of the stationary density. \( \square \)

Now we consider the direct product of two identical probability spaces with Wiener processes, where two independent copies of our Markov process \( (X_t, t \geq 0) \) and \( (\tilde{X}_t, t \geq 0) \) live with the initial values \( x \) and \( \tilde{x} \), respectively. The notation for the probability and expectation remains unchanged. Let \( \gamma \equiv \gamma_{R_1} = \inf(t \geq 0: |X_t| \vee |\tilde{X}_t| \leq R_1) \) and \( \gamma(t) = \min(\gamma, t) \). In the following result, we assume that the constant \( R \) is already chosen and fixed.

Lemma 4. For all \( 0 < m < \nu/2 \), there exist \( R_1 \geq R \) and \( C > 0 \) such that
\begin{equation}
E_{x, \tilde{x}} \exp(\alpha \gamma) \leq C \left( 1 + \left| \frac{x - \mu}{\delta} \right|^{2m} + \left| \frac{\tilde{x} - \mu}{\delta} \right|^{2m} \right).
\end{equation}

Proof. Consider the Lyapunov function in variables \( y \) and \( \tilde{y} \), namely
\[
f(t, y, \tilde{y}) = \exp(\alpha t) (\psi(y) + \psi(\tilde{y}))
\]
Similarly to the proof of Lemma 1, we apply Itô’s formula
\[
\begin{align*}
df(t, Y_t, \dot{Y}_t) &= \alpha \exp(\alpha t) \left( \psi(Y_t) + \psi(\dot{Y}_t) \right) dt \\
&\quad + \exp(\alpha t) \psi'(Y_t) dY_t + \exp(\alpha t) \psi'(\dot{Y}_t) d\dot{Y}_t + \frac{1}{2} \exp(\alpha t) \psi''(Y_t) (dY_t)^2 \\
&\quad + \frac{1}{2} \exp(\alpha t) \psi''(\dot{Y}_t)(d\dot{Y}_t)^2.
\end{align*}
\]
To get the desired bound, we must have a negative expression with \(dt\), that is,
\[
1(t < \gamma) \left[ \alpha \left( \psi(Y_t) + \psi(\dot{Y}_t) \right) + C_1 \left( \psi'(Y_t) + \psi'(\dot{Y}_t) \right) \\
+ \frac{1}{2} C_2 \left( \psi''(Y_t) \left( 1 + |Y_t|^2 \right) + \psi''(\dot{Y}_t) \left( 1 + |\dot{Y}_t|^2 \right) \right) \right] \leq 0.
\]

There are three main cases here (with several symmetrical subcases):

(I) \(|Y_t| \geq R_1\) and \(|\dot{Y}_t| \geq R_1\),
or

(II) \(|Y_t| \geq R_1\) and \(R \leq |\dot{Y}_t| < R_1\),
or

(III) \(|Y_t| \geq R_1\) and \(|\dot{Y}_t| < R\).

In cases (I) and (II), the whole expression (20) is negative for all \(R_1 \geq R\) due to the evaluation in the proof of Lemma 1, since by the choice of \(R\), we have
\[
\left[ \alpha |Y_t|^{2m} + 2mC_1 |Y_t|^{2m} + m(2m - 1)(|Y_t|^{2m} + |Y_t|^{2m-2})C_2^2 \right] < 0
\]
and
\[
\left[ \alpha |\dot{Y}_t|^{2m} + 2mC_1 |\dot{Y}_t|^{2m} + m(2m - 1)(|\dot{Y}_t|^{2m} + |\dot{Y}_t|^{2m-2})C_2^2 \right] < 0.
\]

In case (III), we cannot guarantee that the terms with \(\dot{Y}\) provide a negative input: it may turn out to be positive. However, given \(R\), the positive part of all these terms is bounded by
\[
\sup_{|y| \leq R} \left| \alpha \psi(y) + C_1 \psi'(y) + \frac{1}{2} C_2 \psi''(y) \left( 1 + |y|^2 \right) \right| =: C_0 < \infty.
\]
(The constant \(C_0\) on the right hand side does not depend on \(R\), although if it were, the change in the proof would have been minimal.) Let us choose \(R_1\) so large that
\[
\left| 2mC_1 |y|^{2m} + m(2m - 1)(|y|^{2m} + |y|^{2m-2})C_2^2 \right| \gg C_0
\]
for any \(|y| \geq R_1\). Since the negative part due to \(Y\) is of the order of at least \(R_1^{2m}\), its absolute value exceeds \(g(R)\). Hence, the desired statement follows as in the proof of Lemma 1. Lemma 3 is proved. \(\square\)

5. Proof of Theorem 1

Consider two independent copies of a Markov process, \(X\) and \(\tilde{X}\), both being solutions of equation (3) with two different independent Wiener processes, \((W_t, t \geq 0)\) and \((\tilde{W}_t, t \geq 0)\), with deterministic initial values \(X_0 = x \in \mathbb{R}\) and \(\tilde{X}_0 = \tilde{x} \in \mathbb{R}\), respectively. Later on, in certain steps of the proof, \((\tilde{x}, \tilde{y})\) will be chosen randomly. After the change of variables as in (13), denote the corresponding processes by \(Y\) and \(\tilde{Y}\). Fix \(s_0 \geq 0\). Consider a sequence of stopping times, \(\gamma_1 < \gamma_2 < \cdots\), defined as follows:
\[
\gamma_1 = \inf \{ t \geq s_0 : |Y_t| \leq R \text{ and } |\tilde{Y}_t| \leq R \}.
\]
and, for $n \geq 1$,
\[ T_n = \inf (t \geq \gamma_n : |Y_t| \geq R + 1, \text{ or } |	ilde{Y}_t| \geq R + 1) \land (\gamma_n + 1). \]

Then we put
\[ \gamma_{n+1} = \inf (t \geq T_n : |Y_t| \leq R \text{ and } |	ilde{Y}_t| \leq R). \]

By Lemma [1] we have a priori bounds
\[ \mathbb{E} \left( \exp(\alpha(\gamma_1 - s_0)) \mid \tilde{F}_{s_0} \right) \leq C(1 + |Y_{s_0}|^{2m} + |	ilde{Y}_{s_0}|^{2m}) \]
and
\[ \mathbb{E} \left( \exp(C(\gamma_{n+1} - \gamma_n)) \mid \tilde{F}_{\gamma_n} \right) \leq C. \]

Here initial values are irrelevant and hence dropped.

2. We follow the coupling procedure as described in [6] and [7] and use the Harnack inequality. Given $Y_{\gamma_n}$ and $\tilde{Y}_{\gamma_n}$, consider the exit measures of both processes on the parabolic boundary $\Gamma$ of the cylinder
\[ (t, y, \tilde{y}) : \gamma_n \leq t \leq \gamma_n + 1, |y| \leq R + 1, |	ilde{y}| \leq R + 1, \]
that is,
\[ \Gamma = \{(t, y, \tilde{y}) : \gamma_n \leq t \leq \gamma_n + 1, |y| = R + 1, |	ilde{y}| = R + 1 \} \]
\[ \cup \{(t, y, \tilde{y}) : t = \gamma_n + 1, |y| \leq R + 1, |	ilde{y}| \leq R + 1 \}. \]

Let $\delta \in (0, 1)$. Consider, for example, the part of $\Gamma$,
\[ \Gamma_\delta = \{(t, y, \tilde{y}) : \gamma_n + \delta \leq t \leq \gamma_n + 1, |y| = R + 1, |	ilde{y}| = R + 1 \}
\[ \cup \{(t, y, \tilde{y}) : t = \gamma_n + 1, |y| \leq R + 1, |	ilde{y}| \leq R + 1 \}. \]

Using the Harnack inequality [4], the exit measures on $\Gamma_\delta$ of both processes $Y$ and $\tilde{Y}$ are equivalent and have bounded derivative. Therefore,
\[ \inf_{y, \tilde{y} \in B_R} \int_{\Gamma} \left( \frac{P_y((T, Y_T) \in dv)}{P_{\tilde{y}}((T, Y_T) \in dv)} \land 1 \right) \frac{dP_y((T, Y_T) \in dv)}{dP_{\tilde{y}}((T, Y_T) \in dv)} =: c > 0 \]
on $\Gamma$, where $dv$ is the area element on $\Gamma$. The latter inequality is called the Markov–Dobrushin local mixing condition.

Applying the approach of [8, Section 2.4], given $\tilde{F}_{\gamma_n}$, one can find new representations for $Y$ and $\tilde{Y}$ on some probability space at the moments when they first touch $\Gamma$ such that they coincide with probability
\[ P_{Y_{\gamma_n}, \tilde{Y}_{\gamma_n}} ((T_n, Y_{T_n}) = (T_n, \tilde{Y}_{T_n})) \geq c. \]

Such representations are usually constructed on some extensions of the initial probability space. Moreover, given their values at $\gamma_n$ and at the moment of hitting $\Gamma$, say $T_n$, one can recover the part of the trajectory between these two values via the conditional measure.

It is important to note that, for the new representations,
\[ P_{F_{\gamma_n}} (Y_{T_n} = \tilde{Y}_{T_n}) \geq c > 0. \]

Thus we denote the moment of coupling by
\[ L := \inf(T_n : Y_{T_n} = \tilde{Y}_{T_n}). \]

We refer to [8] for more details.

3. In the one-dimensional case, the coupling can be described in terms of a simpler idea of intersection. Consider the following change of variables:
\[ y \mapsto \sqrt{\frac{\nu - 1}{2\nu}} \left( y + \sqrt{1 + y^2} \right). \]
For the new scale, equation \((6)\) is rewritten as follows:
\[ dY_t = \hat{b}(Y_t) \, dt + dW_t, \quad Y_0 = y, \]
with
\[ \hat{b}(y) = -\sqrt{\frac{2\theta(\nu - 1)^2 + \theta}{2(\nu - 1)}} \exp(2y - 1); \]
that is, the diffusion coefficient in this scale is equal to one, while the drift \(\hat{b}\) is bounded.

Now it is useful to consider the levels \(R\) and \(R + K\) instead of the levels \(R\) and \(R + 1\) as in the preceding proof, where \(K\) is a parameter to be chosen later. One of the restrictions is that \(K\) is such that \(K/2 \geq ||\hat{b}||_C\); another restriction is given below. Consider the following random event:
\[ A := \left\{ (Y_{\gamma_n} - \tilde{Y}_{\gamma_n})Y_{\gamma_n+1} \leq 0 \text{ and } (Y_{\gamma_n} - \tilde{Y}_{\gamma_n})\tilde{Y}_{\gamma_n+1} \geq 0 \right\}. \]

For all \(Y_{\gamma_n}\) and \(\tilde{Y}_{\gamma_n}\), the probability of this event can be estimated from below as follows:
\[
P(A) \geq \Phi\left(-R - ||\hat{b}||\right)^2.
\]
Indeed, assume, for example, that \(Y_{\gamma_n} > \tilde{Y}_{\gamma_n}\). Then the probability attains its minimum if \(Y_{\gamma_n} = R\) and \(\tilde{Y}_{\gamma_n} = -R\). Moreover,
\[ A \supset \left\{ W_{\gamma_n+1} - W_{\gamma_n} \leq -R - ||\hat{b}|| \right\} \cap \left\{ \tilde{W}_{\gamma_n+1} - \tilde{W}_{\gamma_n} \geq R + ||\hat{b}|| \right\} \]
and this proves estimate \((22)\).

If the event \(A\) occurs, the trajectories of \(Y\) and \(\tilde{Y}\) intersect on \([\gamma_n, \gamma_n + 1]\). The first intersection in this interval is the stopping time, which means that we can “glue” the processes after the first “meeting”. For the sake of definiteness, let the first \(Y\) follow the second \(\tilde{Y}\). There is only one little nuisance if we do this construction straight. Namely, before this intersection, one or both trajectories may already happen to be far away from the origin, that is, escape from \([-R - K, R + K]\). To avoid the technical problems, it is convenient to stop the processes at the first exit time. In other words, we do not try to couple them until \(\gamma_n + 1\); instead we start the procedure again, i.e., wait until next \(\gamma_n + 1\) when both components are inside \([-R, R]\) and attempt to couple (intersect) them again on \([\gamma_n + 1, \gamma_n + 1 + 1]\). We can choose \(K\) sufficiently large to make the probability of the exit from \([-R - K, R + K]\) during the time \([\gamma_n, \gamma_n + 1]\) small. Thus, for the event,
\[ A_K := \left\{ (Y_{\gamma_n} - \tilde{Y}_{\gamma_n})Y_{\gamma_n+1} \leq 0 \right\} \cap \left\{ (Y_{\gamma_n} - \tilde{Y}_{\gamma_n})\tilde{Y}_{\gamma_n+1} \geq 0 \right\} \cap \left\{ \sup_{\gamma_n \leq t \leq \gamma_n + 1} (|Y_t| \vee |\tilde{Y}_t|) < R + K \right\}, \]
we have, for example,
\[
P(A_K) \geq \Phi\left(-R - ||\hat{b}||\right)^2/2.
\]
Now, the coupling via the intersection at the moment when \(Y\) and \(\tilde{Y}\) meet for the first time is accepted only if this intersection occurs on \([\gamma_n, \gamma_n + 1]\) before the exit time \(\tilde{T}_n\), where
\[ \tilde{T}_n := \inf\{ t \geq \gamma_n : |Y_t| \vee |\tilde{Y}_t| \geq R + K \} \wedge (\gamma_n + 1) < \gamma_n + 1. \]
The probability of the latter event is at least as large as \(P(A_K)\) in \((23)\) for which we obtain the estimate
\[
P(\text{an intersection occurs before } \tilde{T}_n \mid F_{\gamma_n}) \geq \Phi\left(-R - ||\hat{b}||\right)^2/2.
\]
If, however, $\hat{T}_n$ occurs earlier, we wait until the next $\gamma_{n+1}$ and then repeat the procedure. Note that a better lower estimate as compared to (23) can be obtained by considering the probability of the event

$$A' := \left\{ (Y_{\gamma_n} - \hat{Y}_{\gamma_n})(Y_{\gamma_n+1} - \hat{Y}_{\gamma_n+1}) \leq 0 \right\},$$

namely

$$\mathbb{P}(A') \geq \Phi(-\sqrt{2}(R + \|b\|)).$$

It is clear that $A'$ also implies an intersection. In the rest of the proof, we deal with the sequence of stopping times $(\gamma_n, T_n)$ as mentioned in step 2. However, the same can be done similarly with $(\gamma_n, \hat{T}_n)$.

4. Now we analyze the following iterative procedure of “attempts” to couple the two processes at $L = T_1, T_2, \ldots$.

If we use intersections, then we define the coupling time $L$ as

$$L := \inf \left( t \geq s_0 : Y_t = \hat{Y}_t, |Y_t| < R + K, t \in \bigcup_{n \geq 0} [\gamma_n, \gamma_n + 1] \right).$$

Note that if the two processes are not coupled before $\gamma_n$, they can be coupled at the moment $T_n$ with probability (24) (if we use $(\hat{T}_n)$, then we refer to (24)). The moment $L$ when they are coupled is a stopping time. Due to the strong Markov property, after being coupled, the trajectories of two processes can be continued as identical; for example, the first process may follow the second one. This does not influence the strong Markov property for each process which solves equation (9). Our procedure of coupling implies the following bound:

$$\mathbb{P}_y(L_{s_0} > T_n) \leq (1 - c)^n =: \kappa^n. \tag{25}$$

A similar bound holds also in the version of intersections,

$$\mathbb{P}_y(L_{s_0} > \hat{T}_n) \leq (1 - \hat{c})^n =: \hat{\kappa}^n. \tag{26}$$

Now we estimate the difference

$$\mathbb{P}_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - \mathbb{P}_y(Y_{s_0+t} \in B) \tag{27}$$

for all $B \in \mathcal{B}(\mathbb{R}).$

To estimate the difference in (27), we consider the second independent version of our process $X$, say $X$, started at $s_0$ and with distribution $\mu_{s_0}$ that is the same as the distribution of $X_{s_0}$ itself. In terms of the processes $Y$ and $\hat{Y}$, we consider $Y_t := Y_{s_0+t}$, $t \geq 0$, started at $s_0$ and having the distribution $\nu_{s_0}^y$ (we denote by $\nu_{s_0}^y$ the distribution of $Y_{s_0}$ started at $y$). We have

$$\left| \mathbb{P}_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - \mathbb{P}_y(Y_{s_0+t} \in B) \right| = \left| \mathbb{P}_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - \mathbb{P}_{y_{s_0}}(Y_t \in B) \right|.$$

Using our coupling procedure with the coupling time $L$, and assuming for simplicity that $s_0 = 0$, we can estimate

$$\left| \mathbb{P}_y(Y_t \in B) - \mathbb{P}_{y_0}(\hat{Y}_t \in B) \right| \leq \mathbb{P}(L > t) \leq E(1(L > t)1(t < T_0)) + \sum_{n=0}^{\infty} E(1(L > t)1(T_n \leq t < T_{n+1}))$$

$$\leq E(1(t < T_0)) + \sum_{n=0}^{\infty} \mathbb{P}(L > T_n)^{1/p} \mathbb{P}(T_{n+1} > t)^{1/q}, \tag{28}$$
where $p^{-1} + q^{-1} = 1$ and $p, q > 0$. Since $P(L > T_n) \leq \kappa^n$, the Bienaimé–Chebyshev inequality implies that

\[
P(\gamma_{n+1} > t) \leq \exp(-\alpha t) E \exp(\gamma_{n+1}) \leq \exp(-\alpha t) E \exp(\alpha(\gamma_{n+1} - \gamma_n) + (\gamma_n - \gamma_{n-1}) + \cdots + (\gamma_2 - \gamma_1) + \gamma_1)) \equiv \exp(-\alpha t) E \exp(\alpha(\gamma_n - \gamma_{n-1}) + \cdots + (\gamma_2 - \gamma_1) + \gamma_1)) \times E_{\gamma_n, \gamma_{n-1}, \cdots, \gamma_1} \exp(\alpha(\gamma_{n+1} - \gamma_n)).
\]

Due to Lemma 2,

\[
E_{\gamma_n, \gamma_{n-1}, \cdots, \gamma_1} \exp(\alpha(\gamma_{n+1} - \gamma_n)) \leq C.
\]

By induction, we get,

\[
P(\gamma_{n+1} > t) \leq \exp(-\alpha t) C^{n+1} (1 + |y_1|^{2m} + |y_2|^{2m}).
\]

Since $\gamma_{n+1} \leq T_{n+1} \leq \gamma_{n+1} + 1$, we conclude that

\[
P(T_{n+1} > t) \leq \exp(-\alpha(t - 1)) C^n (1 + |y_1|^{2m} + |y_2|^{2m}).
\]

Hence, as $Y_{s_0}$ and $\tilde{Y}_{s_0}$ are actually random, we get

\[
P(L > t) = \mathbb{E} 1(L > t)1(t < T_1) + \sum_{n=1}^{\infty} \mathbb{E} 1(L > t)1(T_n \leq t < T_{n+1}) \leq \mathbb{E} 1(t < T_1) + \sum_{n=1}^{\infty} P(\gamma_{n+1} > t) P(T_{n+1} > t)^{1/q}
\]

\[
\leq P(t < T_1) + \sum_{n=1}^{\infty} \kappa^{n/p} P(T_{n+1} > t)^{1/q} \leq C \exp(-\alpha(t - 1)) \int (1 + |Y_{s_0}|^{2m} + |\tilde{Y}|^{2m}) \mu_{s_0}^y (d\tilde{y})
\]

\[
+ \sum_{n \geq 1} \kappa^{n/p} \exp(-\alpha(t - 1)) (\hat{C}^{n+1})^{1/q} \int (1 + |Y_{s_0}|^{2m} + |\tilde{Y}|^{2m})^{1/q} \mu_{s_0}^y (d\tilde{y}) \leq C \exp(-\alpha t) (1 + |Y_{s_0}|^{2m}).
\]

The latter inequality holds if $p$ and $q$ are chosen so that

\[
k^{1/p} \hat{C}^{1/q} < 1.
\]

We used the notation $\hat{C}$ instead of $C$ in the proof to indicate how $p$ and $q$ are chosen to satisfy (29). With this choice, we estimate the $\beta$-mixing coefficient for the process $Y$ by virtue of Lemma 2

\[
\beta^n_y \leq \sup_{s \geq 0} C \exp(-\alpha t) (1 + |Y_{s_0}|^{2m}) \leq C \exp(-\alpha t) (1 + |y|^{2m}).
\]

In terms of the original process $X$, this reads as follows:

\[
\beta^n_x \leq C \exp(-\alpha t) (1 + \left| \frac{x - \mu}{\delta} \right|^{2m}).
\]

5. The total variation distance is estimated in a similar way. Let $s_0 = 0$. Consider an independent pair of Markov processes $X$ and $\hat{X}$ (equivalently, $Y$ and $\tilde{Y}$), where the
process $\tilde{X}$ is stationary as well as $\tilde{Y}$. The coupling procedure is similar to that considered above in steps 2 and 3. Similarly to (28) we obtain from Lemma 3 that
\[
|P_y(Y_t \in B) - P_{\nu}(\tilde{Y}_t \in B)| \leq P(L > t) \\
\leq C \exp(-\alpha(t - 1)) \int (1 + |y|^{2m} + |\tilde{y}|^{2m}) \nu_y(d\tilde{y}) \\
\leq C \exp(-\alpha t) (1 + |y|^{2m}).
\]
In terms of the process $X$, this implies the bound (9). Theorem 1 is proved.

**Remark 2.** If $1 \leq m < \nu/2$, then one can choose $\psi(y) \equiv |y|$, and this simplifies the reasoning to some extent.

**Proof of Corollary 1** is straightforward; it follows from [2, Theorem 18.4.1], Theorem 1, and the following inequality for the moments:
\[
E|X_0|^{2m} < \infty \text{ for } f \in L_2(\nu). (The assumption that the function $f$ is bounded can be weakened here.)
\]

Then the nonstationary version of the central limit theorem is a consequence of the following remark which reduces the case to the stationary version. We have
\[
\frac{1}{\sqrt{t}} \int_0^t f(X_s) ds - t E f(X_s) \equiv \frac{1}{\sqrt{t}} \int_0^t f(X_s) ds - t E f(X_s) + \int_0^t (f(X_s) - f(\tilde{X}_s)) 1(s \leq L) ds,
\]
where $\tilde{X}$ a stationary version of the process and $L$ is the coupling time.

The first term weakly converges to a Gaussian random variable in view of (10). The second term admits the estimate
\[
\left| \frac{1}{\sqrt{t}} \int_0^t (f(X_s) - f(\tilde{X}_s)) 1(s \leq L) ds \right| \leq E_x,\mu, \int_0^t |f(X_s) - f(\tilde{X}_s)| 1(s \leq L) ds
\leq 2\|f\|_B \int_0^t P_{x,\mu}(L > s) ds \\
\leq 2\|f\|_B C_x \frac{1}{c} (1 - \exp(-ct)) \to 0, \quad t \to \infty.
\]
This implies (11), and the corollary is proved.

**Acknowledgement**

The authors are grateful to the anonymous referee for useful remarks. The second author is supported by the grant RFBR 08-01-00105a. The contents of the paper is a part of the submitted PhD Thesis of the first author.

**Bibliography**


School of Mathematics, University of Leeds, LS2 9JT, Leeds, United Kingdom
E-mail address: niloufar@leeds.ac.uk

Institute for Information Transmission Problems, Moscow, Russia

Current address: School of Mathematics, University of Leeds, LS2 9JT, Leeds, United Kingdom
E-mail address: A.Veretennikov@leeds.ac.uk

Received 28/MAY/2009

Translated by THE AUTHORS