ni_v-pre2a.tex

On exponential mixing bounds and convergence rate for Student processes

N. Abourashchi,* A. Yu. Veretennikov†

May 22, 2009

Abstract

Exponential bounds for the coefficient of β-mixing along with certain polynomial moment bounds for a Student diffusion process, – i.e., a Markov diffusion with a Student distribution as a stationary measure, – are established. The method is based on direct estimation of moments and on polynomial Lyapunov functions for evaluating exponential functionals of hitting times.

MSC subject classification: 60H10, 60J60.
Key

1 Introduction

The idea of using processes with heavy tails in marginal distributions is frequently discussed in various applications. Recently it was proposed to consider some particular new classes of diffusions with a goal to replace the latter as a base for some stochastic financial applications, see [1]. One of those classes is called Student diffusions. The primary reason for replacement of Wiener process as a base of stochastic financial theory is that it is

*University of Leeds, UK, e-mail: niloufar @ leeds.ac.uk
†University of Leeds, UK, and Institute of Information Transmission Problems, Moscow, Russia, e-mail: A.Veretennikov @ leeds.ac.uk

now generally accepted that heavy-tailed distributions occur commonly in practice, with their widespread use in communication networks, risky assets and insurance modeling; this is not well compatible with the properties of Wiener process. Another belief in the mathematical finance world is a so called "long memory" of the market, which is a property that the classical Black–Merton–Scholes market does nod admit. There is no strict definition as to what should be a mathematical definition of the latter notion. One of possibilities is to accept that exponential mixing is a short memory, while any slower, e.g., polynomial mixing rate could be interpreted as a long memory. Of course, this does not pretend to be uniquely determined, for example, because mixing itself may be understood by different mixing coefficients.

In particular, in [1] it was proposed to use so called *Student processes,* a parametric class with Student distributions as stationary measures, which have desirable polynomial tails. The present paper has precisely this goal, so as to provide bounds for mixing rates for such processes. Of course, a process is not determined uniquely by its stationary distribution, and in our further work we are going to discuss some other possible constructions leading to the same stationary laws. In this paper we establish upper bounds for beta-mixing for one of possible models suggested in [2], along with analysing the tails. What to mixing, in the paper we prove that the models from our class possess exponential bounds for beta mixing coefficients. This could be treated as short memory. In our further work we will discuss other models with heavy tails and long memory, too.

One of possible ways to establish upper bounds for beta-mixing was proposed in 80s for diffusion processes by the second author. It is based on "local mixing" provided by Harnack inequality or some similar tool, and two bounds, for recurrence and moments,

$$E_x e^{\alpha\tau} \le h(x), \quad (\alpha > 0), \tag{1}$$

and

$$\sup_{t\ge 0} E_x h(X_t) \le Ch(x), \quad \text{or, at least,} \quad \sup_{t\ge 0} E_x h(X_t) 1(t \le \tau) \le Ch(x), \tag{2}$$

for some function h, where

$$\tau = \inf(t \ge 0 : \, X_t \in D)$$

for some appropriate set (or sets) D, usually balls $B_R = \{x : \, |x| \le R\}$. In certain cases bounds (2) may be further relaxed, however, they suffice for our aims here. Some estimates established by this method may be found in [6], [7], [9].

One may ask, why do we use beta-mixing and not some other coefficient. Among other mixing coefficients such as alpha, phi, et.al., beta is clearly most suitable, because it is stronger than alpha and yet allows some quantitative evaluation, while next stronger coefficients starting from ϕ for the diffusions in non-compact state spaces are useless, as normally they do not decay at all; crucial here is non-compactness of the state space. Why not use a covariance or correlation processes? Here the reason is more gentle. All members in the family of mixing coefficients – unlike covariance – possess a very nice feature: any upper bound established for the decay of any coefficient of the latter remains valid also for any process which is a function of the "underlying process". This makes those coefficients a rather universal tool, in particular, very useful for the market with derivatives, while, say, covariance evaluated for one process normally cannot be easily used to analyse another process. Eventually, a fast enough decay of alpha (or beta) mixing coefficient, along with certain moments of the process, provide an easy tool for establishing a Central Limit Theorem, which is basic for any statistical application (cf. [3, Theorem 18.4.1].

The (non-stationary) β-mixing coefficient is defined as

$$\beta^x(t) = \sup_{s\ge 0} E_x \operatorname*{var}_{B\in\mathcal{F}^X_{\ge t+s}} (P(B|\mathcal{F}^X_{\le s}) - P(B)),$$

where $\mathcal{F}^X_I$ is the σ- field generated by the values X_s, $s \in I$, and E_x means the expectation for the process with the initial value x. The approach we will use is based on the following bounds which are versions of (1) and (2). Let $B_R = (x \in R : |x| \le R)$, and

$$\tau = \inf(t \ge 1 : |X_t| \le R).$$

The first auxiliary bound we would need is of type (1),

$$E_x e^{\alpha\tau} \le C(1 + |x|^{2m}), \tag{3}$$

with some $R > 0, \alpha, m > 0$. Another technical bound which we will establish is a version of (2),

$$\sup_{t\ge 0} E_x|X_t|^{2m} \le C(1 + |x|^{2m}), \tag{4}$$

together with a complementary one,

$$\int |x|^{2m} \mu_\infty(dx) < \infty, \tag{5}$$

where μ_∞ denotes a unique stationary measure of the process; more generally, μ_t^x will denote a marginal distribution of X_t given the initial value $X_0 = x$ for the Markov process X.

In the next Section 2 we define a Student process according to [1] – in particular, some notations from that paper will be used – and state its certain basic properties known from [1]. Section 3 states main results for this process related to mixing. Section 4 is devoted to proofs.

The calculus from [7] and [6] relates to stochastic differential equations with bounded coefficients and non-degenerated diffusion, while the equation (6) in the section 3 has linearly growing drift and diffusion. Hence, to apply that technique, we ought to make sure that unbounded coefficients do not bring any major difficulty. This is the reason why we have to remind all main steps and verify that no new problems arise in the proofs. There is one simplification: coupling can be performed here via intersections of trajectories. This is a feature of 1D case.

2 Student diffusion process – 1

Consider the following stochastic differential equation in R^1,

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{\frac{2\theta\delta^2}{\nu - 1}\left[1 + \left(\frac{X_t - \mu}{\delta}\right)^2\right]}dB_t, \quad X_0 = x. \tag{6}$$

Here

$$\nu > 2, \quad \theta > 0, \quad \delta > 0, \quad \mu \in R, \tag{7}$$

and $B = (B_t, \, t \geq 0)$ is a standard Brownian motion. Due to the classical Itô theorem, the stochastic differential equation above has a unique strong solution. It is known that this solution admits Markov and strong Markov property (cf., e.g., [4]). Student distribution $T(\nu, \delta, \mu)$ with the density

$$st_\nu(x) = \frac{c(\nu)}{\delta} \frac{1}{[1 + (\frac{x-\mu}{\delta})^2]^{(\nu+1)/2}} \qquad (x \in R^1), \qquad c(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$$

is stationary for this Markov process. This is easy to check by verifying the equation $L^*(st_\nu) = 0$ for the adjoint in $L_2(R^1)$ to the generator

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

that is,

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

Notice that parameter θ from the equation (6) does not show up in the expression for the stationary density st_ν. This is because its change only varies the speed scale of the process, and may be removed by a deterministic time change. In particular, if $X_0 \sim T(\nu, \delta, \mu)$, then X is stationary, and

$$E(X_{s+t}|X_s = x) = xe^{-\theta t} + \mu(1-e^{-\theta t}),$$

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

$$r(t) = \mathrm{Corr}(X_{s+t}, X_s) = e^{-\theta t}, t \geq 0, \quad s \geq 0.$$

3 Main results

Theorem 1 *For any $0 < m < \nu/2$ and any $0 < \alpha < 2m\theta(1-\dfrac{2m-1}{\nu-1})$, there exists $C > 0$ such that*

$$\beta^x(t) \leq C(x)e^{-\alpha t}, \quad C(x) = C(1+|(x-\mu)/\delta|^{2m}). \tag{8}$$

Also, there exists $C > 0$ such that

$$\|\mu^x(t) - \mu_\infty\|_{TV} \leq C(x)e^{-\alpha t}, \quad C(x) = C(1+|(x-\mu)/\delta|^{2m}). \tag{9}$$

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

$$\frac{S_t - tE_{st}f(X_0)}{\sqrt{t}} \overset{P_{st}}{\Longrightarrow} Z \sim \mathcal{N}(0,\sigma^2), \tag{10}$$

as well as in the non–stationary one,

$$\frac{S_t - tE_{st}f(X_0)}{\sqrt{t}} \stackrel{P_x}{\Longrightarrow} Z \sim \mathcal{N}(0, \sigma^2), \tag{11}$$

where $0 \le \sigma^2 := 2\int_0^\infty cov_{st}(f(X_0), f(X_t))\, dt < \infty$, *and all* P_{st}, E_{st} *and* cov_{st} *stand for the stationary regime.*

Boundedness of f certainly can be relaxed.

4 Auxiliary results

Firstly we will establish several lemmae.

Lemma 1 *For any* $0 < m < \nu/2$, $0 < \alpha < 2m\theta\left(1 - \frac{2m-1}{\nu-1}\right)$, *for* R *large enough, there exists* $C > 0$ *such that*

$$\sup_{t\ge 0} E_x e^{\alpha(t\wedge\tau)} \left|\frac{X_{t\wedge\tau} - \mu}{\delta}\right|^{2m} \le \left|\frac{x-\mu}{\delta}\right|^{2m}, \tag{12}$$

and

$$E_x e^{\alpha\tau} \left|\frac{X_\tau - \mu}{\delta}\right|^{2m} \le \left|\frac{x-\mu}{\delta}\right|^{2m}. \tag{13}$$

Proof. Denote

$$\frac{X_t - \mu}{\delta} =: Y_t, \tag{14}$$

then

$$dY_t = -\theta Y_t\, dt + \sqrt{\frac{2\theta}{\nu - 1}(1 + Y_t^2)}\, dW_t.$$

Applying Itô's formula to the process $f(t, Y_t) := \exp{(\alpha t)}|Y_t|^{2m}$ for $t < \tau$, we have,

$$df(t, Y_t) = \alpha e^{\alpha t} Y_t^{2m} dt + 2m e^{\alpha t} Y_t^{2m-1} dY_t + 1m(2m-1) e^{\alpha t} Y_t^{2m-2} (dY_t)^2.$$

Here

$$dY_t = C_1 Y_t dt + C_2\sqrt{1 + Y_t^2} dW_t,$$

with $C_1 = -\theta$ and $C_2 = \sqrt{\dfrac{2\theta}{\nu - 1}}$, and, hence,

$$(dY_t)^2 = C_2^2(1 + Y_t^2)dt.$$

Thus,

$$df(t, Y_t) = \alpha e^{\alpha t} Y_t^{2m} dt + (2me^{\alpha t} Y_t^{2m-1})(C_1 Y_t dt + C_2\sqrt{1 + Y_t^2} dW_t)$$

$$+(m(2m-1)C_2^2)e^{\alpha t}(Y_t^{2m-2})(1 + Y_t^2)dt$$

$$= e^{\alpha t}(\alpha + (\alpha + 2mC_1 + m(2m-1)C_2^2)Y_t^{2m} + (C_2^2 m(2m-1))Y_t^{2m-2})dt$$

$$+2mC_2 e^{\alpha t} Y_t^{2m-1}\sqrt{1 + Y_t^2} dW_t$$

$$= e^{\alpha t} Y_t^{2m}(\alpha + 2mC_1 + m(2m-1)C_2^2 + m(2m-1)C_2^2 Y_t^{-2} + \alpha Y_t^{-2m})dt$$

$$+(2me^{\alpha t} C_2 Y_t^{2m-1}\sqrt{1 + Y_t^2})dW_t.$$

Let us assume,

$$\alpha + 2mC_1 + m(2m-1)C_2^2 < 0 \quad \sim \quad \alpha < -m(2C_1 - (2m-1)C_2^2),$$

that is (remind that $C_1 = -\theta$, $C_2 = \sqrt{\dfrac{2\theta}{\nu - 1}}$, and $2m < \nu$),

$$\alpha < m(2\theta - (2m-1)\frac{2\theta}{\nu - 1}) = 2\theta m(1 - \frac{2m-1}{\nu - 1}).$$

With any such value of α and for R large enough, we have,

$$(\alpha + 2mC_1 + m(2m-1)C_2^2 + m(2m-1)C_2^2 Y_t^{-2} + \alpha Y_t^{-2m}) \leq -c < 0.$$

Therefore,

$$\boxed{E_y f(t \wedge \tau, Y_{t\wedge\tau}) - f(0, y) \leq -c\, E_y \int_0^{t\wedge\tau} f(s, Y_s)\, ds.} \tag{15}$$

Hence, in particular, we obtain, for R large enough,

$$Ef(t \wedge \tau, Y_{t\wedge\tau}) - f(0, Y_0) \le 0.$$

This is equivalent to (12). As $t \to \infty$, by virtue of Fatou's lemma we also have,

$$Ef(\tau, Y_\tau) \le f(0, Y_0).$$

This is equivalent to (13). The Lemma 1 is proved.

Remark 1 *A standard procedure of localization can be used to avoid a question about why expectation of the stochastic integral vanishes.*

Lemma 2 *For any $0 < m < \nu/2$, for R large enough, there exists $C > 0$ such that*

$$\sup_{t\ge 0} E_x \left|\frac{X_t - \mu}{\delta}\right|^{2m} \le C\left(1 + \left|\frac{x-\mu}{\delta}\right|^{2m}\right). \qquad (16)$$

Proof. In terms of the process Y, we are to show the inequality,

$$\sup_{t\ge 0} E_y \left|Y_t\right|^{2m} \le C\left(1 + |y|^{2m}\right). \qquad (17)$$

Let us apply $f(t, y) := \exp(\alpha t) y^{2m}$ with $\alpha = 0$; then it is natural to use a notation $f(y)$, $y \in R^1$. Then from the calculus in the proof of the Lemma 1, it follows that, in fact, for any $0 \le t_1 \le t_2 < \infty$,

$$E_y f(Y_{t_2}) - E_y f(Y_{t_1})$$

$$\le -cE_y \int_{t_1}^{t_2} f(Y_s) 1(|Y_s| \ge R)\, ds + C_0 E_y \int_{t_1}^{t_2} f(Y_s) 1(|Y_s| < R)\, ds$$

$$\equiv -cE_y \int_{t_1}^{t_2} f(Y_s)\, ds + E_y \int_{t_1}^{t_2} f(Y_s)(C_0 1(|Y_s| < R) + c1(|Y_s| < R))\, ds$$

$$\le -cE_y \int_{t_1}^{t_2} f(Y_s)\, ds + C \int_{t_1}^{t_2} 1\, ds \equiv -c\int_{t_1}^{t_2} E_y f(Y_s)\, ds + C\,(t_2 - t_1).$$

Therefore, if the derivative of the function $g(t) := E_y f(Y_t) \ge 0$ exists, it must satisfy the inequality

$$g'(t) \le -cg(t) + C. \qquad (18)$$

Clearly, the function g is well defined and is continuous with respect to t. The fact that it is differentiable follows from the calculus similar to that in the proof of the Lemma 1 (with $\alpha = 0$) if we do not use inequalities, but just write down the Itô formula and compute expectations taking into account the already established fact that expectation of the stochastic integral vanishes. Finally, any non-negative bounded continuous function which obeys the inequality (18) ought to satisfy the inequality

$$g(t) \le g(0)\exp(-ct) + \frac{C}{c} \quad \sim \quad E_y Y_t^{2m} \le y^{2m}\exp(-ct) + \frac{C}{c}.$$

This is even slightly stronger than desired (17). The Lemma 2 is proved.

Lemma 3 *For any* $1 < m < \nu/2$,

$$\int |x|^{2m} \mu_\infty(dx) < \infty.$$

Proof follows from the explicit representation of the stationary density.

Now we consider a direct product of two identical probability spaces with Wiener processes, where two independent copies of our Markov process $(X_t,\, t \ge 0)$, and $(\tilde X_t,\, t \ge 0)$, with initial data x and $\tilde x$ are defined. The notations for probability and expectation remains unchanged. Let $\gamma \equiv \gamma_{R_1} = \inf(t \ge 0 : \ |X_t| \vee |\tilde X_t| \le R_1)$, $\gamma(t) = \min(\gamma, t)$. In the following Lemma we assume that the constant R is already chosen and fixed.

Lemma 4 *For any* $0 < m < \nu/2$, *there exist* $R_1 \ge R$, $C > 0$ *such that*

$$E_{x,\tilde x} e^{\alpha\gamma} \le C\left(1 + \left|\frac{x-\mu}{\delta}\right|^{2m} + \left|\frac{\tilde x - \mu}{\delta}\right|^{2m}\right). \tag{19}$$

Proof of the Lemma 4. To show the desired bound, let us consider the Lyapunov function with some $\alpha > 0$ in variables y and $\tilde y$,

$$f(t, y, \tilde y) = \exp(\alpha t)(|y|^{2m} + |\tilde y|^{2m}).$$

Similarly to the proof of the Lemma 1, applying Itô's formula, we obtain,

$$\begin{aligned}
df(t, Y_t, \tilde{Y}_t) &= \alpha \exp(\alpha t)(|Y_t|^{2m} + |\tilde{Y}_t|^{2m})dt + 2m \exp(\alpha t)|Y_t|^{2m-1} dY_t \\
&+2m \exp(\alpha t)|\tilde{Y}_t|^{2m-1} d\tilde{Y}_t + \frac{1}{2}\exp(\alpha t)\, 2m\,(2m-1)|Y_t|^{2m-2}(dY_t)^2 \\
&+\frac{1}{2}\exp(\alpha t)\, 2m\,(2m-1)|\tilde{Y}_t|^{2m-2}(d\tilde{Y}_t)^2 \\
&= \exp(\alpha t)\Big[\alpha(Y_t^{2m} + \tilde{Y}_t^{2m}) + 2mC_1(Y_t^{2m} + \tilde{Y}_t^{2m}) \\
&+m(2m-1)(Y_t^{2m} + Y_t^{2m-2} + \tilde{Y}_t^{2m} + \tilde{Y}_t^{2m-2})C_2^2\Big]\, dt \\
&+ \exp(\alpha t)\, 2m\, \left(Y_t^{2m-1}\sqrt{1+Y_t^2} + (\tilde{Y}_t)^{2m-2}\sqrt{1+(\tilde{Y}_t)^2}\right) C_2 dW_t.
\end{aligned}$$

To get the desired bound, we ought to have a negative expression with dt, that is,

$$\begin{aligned}
&1(t < \gamma)\, \Big[\alpha(Y_t^{2m} + \tilde{Y}_t^{2m}) + 2mC_1(Y_t^{2m} + \tilde{Y}_t^{2m}) \\
&+m(2m-1)(Y_t^{2m} + Y_t^{2m-2} + \tilde{Y}_t^{2m} + \tilde{Y}_t^{2m-2})C_2^2\Big] \le 0.
\end{aligned} \tag{20}$$

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

(I) $$|Y_t| \ge R_1 \ \& \ |\tilde{Y}_t| \ge R_1,$$

or

(II) $$|Y_t| \ge R_1 \ \& \ R \le |\tilde{Y}_t| < R_1,$$

or

(III) $$|Y_t| \ge R_1 \ \& \ |\tilde{Y}_t| < R.$$

In the cases (I) and even (II), the whole expression (20) is negative for any $R_1 \ge R$, precisely due to the calculus in the proof of the Lemma 1, because, due to the choice of R,

$$\left[\alpha Y_t^{2m} + 2mC_1 Y_t^{2m} + m(2m-1)(Y_t^{2m} + Y_t^{2m-2})C_2^2\right] < 0,$$

and

$$\Big[\alpha\tilde Y_t^{2m} + 2mC_1\tilde Y_t^{2m} + m(2m-1)(\tilde Y_t^{2m} + \tilde Y_t^{2m-2})C_2^2\Big] < 0.$$

In the case (III), we cannot guarantee that the terms with $\tilde Y$ provide a negative input: it may turn out to be positive (because $|y|^{2m-2} \geq |y|^m$ for $|y| < 1$). However, given R, the positive part of all these terms is bounded by modulus,

$$\sup_{|y|\le R}\left(2mC_1y^{2m} + m(2m-1)(y^{2m} + y^{2m-2})C_2^2\right) =: C_0 < \infty.$$

(This constant even does not depend on R, although if it were, the change would have been minimal.) Let us choose R_1 so large that for any $|y| \geq R_1$,

$$\left|2mC_1y^{2m} + m(2m-1)(y^{2m} + y^{2m-2})C_2^2\right| >> C_0.$$

Since negative part due to Y_t is of the order at least R_1^{2m}, it exceeds $g(R)$ by modulus. Hence, the desired statement follows as in the proof of the Lemma 1.

5 Proof of the Theorem 1

1. Consider two independent copies of our Markov process X and $\tilde X$, both being solutions of the equation (6) with two different independent Wiener processes, correspondingly, $(W_t,\, t \geq 0)$ and $(\tilde W_t,\, t \geq 0)$, and with deterministic initial values $X_0 = x \in R$ and $\tilde X_0 = \tilde x \in R$. Later on, in certain steps of the proof, $(\tilde x, \tilde y)$ will be chosen randomly. After change of variables as in (14), denote corresponding processes by Y and $\tilde Y$. Fix $s_0 \geq 0$.
Consider a sequence of stopping times, $\gamma_1 < \gamma_2 < ...$, defined as follows:

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

and for $n \geq 1$ by induction,

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

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

Due to the Lemma 1, we have *a priori* bounds

$$E(\exp(\alpha(\gamma_1 - s_0))|\hat F_{s_0}) \leq C(1 + |Y_{s_0}|^{2m} + |\tilde Y_{s_0}|^{2m}).$$

and

$$E(\exp \alpha(\gamma_{n+1} - \gamma_n)|\hat{F}_{\gamma_n}) \le C.$$

Here initial values are irrelevant and, hence, dropped.

2. We will use a coupling procedure as in [6] and [7], based on the Harnack inequality. Given $Y_{\gamma_n}, \tilde{Y}_{\gamma_n}$, consider the exit measures of both processes on the parabolic boundary Γ of the cylinder

$$(t, y, \tilde{y}) : \ \gamma_n \le t \le \gamma_n + 1, \ |y| \le R + 1, \ |\tilde{y}| \le R + 1,$$

that is,

$$\Gamma = \{(t, y, \tilde{y}) : \ \gamma_n \le t \le \gamma_n + 1, \ |y| = R + 1, \ |\tilde{y}| = R + 1\}$$

$$\bigcup \{(t, y, \tilde{y}) : \ t = \gamma_n + 1, \ |y| \le R + 1, \ |\tilde{y}| \le R + 1\}.$$

Let $\delta \in (0, 1)$. Consider, in particular, the part of Γ,

$$\Gamma_\delta = \{(t, y, \tilde{y}) : \ \gamma_n + \delta \le t \le \gamma_n + 1, \ |y| = R + 1, \ |\tilde{y}| = R + 1\}$$

$$\bigcup \{(t, y, \tilde{y}) : \ t = \gamma_n + 1, \ |y| \le R + 1, \ |\tilde{y}| \le R + 1\}.$$

Due to the Harnack inequality [5], exit measures of both our processes Y and $\tilde{Y}$ on Γ_δ are equivalent with a bounded derivative. Therefore, on the whole Γ, we have the following inequality,

$$\inf_{y, \tilde{y} \in B_R} \int_\Gamma \left(\frac{P_y((T, Y_T) \in dv)}{P_{\tilde{x}}((T, Y_T) \in dv)} \wedge 1 \right) P_{\tilde{x}}((T, Y_T) \in dv) =: c > 0,$$

which we call Dobrushin's local mixing condition and where dv is the area element on Γ.

Hence, e.g., by using the technique from [8, Section 2.4]), it is possible to choose some new representations of Y and $\tilde{Y}$ on some extended probability space at their exit on Γ so that they coincide with a probability

$$P_{Y_{\gamma_n}, \tilde{Y}_{\gamma_n}}((T_n, Y_{T_n}) = (T_n, \tilde{Y}_{T_n})) \ge c.$$

Of course, those representations are constructed on some extension of the original of our probability space. Moreover, given values at γ_n and at the exit on Γ, say, T_n, we can recover the trajectory between those two values as a conditional measure.

An important feature is that now, for this new representation,

$$P_{F_{\gamma_n}}(Y_{T_n} = \tilde Y_{T_n}) \ge c > 0. \tag{21}$$

Hence, we can denote the coupling moment

$$L := \inf(T_n : \ Y_{T_n} = \tilde Y_{T_n}).$$

For full details we refer to [8].

3. In dimension one, coupling may be equivalently presented by using a more simple idea of intersection. Consider the change of variables,

$$y \mapsto \sqrt{\frac{\nu-1}{2\theta}}\,(y + \sqrt{1+y^2}).$$

In this new scale, the equation (6) reads (slightly abusing notations, we do not change them for the process in the new coordinate),

$$dY_t = \tilde b(Y_t)\,dt + dW_t, \quad Y_0 = y,$$

with

$$\tilde b(y) = -\sqrt{\frac{2\theta(\nu-1)^2+\theta}{2(\nu-1)}}\,\frac{\exp(2y)-1}{\exp(2y)+1},$$

that is, diffusion coefficient in this scale equals one, while the drift $\tilde b$ is bounded. Instead of levels R and $R+1$ as in the previous paragraph and previous technique, here it is more convenient to introduce levels R and $R+K$, where K is to be adjusted: this value in a minute will be chosen large enough, first of all, so as to dominate the drift, $K/2 \ge \|\tilde b\|_C$. Consider the following event,

$$A := \{(Y_{\gamma_n} - \tilde Y_{\gamma_n})Y_{\gamma_n+1} \le 0, \quad \& \quad (Y_{\gamma_n} - \tilde Y_{\gamma_n})\tilde Y_{\gamma_n+1} \ge 0\}.$$

Given any $Y_{\gamma_n}, \tilde Y_{\gamma_n}$, probability of this event admits the bound

$$P(A) \ge \Phi(-R - \|\tilde b\|)^2. \tag{22}$$

Indeed, assume, for example, $Y_{\gamma_n} > \tilde Y_{\gamma_n}$. Then, the minimum of this probability is attained, apparently, if $Y_{\gamma_n} = R, \tilde Y_{\gamma_n} = -R$, and, moreover,

$$A \supset \{W_{\gamma_n+1} - W_{\gamma_n} \le -R - \|\tilde b\|, \quad \& \quad \tilde W_{\gamma_n+1} - \tilde W_{\gamma_n} \ge R + \|\tilde b\|\},$$

which shows the estimate (22).

If the event A occurs, the trajectories of Y and $\tilde Y$ on $[\gamma_n, \gamma_n + 1]$ must intersect. Their first intersection on this interval is a stopping time, so after this first meeting we may "glue" the two processes, for example, letting the first Y follow the second $\tilde Y$. There is only one little nuisance if we do this construction straight, that before this intersection one or both trajectories may already happen to be far away from the origin, e.g., escape from $[-R-K, R+K]$. To avoid some technicalities, it is convenient to stop the processes at such first exit time. In other words, we do not try to couple them until $\gamma_n + 1$, but start the procedure again, i.e. wait until next γ_{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 large enough, to make such escape from $[-R-K, R+K]$ on time interval $[\gamma_n, \gamma_n + 1]$ unlikely, so that for the event

$$A_K := \{(Y_{\gamma_n} - \tilde Y_{\gamma_n})Y_{\gamma_n+1} \le 0, \quad \& \quad (Y_{\gamma_n} - \tilde Y_{\gamma_n})\tilde Y_{\gamma_n+1} \ge 0,$$

$$\& \quad \sup_{\gamma_n \le t \le \gamma_n+1} (|Y_t| \vee |\tilde Y_t|) < R + K\},$$

we have, say,

$$P(A_K) \ge \Phi(-R - \|\tilde b\|)^2/2. \tag{23}$$

Clearly, this is possible if K is chosen large enough. Now, coupling via intersection L will be accepted only if this intersection occurs on $[\gamma_n, \gamma_n + 1]$ earlier than exit time $\tilde T_n$, where

$$\tilde T_n := \inf(t \ge \gamma_n : \ |Y_t| \vee |\tilde Y_t| \ge R + K) \wedge (\gamma_n + 1) < \gamma_n + 1.$$

Probability of the latter is at least as large as $P(A_K)$ which admits the estimate (23),

$$P(\text{intersection earlier than } \tilde T_n \mid F_{\gamma_n}) \ge \Phi(-R - \|\tilde b\|)^2/2. \tag{24}$$

If, however, $\tilde T_n$ occurs earlier, then we simply wait till next γ_{n+1} to repeat this procedure. Notice that we could achieve a slightly better lower bound than (23) by considering probability of the event,

$$A' := \{(Y_{\gamma_n} - \tilde Y_{\gamma_n})(Y_{\gamma_n+1} - \tilde Y_{\gamma_n+1}) \le 0\},$$

namely,

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

Clearly, A' also implies intersection. In the remaining part of the proof, we work with the sequence of stopping times (γ_n, T_n) as suggested in step 2 above. However, the same can be done similarly with $(\gamma_n, \tilde{T}_n)$.

4. Hence, we now analyze the following iterative procedure of "attempts" to couple the two processes at $L = T_1, T_2, \ldots$. Otherwise, if we wished to use intersections, we could have defined coupling time L as

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

Notice that if the two processes were not coupled before γ_n, they may be coupled at T_n with probability (21) (or via $(\tilde{T}_n)$, in which case we would use (24)). The moment L when they are coupled is a stopping time. Due to strong Markov property, after being coupled, the two processes trajectories *can* be continued as identical; for example, the first process may follow the second one. This does not change the fact that each process is a strong Markov one which solves the equation (6). Our procedure of coupling automatically provides the bound,

$$P_y(L_{s_0} > T_n) \leq (1 - c)^n \equiv q^n. \qquad (25)$$

A similar bound holds true in the version via intersections,

$$P_y(L_{s_0} > \tilde{T}_n) \leq (1 - \tilde{c})^n \equiv \tilde{q}^n. \qquad (26)$$

Now, $\forall B \in \mathcal{B}(R)$, we have to estimate the difference,

$$|P_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - P_y(Y_{s_0+t} \in B)|. \qquad (27)$$

To this aim, we consider the second independent version of our process X, say, $\tilde{X}$, *started at* s_0*, with a distribution* $\mu^x_{s_0}$, i.e., the same as the distribution of X_{s_0} itself; in terms of the processes Y and $\tilde{Y}$, we will consider $\tilde{Y}$ started at s_0 with a distribution $\nu^y_{s_0}$ (we use notation ν^y_s for distribution of Y_s started at y). We have,

$$|P_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - P_y(Y_{s_0+t} \in B)|$$

$$\equiv |P_y(Y_{s_0+t} \in B \mid \hat{F}_{s_0}) - P(\tilde{Y}_{s_0+t} \in B)|.$$

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

$$|P_{y_1}(Y_t \in B) - P_{y_2}(\tilde{Y}_t \in B)| \le P(L > t) \qquad (28)$$

$$\le E(1(L > t)1(t < T_0)) + \sum_{n=0}^{\infty} E(1(L > t)1(T_n \le t < T_{n+1}))$$

$$\le E(1(t < T_0)) + \sum_{n=0}^{\infty} P(L > T_n)^{1/a} P(T_{n+1} > t)^{1/c}.$$

As we already mentioned,

$$P(L > T_n) \le q^n.$$

Due to Bienaimé–Chebyshev's inequality,

$$P(\gamma_{n+1} > t) \le \exp(-\alpha t) E \exp(\gamma_{n+1})$$

$$\le \exp(-\alpha t) E \exp(\alpha[(\gamma_{n+1} - \gamma_n) + (\gamma_n - \gamma_{n-1}) + \ldots + (\gamma_2 - \gamma_1) + \gamma_1])$$

$$\equiv \exp(-\alpha t) E \exp(\alpha[(\gamma_n - \gamma_{n-1}) + \ldots + (\gamma_2 - \gamma_1) + \gamma_1])$$

$$\times E_{\gamma_n, Y_{\gamma_n}, \tilde{Y}_{\gamma_n}} \exp(\alpha(\gamma_{n+1} - \gamma_n)).$$

Due to the Lemma 2,

$$E_{\gamma_n, Y_{\gamma_n}, \tilde{Y}_{\gamma_n}} \exp(\alpha(\gamma_{n+1} - \gamma_n)) \le C.$$

By induction, we get,

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

Since $\gamma_{n+1} \le T_{n+1} \le \gamma_{n+1} + 1$, we also have,

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

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

$$P(L > t) = E1(L > t)1(t < T_1) + \sum_{n=1}^{\infty} E1(L > t)1(T_n \le t < T_{n+1})$$

$$\le E1(t < T_1) + \sum_{n=1}^{\infty} P(L > T_n)^{1/a} P(T_{n+1} > t)^{1/c}$$

$$\le P(t < T_1) + \sum_{n=1}^{\infty} q^{n/a} P(T_{n+1} > t)^{1/c}$$

$$\le C\, \exp(-\alpha(t-1)) \int (1 + |Y_{s_0}|^{2m} + |\tilde{y}|^{2m})\, \mu^y_{s_0}(d\tilde{y})$$

$$+ \sum_{n\ge 1}^{\infty} q^{n/a} \exp(-\alpha(t-1))(C^{n+1})^{1/c} \int (1 + |Y_{s_0}|^{2m} + |\tilde{y}|^{2m})^{1/c}\, \mu^y_{s_0}(d\tilde{y})$$

$$\le C \exp(-\alpha t)(1 + |Y_{s_0}|^{2m}),$$

the latter if we choose a, c so that

$$q^{1/a} C^{1/c} < 1.$$

Whence, with this choice we estimate beta-mixing coefficient for the process Y by virtue of the Lemma 2,

$$\beta^y_t \le \sup_{s\ge 0} C \exp(-\alpha t) E_y (1 + |Y_{s_0}|^{2m})$$

$$\le C \exp(-\alpha t)(1 + |y|^{2m}).$$

In terms of the original process X this reads,

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

5. The total variation distance is estimated in a similar way. Let $s_0 = 0$, and consider an independent couple of Markov processes X and $\tilde{X}$ – or correspondingly Y and $\tilde{Y}$ – where now $\tilde{X}$ is stationary, as well as $\tilde{Y}$. Coupling procedure is similar to step 2 or 3 above. We have, similarly to (28) and due to the Lemma 3,

$$|P_y(Y_t \in B) - P_\nu(\tilde{Y}_t \in B)| \le P(L > t)$$

$$\le C\, \exp(-\alpha(t-1)) \int (1 + |y|^{2m} + |\tilde{y}|^{2m})\, \nu_\infty^y(d\tilde{y})$$

$$\le C\, \exp(-\alpha t)(1 + |y|^{2m}).$$

In terms of the process X, this implies the bound (9). The Theorem 1 is proved.

Proof of Corollary 1 follows directly from [3, Theorem 18.4.1], by virtue of the Theorem 1, moment inequality $E_{st}|X_0|^{2m} < \infty$, under assumption $f \in L_2(\nu)$. (This is where the assumption on boundedness of f can be relaxed.)

The non-stationary version of CLT follows from the following remark which reduces the case to the stationary version. We have, with $\tilde{X}$ a stationary version of the process and L for coupling time,

$$\frac{\int_0^t f(X_s)\, ds - t\, Ef(X_s)}{\sqrt{t}}$$

$$\equiv \frac{\int_0^t f(\tilde{X}_s)\, ds - t\, Ef(\tilde{X}_s)}{\sqrt{t}} + \frac{\int_0^t (f(X_s) - f(\tilde{X}_s))1(s \le L)\, ds}{\sqrt{t}}\,.$$

Here the first term weakly converges to a Gaussian random variable, due to

(10). The second term admits an estimate,

$$\left| \frac{\int_0^t (f(X_s) - f(\tilde{X}_s)) 1(s \le L)\, ds}{\sqrt{t}} \right|$$

$$\le E_{x,\mu} \frac{\int_0^t |f(X_s) - f(\tilde{X}_s)|\, 1(s \le L)\, ds}{\sqrt{t}}$$

$$\le \frac{2||f||_B}{\sqrt{t}} \int_0^t P_{x,\mu}(L \ge s)\, ds$$

$$\le \frac{2||f||_B}{\sqrt{t}}\, C_x\, \frac{1}{c}(1 - \exp(-ct)) \to 0, \quad t \to \infty.$$

Hence, (11) follows, and The Corollary is proved.

Acknowledgements

The second author thanks the grant RFBR 08-01-00105a for support.

References

[1] Heyde, C. C., Leonenko, N. N. Student processes. Adv. in Appl. Probab. 37(2) (2005), 342–365.

[2] Heyde, C. C., Leonenko, N. N., Šuvak, N. Statistical inference for Student diffusion process, submitted.

[3] Ibragimov, I. A., Linnik, Yu. V. Independent and stationary sequences of random variables. Wolters-Noordhoff Publ., Groningen, 1971.

[4] Krylov, N. V. The selection of a Markov process from a Markov system of processes, and the construction of quasidiffusion processes. (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973), 691–708.

[5] Krylov, N. V., Safonov, M. V. A property of the solutions of parabolic equations with measurable coefficients. (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 44(1) (1980), 161–175, 239.

[6] Veretennikov, A. Yu. Estimates of the mixing rate for stochastic equations. (Russian) Teor. Veroyatnost. Primenen. 32 (1987), no. 2, 299–308; Engl. transl.: Theory Probab. Appl. 32 (1987), no. 2, 273–281.

[7] Veretennikov, A. Yu. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Teor. Veroyatn. Primenen. 44 (1999), 2, 312–327; Engl. transl.: Theory Probab. Appl. 44 (2000), 2, 361–374.

[8] Veretennikov, A. Yu. On approximations of diffusions with equilibrium, Helsinki University of Technology, Institute of Mathematics Reports C17 (2004); http://math.tkk.fi/visitors0405/AVslides.pdf

[9] Veretennikov, A. Yu.; Klokov, S. A. On the subexponential rate of mixing for Markov processes. (Russian) Teor. Veroyatn. Primenen. 49 (2004), no. 1, 21–35.