Bernoulli **19**(5B), 2013, 2294–2329
DOI: 10.3150/12-BEJ453

Ergodicity and mixing bounds for the Fisher–Snedecor diffusion

A.M. KULIK[1] and N.N. LEONENKO[2]

[1]*Institute of Mathematics, Ukrainian National Academy of Sciences, 01601 Tereshchenkivska str. 3, Kyiv, Ukraine. E-mail: kulik@imath.kiev.ua*
[2]*School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4AG, UK. E-mail: LeonenkoN@Cardiff.ac.uk*

We consider the Fisher–Snedecor diffusion; that is, the Kolmogorov–Pearson diffusion with the Fisher–Snedecor invariant distribution. In the nonstationary setting, we give explicit quantitative rates for the convergence rate of respective finite-dimensional distributions to that of the stationary Fisher–Snedecor diffusion, and for the β-mixing coefficient of this diffusion. As an application, we prove the law of large numbers and the central limit theorem for additive functionals of the Fisher–Snedecor diffusion and construct P-consistent and asymptotically normal estimators for the parameters of this diffusion given its nonstationary observation.

Keywords: β-mixing coefficient; central limit theorem; convergence rate; Fisher–Snedecor diffusion; law of large numbers

1. Introduction

In this paper, we investigate the Markov process X, valued in $(0,\infty)$, defined by the nonlinear stochastic differential equation

$$\mathrm{d}X_t = -\theta(X_t-\kappa)\,\mathrm{d}t + \sqrt{2\theta X_t\left(\frac{X_t}{\beta/2-1}+\frac{\kappa}{\alpha/2}\right)}\,\mathrm{d}W_t, \qquad t\ge 0. \tag{1.1}$$

Such a process belongs to the class of diffusion processes with invariant distributions from the *Pearson family*, introduced by K. Pearson [24] in 1914 in order to unify some of the most important statistical distributions. The study of such processes was started in the 1930s by A.N. Kolmogorov [17,27], hence it seems appropriate to call this important class of processes the *Kolmogorov–Pearson (KP) diffusions*. For a more detailed discussion of KP diffusions, we refer to recent papers [11,26] and [5].

When $\alpha,\beta>2$, the process X defined by (1.1) is ergodic [12]. Under the particular choice $\kappa=\beta/(\beta-2)$, respective unique invariant distribution coincides with the *Fisher–Snedecor distribution* $\mathcal{FS}(\alpha,\beta)$ with α,β degrees of freedom; that is, its probability density is given by

$$\mathfrak{fs}(x)=\frac{1}{xB(\alpha/2,\beta/2)}\left(\frac{\alpha x}{\alpha x+\beta}\right)^{\alpha/2}\left(\frac{\beta}{\alpha x+\beta}\right)^{\beta/2}, \qquad x>0. \tag{1.2}$$

1350-7265

This is the reason to call the process X defined by (1.1) *the Fisher–Snedecor diffusion*. Together with *the reciprocal gamma* and *the Student* diffusions, the Fisher–Snedecor diffusion forms the class of the so-called *heavy-tailed* KP diffusions. Statistical inference for three heavy-tailed KP diffusions is developed in the recent papers [22,23] and [5] in the situation where the stationary version of the respective diffusion is observed.

In this paper, we consider the Fisher–Snedecor diffusion (1.1) in the nonstationary setting; that is, with arbitrary distribution of the initial value X_0. We give explicit quantative rates for *the convergence rate* of respective finite-dimensional distributions to that of the stationary Fisher–Snedecor diffusion, and for *the β-mixing coefficient* of this diffusion. Same problems for the reciprocal gamma and the Student diffusions were considered in [1] and [2], respectively. Similarly to [1] and [2], our way to treat this problem is based on the general theory developed for (possibly nonsymmetric and nonstationary) Markov processes, although there is a substantial novelty in the form taken by the *Lyapunov-type condition* (typical in the field) in our setting.

As an application, we prove the law of large numbers (LLN) and the central limit theorem (CLT) for additive functionals of the Fisher–Snedecor diffusion. Note that, for the stationary version of the diffusion, these limit theorems are well known: LLN is provided by the Birkhoff–Khinchin theorem, and CLT is available either in the form based on the α-mixing coefficient of a stationary sequence or process (see [14]), or in the form formulated in terms of the L_2-semigroup associated with the Markov process (see [6]). Our considerations are based on the natural idea to extend these results to the nonstationary setting using the bounds for the deviation between the stationary and nonstationary versions of the process. The way we carry out this idea differs, for instance, from those proposed in [6], Theorem 2.6, or in [3], Section 4.II.1.10, and is based on the notion of an *(exponential) ϕ-coupling*, introduced in [19] as a tool for studying convergence rates of L_p-semigroups, generated by a Markov process, and spectral properties of respective generators.

The modified version of the Lyapunov-type condition, mentioned above, implies a substantial difference between the asymptotic properties of the finite-dimensional distributions themselves and their continuous-time averages, see Theorem 3.2 and Remark 3.2 below. An important consequence is that, in the continuous-time version of our CLT, the observable functional *may fail to be square integrable* w.r.t. the invariant distribution of the process. This interesting effect seemingly has not been observed in the literature before.

Finally, we apply the above results and provide a statistical analysis for the Fisher–Snedecor diffusion. In the situation where a nonstationary version of the diffusion X is observed, we prove that respective empirical moments and empirical covariances are P-consistent, asymptotically normal, and (under some additional assumptions on the initial distribution of X) asymptotically unbiased. Then, using the method of moments, we construct P-consistent and asymptotically normal estimators for the parameter $(\alpha, \beta, \kappa, \theta)$ given either the discrete-time or the continuous-time observations of a nonstationary version of the Fisher–Snedecor diffusion. To keep the current paper reasonably short, we postpone the explicit calculation of the asymptotic covariance matrices and a more detailed discussion of other statistical aspects to the subsequent paper [20].

2. Preliminaries

In this section, we introduce briefly main objects, assumptions, and notation.

For the Fisher–Snedecor diffusion (1.1), the drift coefficient $a(x)$ and the diffusion coefficient $\sigma(x)$ are respectively given by

$$a(x) = -\theta(x-\kappa), \qquad \sigma(x) = \sqrt{2\theta x\left(\frac{x}{\beta/2-1} + \frac{\kappa}{\alpha/2}\right)}, \tag{2.1}$$

and our standing assumptions on the parameters are

$$\theta > 0, \qquad \kappa > 0, \qquad \beta > 2, \qquad \alpha > 2. \tag{2.2}$$

We assume that, on a proper probability space $(\Omega, P, \mathcal{F})$, independent Wiener process W and random variable X_0 taking values in $(0,\infty)$ are well defined. Then, because the coefficients (2.1) are continuously differentiable inside $(0,\infty)$, the unique strong solution to equation (1.1) with the initial condition X_0 is well defined up to the random time moment $T_{0,\infty}$ of its exit from $(0,\infty)$.

For $x \in (0,\infty)$, the corresponding *scale density* equals

$$\mathfrak{s}(x) = \exp\left(-\int_1^x \frac{2a(u)}{\sigma^2(u)}\,\mathrm{d}u\right) = Cx^{-\alpha/2}\left(x + \frac{\kappa(\beta-2)}{\alpha}\right)^{\alpha/2+\beta/2-1}. \tag{2.3}$$

Here and below, by C we denote a constant, which can be (but is not) expressed explicitly; the value of C can vary from place to place. It follows from the standing assumption (2.2) that

$$\int_x^\infty \mathfrak{s}(y)\,\mathrm{d}y = \infty, \qquad \int_0^x \mathfrak{s}(y)\,\mathrm{d}y = \infty, \qquad x \in (0,\infty),$$

and consequently both 0 and ∞ are unattainable points for the diffusion X, that is, the random time moment $T_{0,\infty}$ is a.s. infinite for any positive initial condition X_0 (e.g., [16], Chapter 18.6). This means that (1.1) uniquely determines a time-homogeneous strong Markov process X with the state space $\mathbb{X} = (0,\infty)$. In the sequel, we consider $\mathbb{X}$ as a locally compact metric space with the metric $d(x,y) = |x-y| + |x^{-1} - y^{-1}|$.

Let us introduce the notation. By $P_t(x, \mathrm{d}y)$, we denote the transition probabilities of the process X. By $\mathcal{P}$ we denote the class of probability distributions on the Borel σ-algebra on $\mathbb{X}$. For any $\mu \in \mathcal{P}$, we denote by P_μ the distribution in $C(\mathbb{R}^+, \mathbb{X})$ of the solution to (1.1) with the distribution of X_0 equal μ, and write E_μ for the respective expectation. When $\mu = \delta_x$, the measure concentrated at the point $x \in \mathbb{X}$, we write P_x, E_x instead of P_μ, E_μ. For any $\mu \in \mathcal{P}$ we denote by $\mu_{t_1,\dots,t_m}$, $0 \le t_1 < \dots < t_m$, $m \ge 1$ the family of finite-dimensional distributions of the process X with the initial distribution μ; that is,

$$\begin{aligned}\mu_{t_1,\dots,t_m}(A) &= \int_{\mathbb{X}}\int_A P_{t_1}(x,\mathrm{d}x_1)P_{t_2-t_1}(x_1,\mathrm{d}x_2)\cdots P_{t_m-t_{m-1}}(x_{m-1},\mathrm{d}x_m)\mu(\mathrm{d}x) \\ &= P_\mu\big((X_{t_1},\dots,X_{t_m}) \in A\big), \qquad A \in \mathcal{B}\big(\mathbb{X}^m\big).\end{aligned} \tag{2.4}$$

By $\mathbb{F}^X = \{\mathcal{F}_t^X, t \ge 0\}$, we denote the natural filtration of the process X. A measurable function $f : \mathbb{X} \to \mathbb{R}$ is said to belong to the domain of the *extended generator* $\mathcal{A}$ of the process X if there exists a measurable function $g : \mathbb{X} \to \mathbb{R}$ such that the process

$$f(X_t) - \int_0^t g(X_s)\,\mathrm{d}s, \qquad t \in \mathbb{R}^+$$

is well defined and is an $\mathbb{F}^X$-martingale w.r.t. to any measure $P_x, x \in \mathbb{X}$. For such a pair (f, g), we write $f \in \mathrm{Dom}(\mathcal{A})$ and $\mathcal{A} f = g$.

For a measurable function $\phi : \mathbb{X} \to [1, \infty)$ and a signed measure $\varkappa$ on $\mathcal{B}(\mathbb{X}^m)$, define the *weighted total variation norm*

$$\|\varkappa\|_{\phi,\mathrm{var}} = \int_{\mathbb{X}^m} \big(\phi(x_1) + \cdots + \phi(x_m)\big)|\varkappa|(\mathrm{d}x),$$

where $|\varkappa| = \varkappa^+ + \varkappa^-$ and $\varkappa = \varkappa^+ - \varkappa^-$ is the Hahn decomposition of $\varkappa$. Frequently, we will use functions ϕ of the form

$$\phi = \phi_\diamondsuit + \phi_\blacklozenge, \tag{2.5}$$

where $\phi \geq 1$, $\phi_\diamondsuit, \phi_\blacklozenge \in C^2(0, \infty)$, $\phi_\diamondsuit = 0$ on $[2, \infty)$, $\phi_\blacklozenge = 0$ on $(0, 1]$,

$$\phi_\diamondsuit(x) = x^{-\gamma} \quad \text{for } x \text{ small enough}, \qquad \phi_\blacklozenge(x) = x^{\delta} \quad \text{for } x \text{ large enough}$$

with nonnegative γ, δ.

The *β-mixing* (or *complete regularity*, or *the Kolmogorov*) coefficient is defined as

$$\beta^\mu(t) = \sup_{s\geq 0} E_\mu \sup_{B \in \mathcal{F}^X_{\geq t+s}} \big|P_\mu\big(B|\mathcal{F}^X_s\big) - P_\mu(B)\big|, \qquad \mu \in \mathcal{P}, t \in \mathbb{R}^+, \tag{2.6}$$

where $\mathcal{F}^X_{\geq r}$ for a given $r \geq 0$ denotes the σ-algebra generated by the values of the process X at the time moments $v \geq r$. In particular, the state-dependent β-mixing coefficient is defined by

$$\beta_x(t) = \sup_{s\geq 0} E_x \sup_{B \in \mathcal{F}^X_{\geq t+s}} \big|P_x\big(B|\mathcal{F}^X_s\big) - P_x(B)\big|, \qquad x \in \mathbb{X}, t \in \mathbb{R}^+ \tag{2.7}$$

(in this case, the initial distribution $\mu = \delta_x$), and the stationary β-mixing coefficient is defined by

$$\beta(t) = \sup_{s\geq 0} E_\pi \sup_{B \in \mathcal{F}^X_{\geq t+s}} \big|P_\pi\big(B|\mathcal{F}^X_s\big) - P_\pi(B)\big|, \qquad x \in \mathbb{X}, t \in \mathbb{R}^+; \tag{2.8}$$

here and below, π denotes the (unique) invariant distribution for the process X. For more information about various types of mixing coefficients see, for example, [9].

3. Main results

Here, we formulate the main results of the paper. The proofs are postponed to Section 5.

3.1. Distributional properties of the Fisher–Snedecor diffusion

The following two basic properties of the Fisher–Snedecor diffusion will be used in the further analysis of its ergodic behavior.

Proposition 3.1. 1. (*Lyapunov-type condition*). *Let* ϕ *to have the form* (2.5) *with*

$$\gamma < \frac{\alpha}{2} - 1, \qquad \delta < \frac{\beta}{2}. \tag{3.1}$$

Then $\phi \in \operatorname{Dom}(\mathcal{A})$ *and*

$$\mathcal{A}\phi = a\phi' + \frac{1}{2}\sigma^2\phi''. \tag{3.2}$$

In addition, there exist a segment $[u, v] \subset (0, \infty)$ *and positive constants* c, C *such that*

$$\mathcal{A}\phi(x) \le -c\phi(x) + C\mathbf{1}_{[u,v]}(x). \tag{3.3}$$

2. (*Local minorization condition*). *For every segment* $[u, v] \subset \mathbb{X}$ *there exist* $T > 0$, *another segment* $[u', v'] \subset \mathbb{X}$ *and a constant* $c_{u,v,u',v',T} > 0$ *such that for every* $x \in [u, v]$ *and every Borel set* $A \subset [u', v']$

$$P_T(x, A) \ge c_{u,v,u',v',T} \int_A \mathrm{d}y.$$

The following moment bound is a well known corollary of the Lyapunov-type condition (see, e.g., Section 3.2 in [18] and references therein).

Corollary 3.1. *In the conditions and notation of statement* 1 *in Proposition* 3.1, *we have*

$$\int_{\mathbb{X}} \phi \, \mathrm{d}\mu_t \le \frac{C}{c} + \mathrm{e}^{-ct} \int_{\mathbb{X}} \phi \, \mathrm{d}\mu, \qquad t \in \mathbb{R}^+.$$

In addition, there exists an invariant measure $\mu^* \in \mathcal{P}$ *such that*

$$\int_{\mathbb{X}} \phi \, \mathrm{d}\mu^* < +\infty.$$

Because the Fisher–Snedecor diffusion is ergodic, the latter statement can be interpreted as the following fact about its (unique) invariant distribution π:

$$\int_{\mathbb{X}} x^{-\gamma} \pi(\mathrm{d}x) < +\infty, \qquad \int_{\mathbb{X}} x^{\delta} \pi(\mathrm{d}x) < +\infty \tag{3.4}$$

as soon as positive γ, δ satisfy (3.1). On the other hand, the probability density $\mathfrak{p}$ of the invariant distribution π is proportional to $\sigma^{-2}\mathfrak{s}^{-1}$ (e.g., see [5]), and straightforward calculation shows that (3.4) holds true if, and only if,

$$\gamma < \frac{\alpha}{2}, \qquad \delta < \frac{\beta}{2}. \tag{3.5}$$

Clearly, the first bound in (3.5) is weaker than the one in (3.1). Such a discrepancy indicates that, in the current setting, the Lyapunov-type condition (3.3) is not precise, in a sense. This observation motivates the following extension of the above results. Define the family of *Cesàro means* of finite-dimensional distributions of X by

$$\mu^t_{t_1,\dots,t_m} = \frac{1}{t}\int_0^t \mu_{t_1+s,\dots,t_m+s}\,\mathrm{d}s, \qquad t>0, 0\le t_1<\dots<t_m, m\ge 1. \tag{3.6}$$

Proposition 3.2. 1. (*Modified Lyapunov-type condition*). *Let ϕ have the form* (2.5) *with positive γ, δ satisfying* (3.5). *Then there exists a nonnegative function $\psi \in \mathrm{Dom}(\mathcal{A})$, satisfying* (3.3) *and such that*

$$\mathcal{A}\psi \le -c'\phi^{1+\varepsilon} + C' \tag{3.7}$$

with some positive constants c', C', ε.

2. (*Moment bounds for Cesàro means*). *In the conditions and notation of statement* 1, *let c, C be the constants from the relation* (3.3) *for the function ψ. Then, for arbitrary $m\ge 1, 0\le t_1 < \dots < t_m$,*

$$\int_{\mathbb{X}^m} \big(\phi(x_1)+\dots+\phi(x_m)\big)^{1+\varepsilon}\mu^t_{t_1,\dots,t_m}(\mathrm{d}x) \le m^\varepsilon\left(\frac{C'}{c'}+\frac{C}{cc't}+\frac{1}{c't}\right)\int_{\mathbb{X}}\psi\,\mathrm{d}\mu. \tag{3.8}$$

Remark 3.1. Let $\mu=\delta_x$, then (3.8) with $m=1$ and $t_1=0$ yields

$$\sup_{t\ge 1}\frac{1}{t}\int_0^t\int_{\mathbb{X}}\phi\,\mathrm{d}\mu_s\,\mathrm{d}s<\infty.$$

On the other hand, by Theorem 3.1 below we have

$$\frac{1}{t}\int_0^t \mu_s\,\mathrm{d}s \Rightarrow \pi, \qquad t\to\infty.$$

These two observations, combined with the proper version of the Fatoux lemma (e.g., [8], Theorem 5.3) provide that ϕ is integrable w.r.t π. This means that the moment bound (3.8) yields (3.4) under (3.5), and hence resolves the discrepancy discussed above.

3.2. Coupling, ergodicity, and β-mixing

This section collects the results about the ergodic behavior of the Fisher–Snedecor diffusion. For our further needs, it will be convenient to introduce explicitly and discuss separately the notion of an *exponential ϕ-coupling.*

By the common terminology, a *coupling* for a pair of processes U, V is any two-component process $Z=(Z^1,Z^2)$ such that Z^1 has the same distribution with U and Z^2 has the same distribution with V. Following this terminology, for a Markov process X and every $\mu,\nu\in\mathcal{P}$, we consider two versions X^μ, X^ν of the process X with the initial distributions equal to μ and ν, respectively, and call (μ,ν)-*coupling for the process* X any two-component process $Z=(Z^1,Z^2)$ which is a coupling for X^μ, X^ν.

Definition 3.1. *The Markov process X* admits an exponential ϕ-coupling *if there exists an invariant measure π for this process and positive constants C, c such that, for every $\mu \in \mathcal{P}$, there exists a (μ, π)-coupling $Z = (Z^1, Z^2)$ with*

$$E\big[\phi\big(Z_t^1\big) + \phi\big(Z_t^2\big)\big]\mathbf{1}_{Z_t^1 \neq Z_t^2} \leq C\mathrm{e}^{-ct} \int_{\mathbb{X}} \phi \,\mathrm{d}\mu, \qquad t \geq 0. \tag{3.9}$$

The coupling construction is a traditional tool for proving the ergodicity. In [19], it was proposed to introduce a separate notion of an exponential ϕ-coupling, and it was demonstrated that such a notion is a convenient tool for studying convergence rates of L_p-semigroups, generated by a Markov process, and spectral properties of respective generators. In Section 5.5 below, we will see that this notion is also efficient for proving LLN and CLT. With this application in mind, we have changed slightly Definition 3.1, if to compare it with the one given in [19]: here, we consider all probability measures $\mu \in \mathcal{P}$ as possible initial distributions, while in [19] only measures of the form $\mu = \delta_x, x \in \mathbb{X}$ are considered.

Theorem 3.1. *Let ϕ be defined by* (2.5) *with γ, δ satisfying* (3.1). *Then the following statements hold.*

1. *The Fisher–Snedecor diffusion admits an exponential ϕ-coupling.*
2. *Finite-dimensional distributions of the Fisher–Snedecor diffusion admit the following convergence rate in the weighted total variation norm with the weight ϕ: for any $m \geq 1, 0 \leq t_1 < \cdots < t_m$,*

$$\|\mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m}\|_{\phi,\mathrm{var}} \leq mC\mathrm{e}^{-ct} \int_{\mathbb{X}} \phi \,\mathrm{d}\mu, \qquad \mu \in \mathcal{P}, t \geq 0. \tag{3.10}$$

Here the constants C, c are the same as in the bound (3.9) *in the definition of an exponential ϕ-coupling.*

3. *The Fisher–Snedecor diffusion admits the following bound for the β-mixing coefficient*:

$$\beta^{\mu}(t) \leq C'\mathrm{e}^{-ct} \int_{\mathbb{X}} \phi \,\mathrm{d}\mu, \qquad \mu \in \mathcal{P}, t \geq 0. \tag{3.11}$$

Here the constant c is the same as in the bound (3.9), *and C' a positive constant, which can be given explicitly* (*see* (5.15) *below*).

From (3.11) and Corollary 3.1, we get the following bounds for state-dependent and stationary β-mixing coefficients:

$$\beta_x(t) \leq C'\mathrm{e}^{-ct}\phi(x), \qquad x \in \mathbb{X}, t \geq 0,$$
$$\beta(t) \leq C''\mathrm{e}^{-ct}, \qquad t \geq 0, C'' := C' \int_{\mathbb{X}} \phi \,\mathrm{d}\pi < +\infty.$$

Note that the general theory for (possibly nonsymmetric and nonstationary) Markov processes provides convergence rates like (3.10), for example, [10], and bounds for β-mixing coefficients like (3.11), for example, [28], under a proper combination of "recurrence" and "local irreducibility" conditions. In our context, these conditions are provided by Proposition 3.1.

Apart with the convergence rate (3.10), we give the following more specific bound for continuous-time averages of the family $\{\mu_{t_1,\dots,t_m}\}$.

Theorem 3.2. *Let ϕ be defined by (2.5) with γ, δ satisfying (3.5), and ψ be the function from Proposition 3.2.*

Then for every $m \ge 1$ there exists a constant C_m such that

$$\left\| \int_0^T (\mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m})\,\mathrm{d}t \right\|_{\phi,\mathrm{var}} \le C_m \int_{\mathbb{X}} \psi\,\mathrm{d}\mu, \qquad \mu \in \mathcal{P}, T \ge 0. \tag{3.12}$$

Remark 3.2. Clearly, (3.10) provides a bound, similar to (3.12), with ϕ instead of ψ in the right-hand side. This bound is weaker than (3.12) because $\psi(x) = \mathrm{o}(\phi(x))$ as $x \to 0$ or $x \to \infty$. In addition, Theorem 3.2 requires (3.5), which is weaker than respective assumption (3.1) in Theorem 3.1. In this sense, for continuous-time averages of the family $\{\mu_{t_1,\dots,t_m}\}$ Theorem 3.2 provides a substantially more precise information than Theorem 3.1 does.

3.3. The law of large numbers and the central limit theorem

In this section, we formulate LLN and CLT for additive functionals of the Fisher–Snedecor diffusion X. Below, X_t^{st}, $t \in (-\infty, \infty)$ denotes the stationary version of X; that is, the strictly stationary process such that for every $m \ge 1$ and $t_1 < \cdots < t_m$ the joint distribution of $X_{t_1}^{st}, \dots, X_{t_m}^{st}$ equals $\pi_{0,t_2-t_1,\dots,t_m-t_1}$ (heuristically, X^{st} is "a solution to (1.1), which is defined on the whole time axis and starts at $-\infty$ from the invariant distribution π").

We consider separately the discrete-time and the continuous-time cases.

Theorem 3.3 (Discrete-time case). *Let, for some $r, k \ge 1$, a vector-valued function*

$$f = (f_1, \dots, f_k) : \mathbb{X}^r \to \mathbb{R}^k$$

be such that for any $i = 1, \dots, k$ for some γ_i, δ_i satisfying (3.1)

$$\left|f_i(x)\right| \le C \sum_{j=1}^r \left(x_j^{-\gamma_i} + x_j^{\delta_i}\right), \qquad x = (x_1, \dots, x_r) \tag{3.13}$$

with some constant C.

Then the following statements hold true.

1. (*LLN*). *For arbitrary initial distribution μ of X and arbitrary $t_1, \dots, t_r \ge 0$,*

$$\frac{1}{n} \sum_{l=1}^n f(X_{t_1+l}, \dots, X_{t_r+l}) \to a_f \tag{3.14}$$

in probability, where the asymptotic mean vector a_f *equals*

$$a_f = Ef\big(X^{st}_{t_1}, \dots, X^{st}_{t_r}\big).$$

If, in addition, the initial distribution is such that for some positive ε

$$\int_{\mathbb{X}} \big(x^{-\gamma_i-\varepsilon} + x^{\delta_i+\varepsilon}\big)\mu(\mathrm{d}x) < \infty, \qquad i = 1, \dots, k, \tag{3.15}$$

then (3.14) *holds true in the mean sense.*

2. (*CLT*). *Assume in addition that there exists* $\varepsilon > 0$ *such that*

$$E\big\| f\big(X^{st}_{t_1}, \dots, X^{st}_{t_r}\big)\big\|^{2+\varepsilon} < \infty. \tag{3.16}$$

Then

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{n} \big(f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f\big) \Rightarrow \mathcal{N}\big(0, \Sigma^d_f\big), \tag{3.17}$$

where the components of the asymptotic covariance matrix Σ^d_f *equal*

$$\big(\Sigma^d_f\big)_{i,j} = \sum_{l=-\infty}^{\infty} \mathrm{Cov}\big(f_i\big(X^{st}_{t_1+l}, \dots, X^{st}_{t_r+l}\big), f_j\big(X^{st}_{t_1}, \dots, X^{st}_{t_r}\big)\big), \qquad i, j = 1, \dots, k.$$

Theorem 3.4 (Continuous-time case). *Let the components of a vector-valued function* $f : \mathbb{X}^r \to \mathbb{R}^k$ *satisfy* (3.13) *with* γ_i, δ_i *satisfying* (3.5) *for every* $i = 1, \dots, k$.

Then the following statements hold true.

1. (*LLN*). *For arbitrary initial distribution* μ *of* X,

$$\frac{1}{T} \int_0^T f(X_{t_1+t}, \dots, X_{t_r+t})\,\mathrm{d}t \to a_f \tag{3.18}$$

in probability. If, in addition, the initial distribution is such that for some positive ε

$$\int_{\mathbb{X}} \big(x^{-(\gamma_i-1)\vee 0-\varepsilon} + x^{\delta_i+\varepsilon}\big)\mu(\mathrm{d}x) < \infty, \qquad i = 1, \dots, k, \tag{3.19}$$

then (3.18) *holds true in the mean sense.*

2. (*CLT*). *Assume in addition that*

$$\gamma_i < \frac{\alpha}{4} + \frac{1}{2}, \qquad \delta_i < \frac{\beta}{4}, \qquad i = 1, \dots, k. \tag{3.20}$$

Then, for arbitrary initial distribution μ *of* X,

$$\frac{1}{\sqrt{T}} \int_0^T \big(f(X_{t_1+t}, \dots, X_{t_r+t}) - a_f\big)\,\mathrm{d}t \Rightarrow \mathcal{N}\big(0, \Sigma^c_f\big), \tag{3.21}$$

where the components of the asymptotic covariance matrix Σ_f^c *equal*

$$(\Sigma_f^c)_{i,j} = \int_{-\infty}^{\infty} \mathrm{Cov}\big(f_i\big(X_{t_1+t}^{st}, \dots, X_{t_r+t}^{st}\big), f_j\big(X_{t_1}^{st}, \dots, X_{t_r}^{st}\big)\big)\,\mathrm{d}t, \qquad i, j = 1, \dots, k. \quad (3.22)$$

For the limit theorems above, respective functional versions are available, as well. In order to keep the exposition reasonably short, we formulate here only one functional limit theorem of such a kind, which corresponds to the CLT (3.21).

Theorem 3.5. *Let the components of a vector-valued function* $f : \mathbb{X}^r \to \mathbb{R}^k$ *satisfy* (3.13) *with*

$$\gamma_i < \frac{\alpha}{4}, \qquad \delta_i < \frac{\beta}{4}, \qquad i = 1, \dots, k. \quad (3.23)$$

Then

$$Y_T(\cdot) \equiv \frac{1}{\sqrt{T}} \int_0^T \big(f(X_{t_1+t}, \dots, X_{t_r+t}) - a_f\big)\,\mathrm{d}t \Rightarrow B, \qquad T \to \infty \quad (3.24)$$

weakly in $C([0, 1])$, *where* B *is the Brownian motion in* $\mathbb{R}^k$ *with the covariance matrix of* $B(1)$ *equal to* Σ_f^c.

4. Examples and statistical applications

4.1. Examples

In this section, we illustrate the above limit theorems and use them to derive the asymptotic properties of *empirical mixed moments*

$$\overline{m}_{\upsilon,\chi,c}(t) = \frac{1}{T} \int_0^T X_s^{\upsilon} X_{t+s}^{\chi}\,\mathrm{d}s, \qquad \overline{m}_{\upsilon,\chi,d}(t) = \frac{1}{n} \sum_{l=1}^{n} X_l^{\upsilon} X_{t+l}^{\chi}, \qquad t > 0$$

both in the continuous-time and in the discrete-time settings. Below we use statistical terminology because such functionals are particularly important for the statistic inference. For instance, usual *empirical moments*

$$\overline{m}_{\upsilon,c} = \frac{1}{T} \int_0^T X_s^{\upsilon}\,\mathrm{d}s, \qquad \overline{m}_{\upsilon,d} = \frac{1}{n} \sum_{l=1}^{n} X_l^{\upsilon} \quad (4.1)$$

equal the empirical mixed moments with $\chi = 0$, and *empirical covariances*

$$\begin{aligned} \overline{R}_c(t) &= \frac{1}{T} \int_0^T X_s X_{t+s}\,\mathrm{d}s - \left(\frac{1}{T} \int_0^T X_s\,\mathrm{d}s\right)^2, \\ \overline{R}_d(t) &= \frac{1}{n} \sum_{l=1}^{n} X_l X_{t+l} - \left(\frac{1}{n} \sum_{l=1}^{n} X_l\right)^2, \end{aligned} \quad (4.2)$$

can be written as

$$\overline{R}_c(t) = \overline{m}_{1,1,c}(t) - (\overline{m}_{1,c})^2, \qquad \overline{R}_d(t) = \overline{m}_{1,1,d}(t) - (\overline{m}_{1,d})^2. \tag{4.3}$$

Denote $\upsilon_- = -(\upsilon \wedge 0)$, $\upsilon_+ = \upsilon \vee 0$.

Example 4.1 (Discrete-time case). Let there exist $p, q > 1$ with $1/p + 1/q = 1$ such that

$$\{p\upsilon, q\chi\} \subset \left(-\frac{\alpha}{2} + 1, \frac{\beta}{2}\right). \tag{4.4}$$

Then for arbitrary initial distribution μ of X the discrete-time empirical mixed moment $\overline{m}_{\upsilon,\chi,d}(t)$ is a P-consistent estimator of the parameter

$$m_{\upsilon,\chi}(t) = E\left(X_0^{st}\right)^{\upsilon}\left(X_t^{st}\right)^{\chi}.$$

If, in addition, the initial distribution μ satisfies

$$\int_0^1 x^{-(p\upsilon_-)\vee(q\chi_-)-\varepsilon}\mu(\mathrm{d}x) + \int_1^\infty x^{(p\upsilon_+)\vee(q\chi_+)+\varepsilon}\mu(\mathrm{d}x) < \infty$$

for some $\varepsilon > 0$, then $\overline{m}_{\upsilon,\chi,d}(t)$ is an asymptotically unbiased estimator of $m_{\upsilon,\chi}(t)$.

Under the assumption

$$\{p\upsilon, q\chi\} \subset \left(-\left(\frac{\alpha}{2} - 1\right) \wedge \left(\frac{\alpha}{4}\right), \frac{\beta}{4}\right) \tag{4.5}$$

for arbitrary initial distribution μ of X the discrete-time empirical mixed moment $\overline{m}_{\upsilon,\chi,d}(t)$ is an asymptotically normal estimator of $m_{\upsilon,\chi}(t)$; that is,

$$\sqrt{n}\left(\overline{m}_{\upsilon,\chi,d}(t) - m_{\upsilon,\chi}(t)\right) \Rightarrow \mathcal{N}\left(0, \sigma^2_{\upsilon,\chi,d}(t)\right), \qquad n \to \infty$$

with

$$\sigma^2_{\upsilon,\chi,d}(t) = \sum_{l=-\infty}^{\infty} \mathrm{Cov}\left(\left(X_l^{st}\right)^{\upsilon}\left(X_{t+l}^{st}\right)^{\chi}, \left(X_0^{st}\right)^{\upsilon}\left(X_t^{st}\right)^{\chi}\right).$$

These results follow immediately from Theorem 3.3 with $k = 1$, $r = 2$, and

$$f(x_1, x_2) = x_1^{\upsilon} x_2^{\chi}.$$

Indeed, by the Young inequality,

$$f(x_1, x_2) \le \frac{x_1^{p\upsilon}}{p} + \frac{x_2^{q\chi}}{q}.$$

Then (3.13) holds true with $\gamma = (p\upsilon_-) \vee (q\chi_-)$ and $\delta = (p\upsilon_+) \vee (q\chi_+)$. Respectively, (4.4) coincides with the assumption (3.1), imposed on γ, δ in Theorem 3.3. The additional integrability assumption (3.16) now is equivalent to the following: for some positive ε,

$$-2(p\upsilon_-) \vee (q\chi_-) - \varepsilon > -\frac{\alpha}{2}, \qquad 2(p\upsilon_+) \vee (q\chi_+) + \varepsilon < \frac{\beta}{2}.$$

Clearly, this means that $\{p\upsilon, q\chi\} \subset (-\alpha/4, \beta/4)$, which together with (4.4) gives (4.5).

Similarly, using Theorem 3.4 under the same choice of f, γ, δ we obtain the following.

Example 4.2 (Continuous-time case). Let there exist $p, q > 1$ with $1/p + 1/q = 1$ such that

$$\{p\upsilon, q\chi\} \subset \left(-\frac{\alpha}{2}, \frac{\beta}{2}\right). \tag{4.6}$$

Then for arbitrary initial distribution μ of X the continuous-time empirical mixed moment $\overline{m}_{\upsilon,\chi,c}(t)$ is a P-consistent estimator of the $m_{\upsilon,\chi}(t)$.

If, in addition, the initial distribution μ satisfies

$$\int_0^1 x^{-((p\upsilon_-)\vee(q\chi_-)-1)_+-\varepsilon}\mu(\mathrm{d}x) + \int_1^\infty x^{(p\upsilon_+)\vee(q\chi_+)+\varepsilon}\mu(\mathrm{d}x) < \infty$$

for some $\varepsilon > 0$, then $\overline{m}_{\upsilon,\chi,c}(t)$ is an asymptotically unbiased estimator of $m_{\upsilon,\chi}(t)$.

Under the assumption

$$\{p\upsilon, q\chi\} \subset \left(-\frac{\alpha}{4} - \frac{1}{2}, \frac{\beta}{4}\right) \tag{4.7}$$

for arbitrary initial distribution μ of X the continuous-time empirical mixed moment $\overline{m}_{\upsilon,\chi,c}(t)$ is an asymptotically normal estimator of $m_{\upsilon,\chi}(t)$; that is,

$$\sqrt{T}\big(\overline{m}_{\upsilon,\chi,c}(t) - m_{\upsilon,\chi}(t)\big) \Rightarrow \mathcal{N}\big(0, \sigma^2_{\upsilon,\chi,c}(t)\big), \qquad T \to \infty$$

with

$$\sigma^2_{\upsilon,\chi,c}(t) = \int_{-\infty}^{\infty} \mathrm{Cov}\big(\big(X_s^{st}\big)^{\upsilon}\big(X_{t+s}^{st}\big)^{\chi}, \big(X_0^{st}\big)^{\upsilon}\big(X_t^{st}\big)^{\chi}\big)\,\mathrm{d}s.$$

The following statements can be obtained easily either by taking in the above examples $\chi = 0$ and $p > 1$ close enough to 1, or by using Theorem 3.3 and Theorem 3.4 with $k = r = 1$, $f(x) = x^{\upsilon}$, and $\gamma = \upsilon_-$, $\delta = \upsilon_+$.

Example 4.3 (Empirical moments). The discrete-time empirical moment $\overline{m}_{\upsilon,d}$, considered as an estimator of the parameter

$$m_{\upsilon} = E\big(X_0^{st}\big)^{\upsilon} = \int_{\mathbb{X}} x^{\upsilon}\pi(\mathrm{d}x),$$

has the following properties:

(i) if

$$\upsilon \in \left(-\frac{\alpha}{2}+1, \frac{\beta}{2}\right), \tag{4.8}$$

then $\overline{m}_{\upsilon,d}$ is P-consistent;

(ii) if, in addition, the initial distribution μ satisfies

$$\int_0^1 x^{-\upsilon_- -\varepsilon}\mu(\mathrm{d}x) + \int_1^\infty x^{\upsilon_+ +\varepsilon}\mu(\mathrm{d}x) < \infty$$

for some $\varepsilon > 0$, then $\overline{m}_{\upsilon,d}$ is asymptotically unbiased;

(iii) if

$$\upsilon \in \left(-\left(\frac{\alpha}{2}-1\right)\wedge\left(\frac{\alpha}{4}\right), \frac{\beta}{4}\right), \tag{4.9}$$

then $\overline{m}_{\upsilon,d}$ is asymptotically normal.

Similarly, the continuous-time empirical moment $\overline{m}_{\upsilon,c}$, considered as an estimator of the same parameter, satisfies the following:

(i) if

$$\upsilon \in \left(-\frac{\alpha}{2}, \frac{\beta}{2}\right), \tag{4.10}$$

then $\overline{m}_{\upsilon,c}$ is P-consistent;

(ii) if, in addition, the initial distribution μ satisfies

$$\int_0^1 x^{-(\upsilon_- -1)_+ -\varepsilon}\mu(\mathrm{d}x) + \int_1^\infty x^{\upsilon_+ +\varepsilon}\mu(\mathrm{d}x) < \infty$$

for some $\varepsilon > 0$, then $\overline{m}_{\upsilon,d}$ is asymptotically unbiased;

(iii) if

$$\upsilon \in \left(-\frac{\alpha}{4}-\frac{1}{2}, \frac{\beta}{4}\right), \tag{4.11}$$

then $\overline{m}_{\upsilon,c}$ is asymptotically normal.

Comparing (4.8) with (4.10) and (4.9) with (4.11), one can see clearly the difference between the conditions of Theorem 3.4 and the conditions of Theorem 3.3. The particularly interesting case here is

$$\upsilon \in \left(-\frac{\alpha}{4}-\frac{1}{2}, -\frac{\alpha}{4}\right].$$

In this case, the function $f(x) = x^\upsilon$ satisfies conditions of Theorem 3.4 with $r = k = 1$, while the additional integrability assumption (3.16) in Theorem 3.3 fails because f is not square integrable w.r.t. π. This observation reveals a new effect, already mentioned in the Introduction, which seemingly has not been observed in the literature before: a functional f, which is not square integrable w.r.t. the invariant distribution, still may lead to the CLT in its continuous-time form (3.21).

Example 4.4 (Empirical covariances). Both the discrete-time empirical covariance $\overline{R}_d(t)$ and the continuous-time empirical covariance $\overline{R}_c(t)$, considered as estimators of the parameter

$$R(t) = \mathrm{Cov}\big(X_t^{st}, X_0^{st}\big),$$

have the following properties:

(i) if $\beta > 4$ then $\overline{R}_d(t)$ and $\overline{R}_c(t)$ are P-consistent;
(ii) if, in addition, the initial distribution μ satisfies

$$\int_1^\infty x^{2+\varepsilon} \mu(\mathrm{d}x) < \infty$$

for some $\varepsilon > 0$, then $\overline{R}_d(t)$ and $\overline{R}_c(t)$ are asymptotically unbiased;
(iii) if $\beta > 8$ then $\overline{R}_d(t)$ and $\overline{R}_c(t)$ are asymptotically normal.

These results follow from the representation (4.3) and Theorems 3.3, 3.4 with $k = r = 2$, $f = (f_1, f_2)$,

$$f_1(x_1, x_2) = x_1, \qquad f_2(x_1, x_2) = x_1 x_2.$$

Similarly to Example 4.1 and Example 4.2 (in this particular case one should take $p = q = 2$), one can verify that both $(\overline{m}_{1,d}, \overline{m}_{1,1,d}(t))$ and $(\overline{m}_{1,c}, \overline{m}_{1,1,c}(t))$ are P-consistent if $\beta > 4$ and asymptotically normal if $\beta > 8$, when considered as estimators of the vector parameter $(m_1, m_{1,1}(t))$. Then properties (i) and (iii) follow by the continuity mapping theorem and the functional delta method (see [25], Theorem 3.3.A). Under the additional integrability assumption on μ both $\overline{m}_{1,1,d}(t)$ and $\overline{m}_{1,1,c}(t)$ are asymptotically unbiased. On the other hand, under the same assumption both $(\overline{m}_{1,d})^2$ and $(\overline{m}_{1,c})^2$ are uniformly integrable w.r.t. P_μ; this follows from the Hölder inequality and Corollary 3.1:

$$E_\mu(\overline{m}_{1,d})^{2+\varepsilon} = E_\mu\left(\frac{1}{n}\sum_{l=1}^n X_l\right)^{2+\varepsilon} \le \frac{1}{n}\sum_{l=1}^n E_\mu X_l^{2+\varepsilon} \le C,$$

the inequality for the continuous-time case is similar and omitted. This implies that $(\overline{m}_{1,d})^2$ and $(\overline{m}_{1,c})^2$ are asymptotically unbiased, which completes the proof of the property (ii).

Similarly, the properties of the empirical estimates of the vector-valued parameters of the type $(m_{\upsilon_1}, \dots, m_{\upsilon_k})$ or $(m_{\upsilon_1}, \dots, m_{\upsilon_k}, R(t))$ can be derived. For such parameters, the component-wise properties of P-consistency and asymptotic unbiasedness are already studied in the previous examples. Hence, in the following example, we address the asymptotic normality only.

Example 4.5 (Multivariate estimators). *I. (Discrete-time case).* Let

$$\upsilon_1, \dots, \upsilon_k \in \left(-\left(\frac{\alpha}{2}-1\right) \wedge \left(\frac{\alpha}{4}\right), \frac{\beta}{4}\right).$$

Then, for arbitrary initial distribution μ of X, the estimator $\overline{m}_{\upsilon_1,\dots,\upsilon_k,d} = (\overline{m}_{\upsilon_1,d}, \dots, \overline{m}_{\upsilon_k,d})$ of the vector-valued parameter $m_{\upsilon_1,\dots,\upsilon_k} = (m_{\upsilon_1}, \dots, m_{\upsilon_k})$ is asymptotically normal; that is,

$$\sqrt{n}(\overline{m}_{\upsilon_1,\dots,\upsilon_k,d} - m_{\upsilon_1,\dots,\upsilon_k}) \Rightarrow \mathcal{N}(0, \Sigma), \qquad n \to \infty$$

with some positive semi-definite matrix Σ.

If, in addition, $\beta > 8$, then $(\overline{m}_{\upsilon_1,d}, \dots, \overline{m}_{\upsilon_k,d}, \overline{R}_d(t))$ is an asymptotically normal estimator of $(m_{\upsilon_1}, \dots, m_{\upsilon_k}, R(t))$ for any $t > 0$.

II. (Continuous-time case). Let

$$\upsilon_1, \dots, \upsilon_k \in \left(-\frac{\alpha}{4} - \frac{1}{2}, \frac{\beta}{4}\right).$$

Then, for arbitrary initial distribution μ of X, the estimator $\overline{m}_{\upsilon_1,\dots,\upsilon_k,c} = (\overline{m}_{\upsilon_1,c}, \dots, \overline{m}_{\upsilon_k,c})$ of the vector-valued parameter $m_{\upsilon_1,\dots,\upsilon_k} = (m_{\upsilon_1}, \dots, m_{\upsilon_k})$ is asymptotically normal; that is,

$$\sqrt{T}(\overline{m}_{\upsilon_1,\dots,\upsilon_k,c} - m_{\upsilon_1,\dots,\upsilon_k}) \Rightarrow \mathcal{N}(0, \Sigma), \qquad T \to \infty$$

with some positive semi-definite matrix Σ.

If, in addition, $\beta > 8$, then $(\overline{m}_{\upsilon_1,c}, \dots, \overline{m}_{\upsilon_k,c}, \overline{R}_c(t))$ is an asymptotically normal estimator of $(m_{\upsilon_1}, \dots, m_{\upsilon_k}, R(t))$ for any $t > 0$.

4.2. Parameter estimation for the Fisher–Snedecor diffusion

In this section, we give an application of the above results to the parameter estimation of the Fisher–Snedecor diffusion. We use the method of moments and the asymptotic properties of the empirical moments (4.1) and the empirical covariances (4.2), exposed in Examples 4.3–4.5, in order to provide the statistical analysis of the *autocorrelation parameter* θ and the *shape parameters* α, β, and κ of the Fisher–Snedecor diffusion. We put

$$\begin{aligned}
\widehat{\alpha}_c &= \frac{2(\overline{m}_{-1,c}\overline{m}_{1,c}\overline{m}_{2,c} - \overline{m}_{1,c}^2)}{\overline{m}_{-1,c}\overline{m}_{1,c}\overline{m}_{2,c} - 2\overline{m}_{2,c} + \overline{m}_{1,c}^2}, \qquad &\widehat{\beta}_c &= \frac{4\overline{m}_{-1,c}(\overline{m}_{2,c} - \overline{m}_{1,c}^2)}{\overline{m}_{-1,c}\overline{m}_{2,c} - 2\overline{m}_{-1,c}\overline{m}_{1,c}^2 + \overline{m}_{1,c}}, \\
\widehat{\kappa}_c &= \frac{4\overline{m}_{-1,c}\overline{m}_{1,c}(\overline{m}_{2,c} - \overline{m}_{1,c}^2)}{\overline{m}_{-1,c}\overline{m}_{2,c} - 2\overline{m}_{-1,c}\overline{m}_{1,c}^2 + \overline{m}_{1,c}}, \qquad &\widehat{\theta}_c &= -\frac{1}{t}\log\left(\frac{\overline{R}_c(t)}{\overline{m}_{2,c} - \overline{m}_{1,c}^2}\right)
\end{aligned} \tag{4.12}$$

for a given $t > 0$, and define $\widehat{\alpha}_d, \widehat{\beta}_d, \widehat{\kappa}_d, \widehat{\theta}_d$ by similar relations with $\overline{m}_{i,d}$, $i = -1, 1, 2$, and $\overline{R}_d(t)$ instead of $\overline{m}_{i,c}$, $i = -1, 1, 2$, and $\overline{R}_c(t)$, respectively.

Theorem 4.1. *Let* $\beta > 8$. *Then, for arbitrary initial distribution of the Fisher–Snedecor diffusion*, $(\widehat{\alpha}_c, \widehat{\beta}_c, \widehat{\kappa}_c, \widehat{\theta}_c)$ *is a* P*-consistent and asymptotically normal estimator of the parameter* $(\alpha, \beta, \kappa, \theta)$; *that is*,

$$\sqrt{T}(\widehat{\alpha}_c - \alpha, \widehat{\beta}_c - \beta, \widehat{\kappa}_c - \kappa, \widehat{\theta}_c - \theta) \Rightarrow \mathcal{N}\big(0, \Sigma_c(\alpha, \beta, \kappa, \theta)\big), \qquad T \to \infty.$$

For the estimator $(\widehat{\alpha}_d, \widehat{\beta}_d, \widehat{\kappa}_d, \widehat{\theta}_d)$, *the similar statement holds true under the additional assumption* $\alpha > 4$. *In that case*,

$$\sqrt{n}(\widehat{\alpha}_d - \alpha, \widehat{\beta}_d - \beta, \widehat{\kappa}_d - \kappa, \widehat{\theta}_d - \theta) \Rightarrow \mathcal{N}\big(0, \Sigma_d(\alpha, \beta, \kappa, \theta)\big), \qquad n \to \infty.$$

The matrices $\Sigma_c(\alpha, \beta, \kappa, \theta)$, $\Sigma_d(\alpha, \beta, \kappa, \theta)$ are completely identifiable. To keep the current paper reasonably short, we postpone their explicit calculation, together with a more detailed discussion of the statistical aspects, to the subsequent paper [20].

Remark 4.1. The estimators (4.12) can be simplified significantly if either exact values of some parameters α, β, κ are known, or these parameters possess some functional relation. Let, for instance, $\kappa = \beta/(\beta - 2)$; this particular case is of a separate interest because the invariant distribution π then coincides with the Fisher–Snedecor distribution $\mathcal{FS}(\alpha, \beta)$. In this case, one can replace in (4.12) the identities for $\widehat{\alpha}_c, \widehat{\beta}_c$ by either

$$\widehat{\alpha}_c = \frac{2\overline{m}_{1,c}^2}{\overline{m}_{2,c}(2 - \overline{m}_{1,c}) - \overline{m}_{1,c}^2}, \qquad \widehat{\beta}_c = \frac{2\overline{m}_{1,c}}{\overline{m}_{1,c} - 1} \tag{4.13}$$

or

$$\widehat{\alpha}_c = \frac{2\overline{m}_{-1,c}}{\overline{m}_{-1,c} - 1}, \qquad \widehat{\beta}_c = \frac{2\overline{m}_{1,c}}{\overline{m}_{1,c} - 1}. \tag{4.14}$$

For the estimator $(\widehat{\alpha}_c, \widehat{\beta}_c, \widehat{\theta}_c)$, defined in such a way, and its discrete-time analogue $(\widehat{\alpha}_d, \widehat{\beta}_d, \widehat{\theta}_d)$, the statements of Theorem 4.1 hold true; see more detailed discussion in [20].

5. Proofs

5.1. Proof of Proposition 3.1

Statement 1. Let the initial value $X_0 = x \in \mathbb{X}$ be fixed. Note that the process

$$H_t^{\phi,X} = \phi(X_t) - \int_0^t \mathcal{A}\phi(X_s)\, ds, \qquad t \in \mathbb{R}^+, \tag{5.1}$$

with $\mathcal{A}\phi$ defined by (3.2), is an $\mathbb{F}^X$-local martingale w.r.t. the measure P_x. The argument here is quite standard, we explain it briefly in order to keep the exposition self-sufficient. Introduce the sequence of $\mathbb{F}^X$-stopping times $T_n = \inf\{t\colon\ X_t \le 1/n\}, n \in \mathbb{N}$, and consider auxiliary functions $\phi_n \in C^2(\mathbb{R})$ such that $\phi_n = \phi$ on $[1/n, \infty)$. For any given $n \in \mathbb{N}$, by the Ito formula (e.g., [15], Chapter II, Theorem 5.1) we have that the process $H^{\phi_n,X}$, defined by the relation (5.1) with ϕ_n instead of ϕ, is an $\mathbb{F}^X$-local martingale. This means that, for any given $n \in \mathbb{N}$, there exists a sequence of $\mathbb{F}^X$-stopping times $T_{n,m}, m \in \mathbb{N}$ such that every process

$$t \mapsto H^{\phi_n,X}(t \wedge T_{n,m}), \qquad m \in \mathbb{N}$$

is an $\mathbb{F}^X$-martingale w.r.t. the measure P_x, and

$$T_{n,m} \to \infty, \qquad m \to \infty, P_x\text{-a.s.}$$

The last relation provides that for every $n \in \mathbb{N}$ there exists m_n such that

$$P_x(T_{n,m_n} \le n) < 2^{-n}.$$

Consequently, by the Borel–Cantelli lemma,

$$T_{n,m_n} \to \infty, \qquad n \to \infty, P_x\text{-a.s.}$$

On the other hand, since the point 0 is unattainable for X, we have $T_n \to \infty$ P_x-a.s. Consequently, for $S_n = T_n \wedge T_{n,m_n}, n \in \mathbb{N}$ we have

$$S_n \to \infty, \qquad n \to \infty, P_x\text{-a.s.}$$

By the Doob optional sampling theorem, the process

$$t \mapsto H^{\phi_n,X}(t \wedge S_n)$$

is an $\mathbb{F}^X$-martingale w.r.t. the measure P_x. On the other hand, the processes $H^{\phi_n,X}$ and $H^{\phi,X}$ coincide up to the time moment T_n because the values of ϕ_n and its derivatives on $[1/n, \infty)$ coincide with respective values of ϕ. Hence, the process

$$t \mapsto H^{\phi,X}(t \wedge S_n)$$

is an $\mathbb{F}^X$-martingale w.r.t. the measure P_x, which completes the proof of the fact that $H^{\phi,X}$ is a $\mathbb{F}^X$-local martingale.

Next, we show that the function $\mathcal{A}\phi$ defined by (3.2) satisfies (3.3) for properly chosen positive u, v, c, C. We have for x large enough:

$$\begin{aligned}\mathcal{A}\phi(x) &= -\theta\delta(x-\kappa)x^{\delta-1} + \theta\delta(\delta-1)x\left(\frac{x}{\beta/2-1} + \frac{\kappa}{\alpha/2}\right)x^{\delta-2} \\ &= -\theta\delta\phi(x)\left[\left(1-\frac{\kappa}{x}\right) - (\delta-1)\left(\frac{1}{\beta/2-1} + \frac{\kappa}{x\alpha/2}\right)\right].\end{aligned} \tag{5.2}$$

The term $[\cdots]$ tends to $1 - \frac{\delta-1}{\beta/2-1}$ as $x \to \infty$, and it was assumed that $\delta < \beta/2$. Hence (3.3) holds true for any $x > v$ assuming $v > 0$ is chosen large enough and $c > 0$ is chosen small enough.

We have for x small enough:

$$\begin{aligned}\mathcal{A}\phi(x) &= \theta\gamma(x-\kappa)x^{-\gamma-1} + \theta\gamma(\gamma+1)x\left(\frac{x}{\beta/2-1} + \frac{\kappa}{\alpha/2}\right)x^{-\gamma-2} \\ &= -\theta\gamma\phi(x)\left\{\left(\frac{\kappa}{x}-1\right) - (\gamma+1)\left(\frac{1}{\beta/2-1} + \frac{\kappa}{x\alpha/2}\right)\right\}.\end{aligned} \tag{5.3}$$

The term $\{\cdots\}$ is equivalent to

$$\frac{\kappa}{x}\left(1 - \frac{\gamma+1}{\alpha/2}\right)$$

as $x \to 0+$, and it tends to $+\infty$ because it was assumed that $\gamma + 1 < \alpha/2$. Hence, (3.3) holds true for any $x \in (0, u)$ assuming $u, c > 0$ are chosen small enough. Finally, for given u, v, c (3.3) holds true for $x \in [u, v]$ under appropriate choice of (large) C.

Finally, we show that the process (5.1) is an $\mathbb{F}^X$-martingale. This proof is quite standard, again. For any $n \in \mathbb{N}$, we have

$$E_x H^{\phi,X}(t \wedge S_n) = \phi(x), \qquad t \geq 0; \tag{5.4}$$

here $S_n, n \in \mathbb{N}$ is the sequence of stopping times constructed in the first part of the proof. Recall that it is supposed that $\phi(x) \geq 1$, and therefore $\phi(x)$ is positive. This, together with (3.3), provides that $[\mathcal{A}\phi]_+(x) = (\mathcal{A}\phi(x)) \vee 0$ is a bounded function. Then

$$E_x\phi(X_{t\wedge S_n}) = \phi(x) + E_x\int_0^{t\wedge S_n} \mathcal{A}\phi(X_s)\,\mathrm{d}s \leq \phi(x) + t\sup_{x'}[\mathcal{A}\phi]_+(x'), \qquad t \geq 0, n \in \mathbb{N}.$$

Consequently, we have from (5.4) that for any $T \geq 0$

$$\sup_{t\leq T}\sup_{n\in\mathbb{N}} E_x\phi(X_{t\wedge S_n}) < \infty. \tag{5.5}$$

Denote $[\mathcal{A}\phi]_-(x) = (-\mathcal{A}\phi(x)) \vee 0$; then (5.4) can be written as

$$E_x\int_0^{t\wedge S_n}[\mathcal{A}\phi]_-(X_s)\,\mathrm{d}s = \phi(x) - E_x\phi(X_{t\wedge S_n}) + E_x\int_0^{t\wedge S_n}[\mathcal{A}\phi]_+(X_s)\,\mathrm{d}s.$$

Combined with (5.5) and the fact that $[\mathcal{A}\phi]_+$ is bounded, this yields

$$E_x\int_0^t [\mathcal{A}\phi]_-(X_s)\,\mathrm{d}s < \infty.$$

In particular, the Lebesgue dominated convergence theorem and boundedness of $[\mathcal{A}\phi]_+$ provide that the sequence

$$\int_0^{t\wedge S_n}\mathcal{A}\phi(X_s)\,\mathrm{d}s, \qquad n \in \mathbb{N}$$

is uniformly integrable w.r.t. P_x.

Note that the above argument can be repeated with the function ϕ replaced by the function $\tilde{\phi} = \phi^{\upsilon}$, where $\upsilon > 1$ is chosen in such a way that

$$\upsilon\gamma < \frac{\alpha}{2} - 1, \qquad \upsilon\delta < \frac{\beta}{2}.$$

Then, similarly to (5.5), we will have

$$\sup_{t \le T} \sup_{n \in \mathbb{N}} E_x \big(\phi(X_{t\wedge \tilde{S}_n})\big)^{\upsilon} < \infty \tag{5.6}$$

with some sequence of stopping times $\tilde{S}_n$ such that $\tilde{S}_n \to \infty$ P_x-a.s. This means that the sequence $\phi(X_{t\wedge S_n \wedge \tilde{S}_n}), n \in \mathbb{N}$ of the processes on $[0, T]$ is uniformly integrable w.r.t. P_x, and hence the sequence $H^{\phi,X}(t \wedge S_n \wedge \tilde{S}_n), n \in \mathbb{N}$ is uniformly integrable, as well. Then $H^{\phi,X}$ is a martingale as an a.s. limit of a uniformly integrable sequence of martingales.

Statement 2. Take a segment $[w, z] \in \mathbb{X}$ such that $[u, v] \subset (w, z)$, and consider the process $X^{[w,z]}$ obtained from X by killing at the exit from (w, z). Clearly, for any x inside (w, z) the transition probability $P_t(x, \mathrm{d}y)$ is minorized by the transition probability $P_t^{[w,z]}(x, \mathrm{d}y)$ of the process $X^{[w,z]}$. The latter function is the fundamental solution to the Cauchy problem for the linear 2nd order parabolic equation

$$\partial_t u(x, y) = \mathcal{L}u(t, x), \qquad x \in (w, z), u(t, w) = u(t, z) = 0, t > 0,$$

where

$$\mathcal{L} = a(x)\partial_x + \tfrac{1}{2}\sigma^2(x)\partial^2_{xx}.$$

Because the coefficients a, σ are smooth in $[w, z]$ and σ is positive, the general analytic results from the theory of linear 2-nd order parabolic equations (e.g., [21], Chapter IV, Sections 11–14) yield representation

$$P_t^{[w,z]}(x, \mathrm{d}y) = Z_t(x, y)\,\mathrm{d}y$$

with a continuous function $Z : (0, +\infty) \times (w, z) \times (w, z) \to [0, \infty)$. Because Z is continuous and is not an identical zero, there exist $t_1 > 0, x_1 \in (w, z), y_1 \in (w, z)$, and $\varepsilon > 0$ such that

$$c_1 := \inf_{|x - x_1| \le \varepsilon, |y - y_1| \le \varepsilon} Z_{t_1}(x, y) > 0.$$

In other words, we have constructed $t_1 > 0$ and segments $[u', v'] = [y_1 - \varepsilon, y_1 + \varepsilon]$ and $[u'', v''] = [x_1 - \varepsilon, x_1 + \varepsilon]$ such that

$$P_{t_1}(x, A) \ge P_{t_1}^{[w,z]}(x, A) \ge c_1 \int_A \mathrm{d}y \tag{5.7}$$

for any $x \in [u'', v'']$ and Borel measurable set $A \subset [u', v']$. Take $t_2 > 0$ and put $T = t_1 + t_2$. The Chapmen–Kolmogorov equation and (5.7) yields for every $x \in [u, v]$ and Borel measurable set

$A \subset [u', v']$

$$P_T(x, A) \geq \int_{[u'', v'']} P_{t_1}(x', A) P_{t_2}(x, \mathrm{d}x') \geq c_1 \inf_{x \in [u,v]} P_{t_2}\big(x, (u'', v'')\big) \int_A \mathrm{d}y.$$

The reason for us to replace in the last inequality the segment $[u'', v'']$ by the open interval (u'', v'') is that the indicator of this interval can be obtained as a limit of an increasing sequence of continuous functions $f_n : \mathbb{X} \to \mathbb{R}^+$, $n \geq 1$. The process X is a Feller one; this follows from the standard theorem on continuity of a solution to an SDE w.r.t. its initial value, for example, [13], Chapter II. Therefore, every function

$$x \mapsto \int_{\mathbb{X}} f_n(y) P_{t_2}(x, \mathrm{d}y)$$

is continuous, which implies that the function

$$x \mapsto P_{t_2}\big(x, (u'', v'')\big)$$

is lower semicontinuous as a point-wise limit of an increasing sequence of continuous functions. Then there exists $x_\diamond \in [u, v]$ such that

$$\inf_{x \in [u,v]} P_{t_2}\big(x, (u'', v'')\big) = P_{t_2}\big(x_\diamond, u'', v''\big).$$

On the other hand, for any $t > 0$, $x \in \mathbb{X}$ the support of the measure $P_t(x, \cdot)$ coincides with whole $\mathbb{X}$; because the diffusion coefficient is positive, this follows from the Stroock–Varadhan support theorem (e.g., [15], Chapter VI, Theorem 8.1). Hence $P_{t_2}(x_\diamond, (v'', v'')) > 0$, and the required statement holds true with

$$c_{u,v,u',v',T} = c_1 \inf_{x \in [u,v]} P_{t_2}\big(x, (u'', v'')\big) > 0.$$

5.2. Proof of Proposition 3.2

Statement 1. Take, analogously to (2.5), a function $\psi : \mathbb{X} \to [1, +\infty)$ of the form

$$\psi = \psi_\diamond + \psi_\blacklozenge,$$

where $\psi_\diamond, \psi_\blacklozenge \in C^2(0, \infty)$, $\psi_\diamond = 0$ on $[2, \infty)$, $\psi_\blacklozenge = 0$ on $(0, 1]$,

$$\psi_\diamond(x) = x^{-\gamma'} \quad \text{for } x \text{ small enough}, \qquad \psi_\blacklozenge(x) = x^{\delta'} \quad \text{for } x \text{ large enough},$$

with

$$\gamma' \in \left((\gamma - 1) \vee 0, \frac{\alpha}{2} - 1\right), \qquad \delta' \in \left(\delta, \frac{\beta}{2}\right).$$

Then, by the statement 1 of Proposition 3.1, $\psi \in \mathrm{Dom}(\mathcal{A})$ and ψ satisfies (3.3). By (5.2), one has

$$\mathcal{A}\psi(x) \sim -C_\infty x^{\delta'} = -C_\infty\big(\phi(x)\big)^{\delta'/\delta}, \qquad x \to \infty$$

with

$$C_\infty = \theta\delta'\left(1 - \frac{\delta' - 1}{\beta/2 - 1}\right) > 0.$$

By (5.3), one has

$$\mathcal{A}\psi(x) \sim -C_0 x^{-\gamma'-1} = -C_0\big(\phi(x)\big)^{(\gamma'+1)/\gamma}, \qquad x \to 0$$

with

$$C_0 = \theta\gamma'\kappa\left(1 - \frac{\gamma' + 1}{\alpha/2}\right) > 0.$$

Finally, for every segment $[u, v] \subset (0, \infty)$ and every $\varepsilon > 0$ one has

$$\sup_{x\in[u,v]} \phi(x) < \infty, \qquad \sup_{x\in[u,v]} \frac{|\mathcal{A}\psi(x)|}{\phi^{1+\varepsilon}(x)} < \infty,$$

because $\phi, \mathcal{A}\psi \in C(0, \infty)$ and $\phi \ge 1$. These observations provide (3.7) with small enough c', ε and large enough C'.

Statement 2. By the elementary inequality $(\sum_{k=1}^m a_k)^{1+\varepsilon} \le m^\varepsilon \sum_{k=1}^m a_k^{1+\varepsilon}$, we have

$$\int_{\mathbb{X}} \big(\phi(x_1) + \cdots + \phi(x_m)\big)^{1+\varepsilon} \mu^t_{t_1,\dots,t_m}(\mathrm{d}x) \le m^\varepsilon \sum_{k=1}^m \int_{\mathbb{X}} \phi^{1+\varepsilon}\,\mathrm{d}\mu^t_{t_k}. \tag{5.8}$$

By the definition of $\mathcal{A}$, we have for arbitrary $\mu \in \mathcal{P}$

$$E_\mu\psi(X_t) = E_\mu\psi(X_0) + E_\mu \int_0^t \mathcal{A}\psi(X_s)\,\mathrm{d}s.$$

Together with (3.7), this yields

$$\begin{aligned} \int_{\mathbb{X}} \phi^{1+\varepsilon}\,\mathrm{d}\mu^t &= \frac{1}{t}\int_0^t E_\mu\phi^{1+\varepsilon}(X_s)\,\mathrm{d}s \le \frac{1}{c't} E_\mu\left[\int_0^t C'\,\mathrm{d}s - \int_0^t \mathcal{A}\psi(X_s)\,\mathrm{d}s\right] \\ &= \frac{C'}{c'} + \frac{1}{c't}E_\mu\psi(X_0) - \frac{1}{c't}E_\mu\psi(X_t) \\ &\le \frac{C'}{c'} + \frac{1}{c't}E_\mu\psi(X_0) = \frac{C'}{c'} + \frac{1}{c't}\int_{\mathbb{X}}\psi\,\mathrm{d}\mu; \end{aligned} \tag{5.9}$$

in the second inequality, we have used that ψ is nonnegative. By Corollary 3.1 with ψ instead of ϕ, we have

$$\int_{\mathbb{X}} \psi \, d\mu_{t_k} \le \frac{C}{c} + \int_{\mathbb{X}} \psi \, d\mu \le \left(\frac{C}{c} + 1\right) \int_{\mathbb{X}} \psi \, d\mu, \qquad k = 1, \dots, m$$

because $\psi \ge 1$. Using (5.8) and (5.9) with $\mu_{t_k}, k = 1, \dots, m$ instead of μ, we obtain (3.8).

5.3. Proof of Theorem 3.1

Statement 1. In [19], Theorem 2.1, it is proved that a Markov process X admits an exponential ϕ-coupling under the following assumptions:

(i) $\phi \in \mathrm{Dom}(\mathcal{A})$ and (3.3) holds true;
(ii) every *level set* $\{\phi \le R\}$, $R \ge 1$ has a compact closure in $\mathbb{X}$;
(iii) for every compact $K \subset \mathbb{X}$ there exists $T > 0$ such that

$$\sup_{x, x' \in K} \| P_T(x, \cdot) - P_T(x', \cdot) \|_{\mathrm{var}} < 2, \tag{5.10}$$

where $\| \cdot \|_{\mathrm{var}}$ denotes the total variation norm.

In our setting, (i) and (iii) are provided by Proposition 3.1 (statements 1 and 2, resp.). Assumption (ii) holds true trivially because $\phi(x) \to +\infty$ when either $x \to 0$ or $x \to \infty$. Hence, the required statement follows by Theorem 2.1 in [19].

Remark 5.1. In [19], the notion of an exponential ϕ-coupling was introduced in a form, slightly weaker than the one from Definition 3.1; see the discussion after Definition 3.1. One can see easily that the proof of Theorem 2.1 in [19] can be extended straightforwardly to provide an exponential ϕ-coupling in the sense of Definition 3.1.

Statement 2. By statement 1, for a given $\mu \in \mathcal{P}$ there exists a (μ, π)-coupling which satisfies (3.9). From this fact, we will deduce (3.10). In a particular case $\phi \equiv 1, m = 1$ such an implication is well known, and the proof for general ϕ, m does not require any substantial changes when compared with the standard one. To keep the exposition self-sufficient, we explain the argument briefly. Denote $\varkappa_t = \mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m}$,

$$\nu_{i,t}(dy) = P\big((Z^i_{t_1+t}, \dots, Z^i_{t_m+t}) \in dy,$$
$$(Z^1_{t_1+t}, \dots, Z^1_{t_m+t}) \ne (Z^2_{t_1+t}, \dots, Z^2_{t_m+t})\big), \qquad i = 1, 2.$$

For arbitrary measurable function $f : \mathbb{X}^m \to [0, +\infty)$, one has

$$\begin{aligned} \int_{\mathbb{X}^m} f \, d\varkappa_t &= Ef(Z^1_{t_1+t}, \dots, Z^1_{t_m+t}) - Ef(Z^2_{t_1+t}, \dots, Z^2_{t_m+t}) \\ &= \int_{\mathbb{X}^m} f \, d\nu_{1,t} - \int_{\mathbb{X}^m} f \, d\nu_{2,t} \le \int_{\mathbb{X}^m} f \, d\nu_{1,t}. \end{aligned} \tag{5.11}$$

Denote by A_t^+ a set such that $\varkappa_t^+$ is supported by A_t^+ and $\varkappa_t^-(A_t^+) = 0$. By (5.11), we have for any measurable $A \subset A_t^+$:

$$\varkappa_t^+(A) = \varkappa_t(A) \le \nu_{1,t}(A).$$

Because $\varkappa_t^+$ is supported by A_t^+, this gives finally

$$\varkappa_t^+ \le \nu_{1,t}.$$

Similarly,

$$\varkappa_t^- \le \nu_{2,t}.$$

From these inequalities, we have

$$\begin{aligned}\|\varkappa_t\|_{\phi,\mathrm{var}} &\le \int_{\mathbb{X}^m} \big(\phi(x_1)+\cdots+\phi(x_m)\big)\nu_{1,t}(\mathrm{d}x) + \int_{\mathbb{X}^m} \big(\phi(x_1)+\cdots+\phi(x_m)\big)\nu_{1,t}(\mathrm{d}x)\\ &= E\left(\sum_{j=1}^m \big[\phi\big(Z^1_{t+t_j}\big)+\phi\big(Z^2_{t+t_j}\big)\big]\right)\mathbf{1}_{(Z^1_{t_1+t},\dots,Z^1_{t_m+t})\ne(Z^1_{t_1+t},\dots,Z^1_{t_m+t})}\\ &\le \sum_{j=1}^m E\big[\phi\big(Z^1_{t+t_j}\big)+\phi\big(Z^2_{t+t_j}\big)\big]\mathbf{1}_{Z^1_{t+t_j}\ne Z^2_{t+t_j}} \le mC\mathrm{e}^{-ct}\int_{\mathbb{X}}\phi\,\mathrm{d}\mu,\end{aligned}$$

where the last inequality comes from the assumption (3.9).

Statement 3. Estimate (3.10) with $m=1$ provides similar and weaker estimate with $\|\cdot\|_{\mathrm{var}}$ instead of $\|\cdot\|_{\phi,\mathrm{var}}$. It is another standard observation that such an estimate, together with an estimate of the form

$$\int_{\mathbb{X}}\phi\,\mathrm{d}\mu_t \le \tilde{C}\int_{\mathbb{X}}\phi\,\mathrm{d}\mu, \qquad \mu\in\mathcal{P}, t\ge 0, \tag{5.12}$$

provide (3.11). Again, we explain this argument briefly.

The σ-algebra $\mathcal{F}^X_{\ge r}$ is generated by the algebra $\mathcal{F}^{X,cyl}_{\ge r}$ of the sets of the form

$$B = \big\{\big(X(v_1),\dots,X(v_m)\big)\in C\big\}, \qquad v_1,\dots,v_m\ge r, C\in\mathcal{B}\big(\mathbb{X}^m\big), m\ge 1. \tag{5.13}$$

Hence, in the identity (2.6), we can replace $\sup_{B\in\mathcal{F}^X_{\ge t+s}}$ by $\sup_{B\in\mathcal{F}^{X,cyl}_{\ge t+s}}$. On the other hand, for every B of the form (5.13) with $r=t+s$, we have

$$P_\mu\big(B|\mathcal{F}^X_s\big) = T_t f(X_s), \qquad P_\mu(B) = \int_{\mathbb{X}} T_{t+s} f\,\mathrm{d}\mu$$

with

$$f(x) = P_x\big(\big(X(v_1-t-s),\dots,X(v_m-t-s)\big)\in C\big), \qquad x\in\mathbb{X}$$

and

$$T_r f(x) = \int_{\mathbb{X}} f(y)P_r(x,\mathrm{d}y) = E_x f(X_r),$$

the usual notation for the semigroup generated by the Markov process X. We have

$$\begin{aligned}\left|P_\mu(B|\mathcal{F}_s)-P_\mu(B)\right| &\le \left|T_t f(X_s)-\int_{\mathbb{X}} f\,\mathrm{d}\pi\right|+\left|\int_{\mathbb{X}} f\,\mathrm{d}\pi-\int_{\mathbb{X}} T_{t+s} f\,\mathrm{d}\mu\right| \\ &\le \left\|P_t(X_s,\cdot)-\pi\right\|_{\mathrm{var}}+\|\mu_{t+s}-\pi\|_{\mathrm{var}},\end{aligned}$$

here we have used that $\|f\|\le 1$. Therefore, we have

$$\beta^\mu(t)\le\sup_{s\ge 0}\left(\|\mu_{t+s}-\pi\|_{\mathrm{var}}+E_\mu\left\|P_t(X_s,\cdot)-\pi\right\|_{\mathrm{var}}\right). \tag{5.14}$$

Note that (the weaker version of) (3.10) gives

$$\|\mu_{t+s}-\pi\|_{\mathrm{var}}\le C\mathrm{e}^{-ct}\int_{\mathbb{X}}\phi\,\mathrm{d}\mu,\qquad \left\|P_t(X_s,\cdot)-\pi\right\|_{\mathrm{var}}\le C\mathrm{e}^{-ct}\phi(X_s).$$

These observations combined with (5.12) provide (3.11) with $C'=C(1+\tilde{C})$.

Recall that ϕ satisfies a condition of the form (3.3); denote respective constants by c_L, C_L. Then Corollary 3.1 yields (5.12) with $\tilde{C}=\frac{C_L}{c_L}+1$ because it is supposed that $\phi\ge 1$. These observations finally lead to (3.11) with

$$C'=C\left(2+\frac{C_L}{c_L}\right). \tag{5.15}$$

5.4. Proof of Theorem 3.2

Let γ',δ' be the values introduced in the construction of the function ψ, see Section 5.2. Denote

$$\lambda=\left(\frac{\gamma'}{\gamma}\wedge\frac{\delta'}{\delta}\right)^{-1}.$$

For any signed measure $\varkappa$ on $\mathcal{B}(\mathbb{X}^m)$, by the Hölder inequality, we have

$$\|\varkappa\|_{\phi,\mathrm{var}}\le\left(\int_{\mathbb{X}^m}\left(\sum_{j=1}^m\phi(x_j)\right)^{\sigma p}|\varkappa|(\mathrm{d}x)\right)^{1/p}\left(\int_{\mathbb{X}^m}\left(\sum_{j=1}^m\phi(x_j)\right)^{(1-\sigma)q}|\varkappa|(\mathrm{d}x)\right)^{1/q}$$

for any $\sigma>0$ and any $p,q>1$ with $1/p+1/q=1$. We put $p=(\lambda\sigma)^{-1}$ and take σ close enough to 0, so that $p>1$. Then $\phi^{\sigma p}=\phi^{1/\lambda}$, and

$$\phi^{1/\lambda}(x)=x^{-\gamma((\gamma'/\gamma)\wedge(\delta'/\delta))}\le x^{-\gamma(\gamma'/\gamma)}=\psi(x)$$

for x small enough,

$$\phi^{1/\lambda}(x)=x^{\delta((\gamma'/\gamma)\wedge(\delta'/\delta))}\le x^{\delta(\delta'/\delta)}=\psi(x)$$

for x large enough. Because ϕ is continuous and $\psi \geq 1$, this means that

$$\left(\sum_{j=1}^{m} \phi(x_j)\right)^{\sigma p} \leq C \sum_{j=1}^{m} \psi(x_j) \tag{5.16}$$

with some constant C. We have

$$\frac{1}{q} = 1 - \lambda\sigma, \qquad (1-\sigma)q = \frac{1-\sigma}{1-\lambda\sigma},$$

and in the above construction σ can be taken close enough to 0 in order to provide inequality $(1-\sigma)q \leq 1+\varepsilon$. Then we obtain, finally,

$$\|\varkappa\|_{\phi,\mathrm{var}} \leq C\|\varkappa\|_{\psi,\mathrm{var}}^{1/p} \left(\int_{\mathbb{X}^m} \left(\sum_{j=1}^{m} \phi(x_j)\right)^{1+\varepsilon} |\varkappa|(\mathrm{d}x)\right)^{1/q}. \tag{5.17}$$

Because the weighted total variation norm is a norm indeed, we have

$$\begin{aligned}
&\left\| \int_0^T (\mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m})\,\mathrm{d}t \right\|_{\phi,\mathrm{var}} \\
&\quad \leq \sum_{k=0}^{[T]-1} \left\| \int_k^{k+1} (\mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m})\,\mathrm{d}t \right\|_{\phi,\mathrm{var}} \\
&\qquad + \left\| \int_{[T]}^{T} (\mu_{t+t_1,\dots,t+t_m} - \pi_{t_1,\dots,t_m})\,\mathrm{d}t \right\|_{\phi,\mathrm{var}} \\
&\quad = \sum_{k=0}^{[T]-1} \left\| (\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|_{\phi,\mathrm{var}} + \big(T-[T]\big) \left\| (\mu_{[T]})^{T-[T]}_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|_{\phi,\mathrm{var}};
\end{aligned}$$

recall that μ_t denotes the one-dimensional distribution, see (2.4), and $\mu^t_{t_1,\dots,t_m}$ denotes the Cesàro mean, see (3.6). By (5.17), we have

$$\begin{aligned}
&\left\| (\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|_{\phi,\mathrm{var}} \\
&\quad \leq C \left\| (\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|^{1/p}_{\psi,\mathrm{var}} \left\| (\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|^{1/q}_{\phi^{1+\varepsilon},\mathrm{var}} \\
&\quad \leq C \left\| (\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m} \right\|^{1/p}_{\psi,\mathrm{var}} \\
&\qquad \times \left(\int_{\mathbb{X}^m} \left(\sum_{j=1}^{m} \phi(x_j)\right)^{1+\varepsilon} \big[(\mu_k)^1_{t_1,\dots,t_m} + \pi_{t_1,\dots,t_m}\big](\mathrm{d}x)\right)^{1/q}.
\end{aligned}$$

Recall that ψ satisfies conditions of Proposition 3.1. In addition, it has compact level sets; see condition (ii) in Section 5.3. Then (3.10) with ψ instead of ϕ holds true, and we have

$$\begin{aligned}\|(\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m}\|^{1/p}_{\psi,\mathrm{var}} &= \left\| \int_k^{k+1} (\mu_{t_1+t,\dots,t_m+t} - \pi_{t_1,\dots,t_m})\,\mathrm{d}t \right\|^{1/p}_{\psi,\mathrm{var}} \\ &\le \left(\int_k^{k+1} \|\mu_{t_1+t,\dots,t_m+t} - \pi_{t_1,\dots,t_m}\|_{\psi,\mathrm{var}}\,\mathrm{d}t \right)^{1/p} \\ &\le m^{1/p} C^{1/p} \mathrm{e}^{-ck/p} \left(\int_{\mathbb{X}} \psi\,\mathrm{d}\mu \right)^{1/p}\end{aligned}$$

with the constants c, C from (3.10). Note that $\phi^{1+\varepsilon}$ is integrable w.r.t. π; see Remark 3.1. Then

$$\begin{aligned}\int_{\mathbb{X}^m} \left(\sum_{j=1}^m \phi(x_j) \right)^{1+\varepsilon} \pi_{t_1,\dots,t_m}(\mathrm{d}x) &\le m^{\varepsilon} \int_{\mathbb{X}^m} \sum_{j=1}^m \phi^{1+\varepsilon}(x_j) \pi_{t_1,\dots,t_m}(\mathrm{d}x) \\ &= m^{1+\varepsilon} \int_{\mathbb{X}} \phi^{1+\varepsilon}\,\mathrm{d}\pi < \infty.\end{aligned}$$

On the other hand, by (3.8) with $t = 1$ we have

$$\begin{aligned}\int_{\mathbb{X}^m} \left(\sum_{j=1}^m \phi(x_j) \right)^{1+\varepsilon} (\mu_k)^1_{t_1,\dots,t_m}(\mathrm{d}x) &= \int_{\mathbb{X}^m} \left(\sum_{j=1}^m \phi(x_j) \right)^{1+\varepsilon} \mu^1_{t_1+k,\dots,t_m+k}(\mathrm{d}x) \\ &\le C \int_{\mathbb{X}} \psi\,\mathrm{d}\mu.\end{aligned}$$

Using the elementary inequality

$$(x+y)^{1/q} \le x^{1/q} + y^{1/q}, \qquad x, y > 0, q > 1$$

and the assumption $\psi \ge 1$, we get from the above estimates

$$\|(\mu_k)^1_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m}\|_{\phi,\mathrm{var}} \le \tilde{C}_m \mathrm{e}^{-ck/p} \int_{\mathbb{X}} \psi\,\mathrm{d}\mu \tag{5.18}$$

with some explicitly calculable $\tilde{C}_m$. Similarly to (5.18) (we omit the details), one can show that

$$\big(T - [T]\big) \|(\mu_{[T]})^{T-[T]}_{t_1,\dots,t_m} - \pi_{t_1,\dots,t_m}\|_{\phi,\mathrm{var}} \le \tilde{C}_m \mathrm{e}^{-c[T]/p} \int_{\mathbb{X}} \psi\,\mathrm{d}\mu. \tag{5.19}$$

From (5.18) and (5.19), we obtain the required inequality with $C_m = \tilde{C}_m \sum_{k=0}^{\infty} \mathrm{e}^{-ck/p}$.

5.5. Proof of Theorem 3.3

In order to simplify the notation, we assume $k = 1$ and remove respective subscripts, that is, write f, γ, δ instead of f_i, γ_i, δ_i. One can see that the proof below can be extended to the multidimensional case easily; to do that, it is enough to replace the one-dimensional "deviation inequalities" (5.20) and (5.22) by completely analogous inequalities for the components $f_i, i = 1, \dots, k$ of the multidimensional function f.

We proceed in two steps: the "coupling" one and the "truncation" one.

The "coupling" step deals with the case where for some positive ε the initial distribution μ satisfies (3.15). Let ϕ be defined by (2.5) with γ, δ from (3.13). Then Theorem 3.1 provides that there exists a (μ, π)-coupling (Z^1, Z^2) for the process X, which satisfies (3.9). We have

$$
\begin{aligned}
E_\mu \left| \frac{1}{n} \sum_{l=1}^n f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f \right| &= E \left| \frac{1}{n} \sum_{l=1}^n f\big(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}\big) - a_f \right| \\
&\le E \left| \frac{1}{n} \sum_{l=1}^n f\big(Z^2_{t_1+l}, \dots, Z^2_{t_r+l}\big) - a_f \right| \\
&\quad + \frac{1}{n} \sum_{l=1}^n E\big| f\big(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}\big) - f\big(Z^2_{t_1+l}, \dots, Z^2_{t_r+l}\big)\big|
\end{aligned}
$$

because Z^2 has the same distribution with $\{X^{st}(t), t \ge 0\}$. Recall that X is ergodic, see [12]. Then, by the Birkhoff–Khinchin theorem,

$$
E \left| \frac{1}{n} \sum_{l=1}^n f\big(X^{st}_{t_1+l}, \dots, X^{st}_{t_r+l}\big) - a_f \right| \to 0, \qquad n \to \infty.
$$

On the other hand, by (3.13) we have

$$
\begin{aligned}
& E\big| f\big(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}\big) - f\big(Z^2_{t_1+l}, \dots, Z^2_{t_r+l}\big)\big| \\
&\quad \le C \sum_{j=1}^r E\big(\phi\big(Z^1_{t_j+l}\big) + \phi\big(Z^2_{t_j+l}\big)\big) \mathbf{1}_{(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}) \ne (Z^2_{t_1+l}, \dots, Z^2_{t_r+l})} \\
&\quad \le C \sum_{j=1}^r \sum_{i=1}^r E\big(\phi\big(Z^1_{t_j+l}\big) + \phi\big(Z^2_{t_j+l}\big)\big) \mathbf{1}_{Z^1_{t_i+l} \ne Z^2_{t_i+l}}
\end{aligned}
$$

(note that C here does not coincide with the constant C in (3.13) because $\phi(x) \ne x^{-\gamma} + x^\delta$). By the Hölder inequality and the elementary inequality $(a+b)^p \le 2^{p-1}(a^p + b^p)$, $a, b \ge 0$, $p > 1$, we have for arbitrary $p, q > 1$ with $1/p + 1/q = 1$

$$
\begin{aligned}
& E\big(\phi\big(Z^1_{t_j+l}\big) + \phi\big(Z^2_{t_j+l}\big)\big) \mathbf{1}_{Z^1_{t_i+l} \ne Z^2_{t_i+l}} \\
&\quad \le 2^{(p-1)/p} \big(E\big(\phi^p\big(Z^1_{t_j+l}\big) + \phi^p\big(Z^2_{t_j+l}\big)\big)\big)^{1/p} \big(P\big(Z^1_{t_i+l} \ne Z^2_{t_i+l}\big)\big)^{1/q}.
\end{aligned}
$$

We can take $p > 1$ close enough to 1, so that $\gamma' = \gamma p < \gamma + \varepsilon$, $\delta' = \delta p < \delta + \varepsilon$, and γ', δ' satisfy (3.1). Then $\phi' = \phi^p$ clearly has the form (2.5) with γ', δ' instead of γ, δ. Corollary 3.1 applied to ϕ' instead of ϕ yields that

$$\sup_{t\ge 0} E\phi^p\big(Z_t^1\big) < \infty, \qquad \sup_{t\ge 0} E\phi^p\big(Z_t^2\big) = \int_{\mathbb{X}} \phi^p \,\mathrm{d}\pi < \infty.$$

On the other hand, (3.9) and standing assumption $\phi \ge 1$ yield

$$P\big(Z_t^1 \ne Z_t^2\big) \le C\mathrm{e}^{-ct} \int_{\mathbb{X}} \phi \,\mathrm{d}\mu, \qquad t \ge 0,$$

where c, C are the same as in (3.9). Summarizing all the above, we obtain

$$E\big|f\big(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}\big) - f\big(Z^2_{t_1+l}, \dots, Z^2_{t_r+l}\big)\big| \le C' \sum_{i=1}^r \mathrm{e}^{-c(t_i+l)/q} \tag{5.20}$$

with the same constant c and some constant C' which depends on ϕ, p, μ, and the constants C in (3.13) and (3.9). Therefore

$$E_\mu \left| \frac{1}{n} \sum_{l=1}^n f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f \right| \to 0, \qquad n \to \infty,$$

which completes the proof of statement 1 under the assumption (3.15). To prove statement 2, we need to show that for any bounded Lipschitz continuous function $F : \mathbb{R} \to \mathbb{R}$

$$E_\mu F\big(S_n(X)\big) \to \int_{\mathbb{R}} F(y)\nu_f(\mathrm{d}y), \tag{5.21}$$

where $\nu_f \sim \mathcal{N}(0, \Sigma_f^d)$ and

$$S_n(X) = \frac{1}{\sqrt{n}} \sum_{l=1}^n \big(f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f\big).$$

In [4], Remark 3.1, it was shown that the general result by Genon-Catalot *et al.* (see [12], Corollary 2.1) can be applied to prove that the stationary Fisher–Snedecor diffusion is an α-mixing process with an exponential decay rate. Then the CLT for α-mixing sequences (see [14]) provide

$$EF\big(S_n\big(X^{st}\big)\big) \to \int_{\mathbb{R}} F(y)\nu_f(\mathrm{d}y).$$

On the other hand, the estimates similar to those made above provide that

$$\begin{aligned} &\big|E_\mu F\big(S_n(X)\big) - EF\big(S_n\big(X^{st}\big)\big)\big| \\ &\quad\le \frac{\mathrm{Lip}(F)}{\sqrt{n}} \sum_{l=1}^n E\big|f\big(Z^1_{t_1+l}, \dots, Z^1_{t_r+l}\big) - f\big(Z^2_{t_1+l}, \dots, Z^2_{t_r+l}\big)\big| \le \frac{C' \,\mathrm{Lip}(F)}{\sqrt{n}} \end{aligned} \tag{5.22}$$

with some constant C'. This proves statement 2 under the assumption (3.15).

The "truncation" step removes the assumption (3.15). For an arbitrary μ and any $a \in (0,1)$ there exist $\mu_a, \mu^a \in \mathcal{P}$ such that μ_a is supported in some segment $[u, v] \subset (0, \infty)$, and

$$\mu = (1-a)\mu_a + a\mu^a.$$

Then $P_\mu = (1-a)P_{\mu_a} + aP_{\mu^a}$, and μ_a satisfies (3.15). Hence, for any $\zeta > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} P_\mu\left(\left|\frac{1}{n}\sum_{l=1}^n f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f\right| > \zeta\right) \\
&\quad \le a \limsup_{n\to\infty} P_{\mu^a}\left(\left|\frac{1}{n}\sum_{l=1}^n f(X_{t_1+l}, \dots, X_{t_r+l}) - a_f\right| > \zeta\right) \le a.
\end{aligned}$$

Because a is arbitrary, this proves statement 1 for arbitrary μ. Similar argument proves (5.21) for arbitrary μ, and completes the proof of the theorem.

5.6. Proof of Theorem 3.4

Again, we assume $k = 1$. We note that both statement 1 and statement 2 hold true under the respective conditions of Theorem 3.3. The proof of this fact is analogous to the proof of Theorem 3.3 and therefore is omitted. The only difference is that, in this proof, one requires the continuous-time version of the CLT (3.21) for the stationary version X^{st} of the process X instead of the discrete-time one. This statement can be easily derived from the respective discrete-time one by the standard discretization argument (see, e.g., [8], pages 178–179). Hence, our task is to reduce the conditions of Theorem 3.3 to those of Theorem 3.4.

First, note that we can increase slightly γ, so that the conditions of Theorem 3.4 still hold true. Let ϕ be defined by (2.5) with this new γ and δ from the formulation of the theorem. Because $\alpha > 2$, condition (3.20) yields (3.5). Then we can apply Proposition 3.2 and define respective function ψ, see Section 5.2. While doing that, we can choose γ', δ' larger than, but close enough to $(\gamma - 1) \vee 0, \delta$, respectively, so that $\int_{\mathbb{X}} \psi \, d\mu < \infty$ if μ is supposed to satisfy (3.19) and

$$\gamma' + \gamma < \frac{\alpha}{2}, \qquad \delta' + \delta < \frac{\beta}{2} \tag{5.23}$$

if γ, δ satisfy (3.20). We put

$$\|f\|_\phi = \sup_{x=(x_1,\dots,x_r)} \frac{|f(x)|}{\sum_{j=1}^r \phi(x_j)}, \qquad f_n(x) = f(x)\prod_{j=1}^r \mathbf{1}_{x_j \ge 1/n}, \qquad n \ge 1.$$

For arbitrary $t_1, \dots, t_r \ge 0$ one has

$$E\sum_{j=1}^r \phi\big(X_{t_j}^{st}\big) = r\int_0^\infty \phi(x)\pi(dx) < \infty$$

because γ, δ satisfy (3.5). Then, by (3.13) and the Lebesgue dominated convergence theorem, $a_{f_n} \to a_f$.

We put $\tilde{f}_n = f_n + a_f - a_{f_n}$. Then the condition (3.13) with the initial γ provide that

$$\|f - \tilde{f}_n\|_\phi \to 0, \qquad n \to \infty. \tag{5.24}$$

On the other hand, $a_{\tilde{f}_n} = a_f$, and every $\tilde{f}_n$ satisfy conditions of Theorem 3.3. Hence, for every n

$$\begin{aligned}
&\limsup_{T\to\infty} E_\mu \left| \frac{1}{T} \int_0^T f(X_{t_1+t}, \dots, X_{t_r+t})\,\mathrm{d}t - a_f \right| \\
&\quad \le \limsup_{T\to\infty} \frac{1}{T} E_\mu \int_0^T \left| f(X_{t_1+t}, \dots, X_{t_r+t}) - \tilde{f}_n(X_{t_1+t}, \dots, X_{t_r+t}) \right| \mathrm{d}t \\
&\quad \le \limsup_{T\to\infty} \frac{C\|f - \tilde{f}_n\|_\phi}{T} E_\mu \int_0^T \sum_{j=1}^r \phi(X_{t_j+t})\,\mathrm{d}t \\
&\quad = C\|f - \tilde{f}_n\|_\phi \limsup_{T\to\infty} \sum_{j=1}^r \left(\frac{T + t_j}{T} \int_{\mathbb{X}} \phi\,\mathrm{d}\mu^{T+t_j} - \frac{t_j}{T} \int_{\mathbb{X}} \phi\,\mathrm{d}\mu^{t_j} \right).
\end{aligned}$$

Then from (3.8) with $m = 1$ and $\varepsilon = 0$ we obtain that, when μ satisfies (3.19),

$$\limsup_{T\to\infty} E_\mu \left| \frac{1}{T} \int_0^T f(X_{t_1+t}, \dots, X_{t_r+t})\,\mathrm{d}t - a_f \right| \le C\|f - \tilde{f}_n\|_\phi$$

with some constant C. Because n is arbitrary and (5.24) holds, this proves (3.18) in the mean sense. If (3.19) fails, then (3.18) still holds in the sense of convergence of probability; one can show this using the truncation argument from the previous section. This proves statement 1.

Denote $Q = \max_j t_j - \min_j t_j$ and assume that $T > Q$. Then

$$\begin{aligned}
&E_\mu \left[\frac{1}{\sqrt{T}} \int_0^T \big(f(X_{t_1+t}, \dots, X_{t_r+t}) - \tilde{f}_n(X_{t_1+t}, \dots, X_{t_r+t}) \big)\,\mathrm{d}t \right]^2 \\
&\quad \le \frac{2}{T} \left[\int_0^T \int_s^{T\wedge(s+Q)} + \int_0^{T-Q} \int_{s+Q}^T \right] E_\mu \big(f(X_{t_1+t}, \dots, X_{t_r+t}) - \tilde{f}_n(X_{t_1+t}, \dots, X_{t_r+t}) \big) \\
&\qquad\qquad \times \big(f(X_{t_1+s}, \dots, X_{t_r+s}) - \tilde{f}_n(X_{t_1+s}, \dots, X_{t_r+s}) \big)\,\mathrm{d}t\,\mathrm{d}s \\
&\quad =: I_1 + I_2.
\end{aligned}$$

We estimate I_1, I_2 separately. We explain the estimates in the particular case $r = 2, t_1 = 0, t_2 = Q$; the general case is quite analogous, but the calculations are more cumbersome. We have

$$I_1 \le \frac{C\|f - \tilde{f}_n\|_\phi^2}{T} \int_0^T \int_s^{T\wedge(s+Q)} E_\mu \big(\phi(X_t) + \phi(X_{t+Q})\big)\big(\phi(X_s) + \phi(X_{s+Q})\big)\,\mathrm{d}t\,\mathrm{d}s. \tag{5.25}$$

By the Markov property of the process X,

$$\int_0^T \int_s^{T\wedge(s+Q)} E_\mu \phi(X_t)\phi(X_s)\,\mathrm{d}t\,\mathrm{d}s \le E_\mu \int_0^T \phi(X_s)\biggl(\int_0^Q T_v\phi(X_s)\,\mathrm{d}v\biggr)\,\mathrm{d}s;$$

here we have used the standard notation

$$T_v f(x) = \int_{\mathbb{X}} f(y) P_t(x, \mathrm{d}y).$$

Note that $P_t(x,\cdot) = (\delta_x)_t$. Hence, by (3.8) with $m = 1$, $\varepsilon = 0$, and $\mu = \delta_x$, we have

$$\int_0^Q T_v\phi(x)\,\mathrm{d}v \le QC\psi(x), \qquad x \in \mathbb{X}. \tag{5.26}$$

By the inequalities (5.23), the function $\Phi = \phi\psi$ has the form (2.5) with the parameters satisfying (3.5). Then, using once again (3.8) with Φ instead of ϕ, we get

$$\int_0^T \int_s^{T\wedge(s+Q)} E_\mu \phi(X_t)\phi(X_s)\,\mathrm{d}t\,\mathrm{d}s \le QCE_\mu \int_0^T \Phi(X_s)\,\mathrm{d}s \le TQC' \int_{\mathbb{X}} \Phi\,\mathrm{d}\mu;$$

the constants C, C' here depend on ϕ, ψ, etc., but does not depend on Q, T, and μ. Similar calculations provide estimates for other parts of the integral in the right-hand side of (5.25). For instance, changing the variables $s' = s + Q$ and using the Markov property at the point $t \le s'$, we get

$$\begin{aligned}
&\int_0^T \int_s^{T\wedge(s+Q)} E_\mu \phi(X_t)\phi(X_{s+Q})\,\mathrm{d}t\,\mathrm{d}s\\
&\quad= E_\mu \int_0^T \phi(X_t)\biggl(\int_{Q\vee t}^{t+Q} \phi(X_{s'})\,\mathrm{d}s'\biggr)\,\mathrm{d}t\\
&\quad\le E_\mu \int_0^T \phi(X_t)\biggl(\int_0^Q T_v\phi(X_t)\,\mathrm{d}v\biggr)\,\mathrm{d}t \le TQC' \int_{\mathbb{X}} \Phi\,\mathrm{d}\mu;
\end{aligned}$$

in the last inequality we use (5.26) and (3.8) with Φ instead of ϕ.

Summarising these estimates, we get

$$I_1 \le CQ\|f - \tilde{f}_n\|_\phi^2 \int_{\mathbb{X}} \Phi\,\mathrm{d}\mu.$$

To estimate I_2, we use the Markov property at the time moment $s + Q$ and write

$$I_2 \le \frac{C\|f - \tilde{f}_n\|_\phi}{T} E_\mu \int_0^T \bigl(\phi(X_s) + \phi(X_{s+Q})\bigr) F_s^{n,Q,T}(X_{s+Q})\,\mathrm{d}s$$

with

$$F_s^{n,Q,T}(x) = \biggl|\int_0^{T-s-Q} E_x\bigl(f(X_t, X_{t+Q}) - \tilde{f}_n(X_t, X_{t+Q})\bigr)\,\mathrm{d}t\biggr|.$$

Denote $g_n = f - \tilde{f}_n$. Because, by the construction, $a_f = a_{\tilde{f}_n}$, we have $\int_{\mathbb{X}^2} g_n \, \mathrm{d}\pi_{t,t+Q} = 0$ for every t. Then

$$F_s^{n,Q,T}(x) = \left| \int_{\mathbb{X}} g_n \, \mathrm{d}\left(\int_0^{T-s-Q} \big((\delta_x)_{t,t+Q} - \pi_{t,t+Q} \big) \, \mathrm{d}t \right) \right|.$$

Clearly,

$$\left| \int_{\mathbb{X}^m} g \, \mathrm{d}\varkappa \right| \le \|g\|_\phi \|\varkappa\|_{\phi,\mathrm{var}}$$

for any measurable function g on $\mathbb{X}^m$ and any signed measure $\varkappa$. Then, by (3.12),

$$F_s^{n,Q,T}(x) \le C \|f - \tilde{f}_n\|_\phi \psi(x).$$

Recall that ψ satisfies the Lyapunov-type condition (3.3). Then by the Markov property and the moment bound from Corollary 3.1 we have $E_\mu \phi(X_s)\psi(X_{s+Q}) \le C E_\mu \phi(X_s)\psi(X_s)$, which together with the preceding estimate gives

$$I_2 \le \frac{C\|f - \tilde{f}_n\|_\phi^2}{T} E_\mu \int_0^T \big(\phi(X_s)\psi(X_s) + \phi(X_{s+Q})\psi(X_{s+Q}) \big) \, \mathrm{d}s.$$

Using once again (3.8) with $\Phi = \phi\psi$ instead of ϕ and recalling the estimates for I_1, we get finally

$$E_\mu \left[\frac{1}{\sqrt{T}} \int_0^T \big(f(X_t) - \tilde{f}_n(X_t) \big) \, \mathrm{d}t \right]^2 \le C \|f - \tilde{f}_n\|_\phi^2 \int_{\mathbb{X}} \Phi \, \mathrm{d}\mu. \tag{5.27}$$

By the construction, every f_n satisfies conditions of Theorem 3.3, and therefore (3.21) holds true with f_n instead of f. Then, if Φ is integrable w.r.t. μ, (5.27) and the approximation argument, similar to the one used in the proof of Theorem 3.3, lead to (3.21) for f with

$$\Sigma_f^c = \lim_{n \to \infty} \Sigma_{f_n}^c. \tag{5.28}$$

On the other hand, if we write

$$\Sigma_{f,R}^c = \int_{-R}^R \mathrm{Cov}\big(f\big(X_{t_1+t}^{st}, \dots, X_{t_r+t}^{st}\big), f\big(X_{t_1}^{st}, \dots, X_{t_r}^{st}\big) \big) \, \mathrm{d}t,$$

then

$$\big| \Sigma_{f,R}^c - \Sigma_{\tilde{f}_n,R}^c \big| \le C \|f - \tilde{f}_n\|_\phi^2 \int_{\mathbb{X}} \Phi \, \mathrm{d}\pi; \tag{5.29}$$

the proof of (5.29) is similar to the proof of (5.27) and is omitted. Therefore the integral (3.22) coincides with the limit (5.28). This completes the proof of statement 2 when Φ is integrable w.r.t. μ. For general μ, we use the truncation argument from the previous section.

5.7. Proof of Theorem 3.5

Again, we restrict ourselves by the case $k = 1$. The proof is based on the following auxiliary estimate.

Lemma 5.1. *Under conditions of Theorem 3.5, for any T*

$$E\left(\int_0^T \bigl(f(X_{t_1+t}, \dots, X_{t_r+t}) - a_f\bigr)\,\mathrm{d}t\right)^2 \le CT\|f\|_\phi^2 \int_{\mathbb{X}} \Phi\,\mathrm{d}\mu$$

with some Φ satisfying conditions of statement 1 of Proposition 3.1.

Proof. We assume that f is centered and $r = 1$. The general case can be reduced to this one using the same arguments with those explained Section 5.6.

We proceed like in Section 5.6: take ψ of the form (2.5) with $\gamma' \in ((\gamma - 1) \vee 0, \alpha/2 - 1)$, $\delta' < \beta/2$ such that $\gamma + \gamma' < \alpha/2 - 1$, $\delta + \delta' < \beta/2$ and put $\Phi = \phi\psi$. Then

$$\begin{aligned}E\left(\int_0^T f(X_t)\,\mathrm{d}t\right)^2 &= 2\int_0^T Ef(X_s)\int_s^T f(X_t)\,\mathrm{d}t\,\mathrm{d}s \\ &\le 2\|f\|_\phi \int_0^T E\bigl|f(X_s)\bigr| \left\| \int_0^{T-s} \bigl((\delta_{X_s})_r - \pi\bigr)\,\mathrm{d}r \right\|_{\phi,\mathrm{var}} \mathrm{d}s \\ &\le C\|f\|_\phi \int_0^T E\bigl|f(X_s)\bigr|\psi(X_s)\,\mathrm{d}s,\end{aligned}$$

here we have used the Markov property and Theorem 3.2. On the other hand, Corollary 3.1 applied to Φ instead of ϕ gives

$$\int_0^T E\bigl|f(X_s)\bigr|\psi(X_s)\,\mathrm{d}s \le \|f\|_\phi \int_0^T \left(\int_{\mathbb{X}} \Phi\,\mathrm{d}\mu_s\right)\mathrm{d}s \le C\|f\|_\phi T \int_{\mathbb{X}} \Phi\,\mathrm{d}\mu$$

with some other constant C, which completes the proof. □

Let us proceed with the proof of the theorem. By Theorem 3.4, finite-dimensional distributions of Y_T converge to that of B. Hence, we need to prove the weak compactness, only. In addition, it is sufficient to prove weak compactness in $D([0, 1])$ instead of $C([0, 1])$: when we succeed to do that, we get the weak convergence $Y_T \Rightarrow B$ in $D([0, 1])$. Because both Y_T and B have continuous trajectories, this would imply the weak convergence $Y_T \Rightarrow B$ in $C([0, 1])$.

For the function Φ constructed in the proof of Lemma 5.1, there exists $q > 1$ such that Φ^q still satisfies conditions of Proposition 3.1, statement 1. Then, for p such that $1/p + 1/q = 1$, we have for every $v_1 < v_2 < v_3$

$$\begin{aligned}&E\bigl|Y_T(v_1) - Y_T(v_2)\bigr|^{2/p}\bigl|Y_T(v_2) - Y_T(v_3)\bigr|^2 \\ &\quad \le C\|f\|_\phi^2(v_3 - v_2)E\bigl|Y_T(v_1) - Y_T(v_2)\bigr|^{2/p}\Phi\bigl(X(v_2T)\bigr) \qquad (5.30)\end{aligned}$$

$$\leq C\|f\|_{\phi}^{2}(v_3-v_2)\big(E\big|Y_T(v_1)-Y_T(v_2)\big|^2\big)^{1/p}\big(E\Phi^q\big(X(v_2T)\big)\big)^{1/q}$$
$$\leq C\|f\|_{\phi}^{2+2/p}(v_3-v_2)(v_2-v_1)^{1/p}E\Phi\big(X(v_1T)\big)^{1/p}E\Phi^q\big(X(v_2T)\big)^{1/q}$$
$$\leq C\|f\|_{\phi}^{2+2/p}(v_3-v_1)^{1+1/p}\left(\int_{\mathbb{X}}\Phi\,\mathrm{d}\mu\right)^{1/p}\left(\int_{\mathbb{X}}\Phi^q\,\mathrm{d}\mu\right)^{1/q}.$$

Here we have used subsequently Lemma 5.1, the Hölder inequality, Lemma 5.1 again, and Corollary 3.1 with Φ, Φ^q instead of ϕ. Theorem 15.6 in [8] and (5.30) provide weak compactness in $D([0,1])$ of the family $\{X_T\}$.

5.8. Proof of Theorem 4.1

By Example 4.5, under the assumptions of Theorem 4.1, for any fixed $t>0$ either $(\overline{m}_{-1,c},\overline{m}_{1,c},\overline{m}_{2,c},\overline{R}_c(t))$ or $(\overline{m}_{-1,d},\overline{m}_{1,d},\overline{m}_{2,d},\overline{R}_d(t))$ is an asymptotically normal estimator of $(m_{-1,c},m_{1,c},m_{2,c},R(t))$. Note that the assumption $\alpha>2$, $\beta>8$ (in the continuous-time case) is equivalent to

$$\{-1,1,2\}\in\left(-\frac{\alpha}{4}-\frac{1}{2},\frac{\beta}{4}\right),$$

while the assumption $\alpha>4$, $\beta>8$ (in the discrete-time case) is equivalent to

$$\{-1,1,2\}\in\left(-\left(\frac{\alpha}{2}-1\right)\wedge\left(\frac{\alpha}{4}\right),\frac{\beta}{4}\right).$$

The invariant distribution density for the process X can be written in the form

$$\mathfrak{p}(x)=\frac{1}{xB(\alpha/2,\beta/2)}\left(\frac{\alpha x}{\alpha x+\varrho}\right)^{\alpha/2}\left(\frac{\varrho}{\alpha x+\varrho}\right)^{\beta/2} \tag{5.31}$$

with $\varrho=(\beta-2)\kappa/\beta$. Respective moments are equal

$$m_\upsilon=\int_0^\infty x^\upsilon\mathfrak{p}(x)\,\mathrm{d}x=\left(\frac{\varrho}{\alpha}\right)^\upsilon\frac{\Gamma(\alpha/2+\upsilon)\Gamma(\beta/2-\upsilon)}{\Gamma(\alpha/2)\Gamma(\beta/2)},\qquad \upsilon\in\left(-\frac{\alpha}{2},\frac{\beta}{2}\right). \tag{5.32}$$

In particular,

$$m_{-1}=\frac{\alpha}{(\alpha-2)(\beta-2)\kappa},\qquad m_1=\frac{\kappa}{\beta},\qquad m_2=\frac{(\alpha+2)(\beta-2)\kappa^2}{\alpha(\beta-4)\beta^2}.$$

On the other hand, one has

$$\operatorname{Corr}\big(X_0^{st},X_t^{st}\big)=\mathrm{e}^{-\theta t},$$

see [7], Theorem 2.3(iii). Resolving the above identities for a fixed t, we can write $(\alpha,\beta,\kappa,\theta)=G(m_{-1},m_1,m_2,R(t))$ with

$$G_1(x,y,z,w)=\frac{2(xyz-y^2)}{xyz-2z+y^2},\qquad G_2(x,y,w)=\frac{4x(z-y^2)}{xz-2xy^2+y},$$
$$G_3(x,y,z,w)=\frac{4xy(z-y^2)}{xz-2xy^2+y},\qquad G_4(x,y,z,w)=-\frac{1}{t}\log\left(\frac{w}{z-y^2}\right).$$

Clearly, the function G is well defined and smooth in some neighbourhood of the point

$$\mathbf{x}=\bigl(m_{-1}(\alpha,\beta,\kappa,\theta),m_1(\alpha,\beta,\kappa,\theta),m_2(\alpha,\beta,\kappa,\theta),\bigl[R(t)\bigr](\alpha,\beta,\kappa,\theta)\bigr).$$

Then one can obtain the required statements using the continuity mapping theorem and the functional delta method (see [25], Theorem 3.3.A). Asymptotic covariance matrices for $(\widehat{\alpha}_c,\widehat{\beta}_c,\widehat{\kappa}_c,\widehat{\theta}_c)$ and $(\widehat{\alpha}_d,\widehat{\beta}_d,\widehat{\kappa}_d,\widehat{\theta}_d)$, are given by the formula

$$\Sigma_c(\alpha,\beta,\kappa,\theta)=D\Sigma_c D^\top,\qquad \Sigma_d(\alpha,\beta,\kappa,\theta)=D\Sigma_d D^\top, \tag{5.33}$$

where Σ_c, Σ_d are the asymptotic covariance matrices for

$$\bigl(\overline{m}_{-1,c},\overline{m}_{1,c},\overline{m}_{2,c},\overline{R}_c(t)\bigr),\qquad \bigl(\overline{m}_{-1,d},\overline{m}_{1,d},\overline{m}_{2,d},\overline{R}_d(t)\bigr),$$

respectively, and $D_{ij}=[\frac{\partial G_i}{\partial x_j}](\mathbf{x})$, $i,j\in\{1,2,3,4\}$.

Acknowledgements

The authors express their deep gratitude to the referee for numerous comments and suggestions that were helpful and led for substantial improvement of the paper.

A.M. Kulik was supported in part by the State fund for fundamental researches of Ukraine and the Russian foundation for basic research, Grant F40.1/023.

N.N. Leonenko was supported in part by grant of the European commission PIRSES-GA-2008-230804 (Marie Curie).

References

[1] Abourashchi, N. and Veretennikov, A.Y. (2009). On exponential mixing bounds and convergence rate for reciprocal gamma diffusion processes. *Math. Commun.* **14** 331–339. MR2743181

[2] Aburashchi, N. and Veretennikov, O.Y. (2009). On exponential bounds for mixing and the rate of convergence for Student processes. *Teor. Ĭmovīr. Mat. Stat.* **81** 1–12. MR2667305

[3] Anulova, S.V., Veretennikov, A.Y., Krylov, N.V., Liptser, R.Sh. and Shiryaev, A.N. (1989). Stochastic Calculus: Probability Theory—3. In *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. Fund. Napr.* **45**. Moscow: VINITI. MR1039617

[4] Avram, F., Leonenko, N.N. and Šuvak, N. (2011). On spectral properties and statistical analysis of Fisher–Snedecor diffusion. Preprint, available at arXiv:1007.4909.

[5] Avram, F., Leonenko, N.N. and Šuvak, N. (2011). Parameter estimation for Fisher–Snedecor diffusion. *Statistics* **45** 27–42. MR2772153
[6] Bhattacharya, R.N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60** 185–201. MR0663900
[7] Bibby, B.M., Skovgaard, I.M. and Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* **11** 191–220. MR2132002
[8] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[9] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042
[10] Down, D., Meyn, S.P. and Tweedie, R.L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. MR1379163
[11] Forman, J.L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scand. J. Statist.* **35** 438–465. MR2446729
[12] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079. MR1809735
[13] Gīhman, Ĭ.Ī. and Skorohod, A.V. (1972). *Stochastic Differential Equations*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **72**. New York: Springer. Translated from the Russian by Kenneth Wickwire. MR0346904
[14] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. *Probability and Mathematical Statistics*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0624435
[15] Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. *North-Holland Mathematical Library* **24**. Amsterdam: North-Holland. MR0637061
[16] Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0611513
[17] Kolmogoroff, A. (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104** 415–458. MR1512678
[18] Kulik, A.M. (2009). Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.* **119** 602–632. MR2494006
[19] Kulik, A.M. (2011). Asymptotic and spectral properties of exponentially ϕ-ergodic Markov processes. *Stochastic Process. Appl.* **121** 1044–1075. MR2775106
[20] Kulik, A.M., Leonenko, N.N. and Šuvak, N. (2012). Statistical inference for Fisher–Snedecor diffusion process. Preprint.
[21] Ladyzhenskaja, O.A., Solonnikov, V.A. and Ural'ceva, N.N. (1968). *Linear and Quasilinear Equations of Parabolic Type*. *Translations of Mathematical Monographs* **23**. Providence, RI: Amer. Math. Soc. Translated from the Russian by S. Smith. MR0241822
[22] Leonenko, N.N. and Šuvak, N. (2010). Statistical inference for reciprocal gamma diffusion process. *J. Statist. Plann. Inference* **140** 30–51. MR2568120
[23] Leonenko, N.N. and Šuvak, N. (2010). Statistical inference for student diffusion process. *Stoch. Anal. Appl.* **28** 972–1002. MR2739327
[24] Pearson, K. (1914). *Tables for Statisticians and Biometricians*. Cambridge: Cambridge Univ. Press.
[25] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165
[26] Shaw, T.W. and Munir, A. (2009). Dependency without copulas or ellipticity. *European Journal of Finance* **15** 661–674.
[27] Shiryayev, A.N. (1992). On analytic methods in probability theory. In *Selected Works of A.N. Kolmogorov*, *Volume II* 62–108. Dordrecht: Kluwer Academic Publisher.
[28] Veretennikov, A.Y. (1987). Estimates of the mixing rate for stochastic equations. *Theory Probab. Appl.* **32** 299–308.

Received October 2011 and revised May 2012