Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

SIMULATION-BASED ASSESSMENT OF THE STATIONARY TAIL DISTRIBUTION OF A STOCHASTIC DIFFERENTIAL EQUATION

Krzysztof Bisewski
Daan Crommelin

Michel Mandjes

Centrum Wiskunde & Informatica
Science Park 123
1098 XG, Amsterdam, NETHERLANDS

University of Amsterdam
Science Park 105
1098 XG, Amsterdam, NETHERLANDS

ABSTRACT

A commonly used approach to analyzing stochastic differential equations (SDEs) relies on performing Monte Carlo simulation with a discrete-time counterpart. In this paper we study the impact of such a time-discretization when assessing the stationary tail distribution. For a family of semi-implicit Euler discretization schemes with time-step $h > 0$, we quantify the relative error due to the discretization, as a function of h and the exceedance level x. By studying the existence of certain (polynomial and exponential) moments, using a sequence of prototypical examples, we demonstrate that this error may tend to 0 or ∞. The results show that the original shape of the tail can be heavily affected by the discretization. The cases studied indicate that one has to be very careful when estimating the stationary tail distribution using Euler discretization schemes.

1 INTRODUCTION

Let $(X_t)_{t \geq 0}$ solve the stochastic differential equation (SDE)

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t\,, \tag{1}$$

with an initial condition $X_0 \sim \xi$. The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are called *drift* and *volatility* respectively, while ξ follows an arbitrary, tight (possibly degenerate) probability law concentrated on $\mathbb{R}$. SDEs are used in a variety of application areas, e.g. chemistry (Yang et al. 2006) and climate science (Majda et al. 2009). Under some conditions on f and g, X_t converges to a stationary distribution as $t \to \infty$. Let μ_0 be the corresponding stationary (or *invariant, ergodic*) measure, that is, the unique probability measure such that $X_0 \sim \mu_0$ implies $X_t \sim \mu_0$ for all $t > 0$. In the following, we abbreviate $\bar{\mu}_0(x) := \mu_0((x, \infty))$.

We are interested in determining the *shape of the tail* of μ_0, i.e., the way $\bar{\mu}_0(x)$ decays to 0 as $x \to \infty$. Besides the one-dimensional case, no explicit expressions for $\bar{\mu}_0(x)$ are available, thus motivating the use of simulation-based methods. Ideally, one would sample a path of $(X_t)_{t \geq 0}$ (in continuous time, that is), and estimate $\bar{\mu}_0(x)$ by the fraction of time it spends above level x in a time interval $[0, T]$ (which, by the ergodic theorem, converges to $\bar{\mu}_0(x)$ as $T \to \infty$). It is evidently impossible to sample a continuous and infinitely long path of a process $(X_t)_{t \geq 0}$ on a computer, explaining the need for time-discretization and truncation. Discretization schemes are not exact and may intrinsically change the dynamics of the original continuous-time process defined through (1). As a consequence, the stationary measure pertaining to the discretized process will generally differ from μ_0 (or might not even exist!) A few relevant references on this topic are Roberts and Tweedie (1996), Stramer and Tweedie (1999), and Mattingly et al. (2002).

In this paper we study the effect of discretization on the shape of the tail of the stationary distribution. In order to illustrate the problem that might occur, we use the Ornstein-Uhlenbeck (OU) process as an example. Let $(X_t)_{t \geq 0}$ solve $\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$; it can be shown that $\mu_0 \sim N(0, 1)$. The forward Euler

discretization scheme with time-step $h < 2$ can be shown to have invariant measure $\mu_h \sim N(0, (1 - h/2)^{-1})$. Both distributions may look similar, but for any fixed h the ratio $\bar{\mu}_h(x)/\bar{\mu}_0(x)$ explodes as $x \to \infty$, entailing that in this example the forward Euler scheme creates a stationary distribution with tails *heavier* than those of the original continuous-time process. Whereas in the OU case μ_h and μ_0 belong to the same class of distributions (i.e., Gaussian), we will identify other examples in which they belong to different classes of distributions; we even find an instance in which $\bar{\mu}_0(x)$ has essentially the shape $\exp(-x^2)$, whereas $\bar{\mu}_h(x)$ has power-law decay; see Section 4.4. Thus, an important message from our work is that one should be extremely careful when estimating the stationary tail distribution using time-discretization based simulation.

In earlier studies, it has been observed that the discretization influences the stationary distribution. Talay (1990) showed that under regularity conditions, for a Milstein scheme with time-step $h > 0$, (i) there exists a stationary distribution μ_h and (ii) for a class of functions $r : \mathbb{R} \to \mathbb{R}$, $|\int r(x) \mu_h(\mathrm{d}x) - \int r(x) \mu(\mathrm{d}x)| \to 0$, as $h \downarrow 0$. Recently Abdulle et al. (2014) extended this result to higher order schemes. These results are reassuring, as they show that at least in some sense μ_h is close to μ_0, but unfortunately they do not provide any insight into the level of resemblance between the respective tails.

Schurz (1999) proposes a family of θ-implicit Euler schemes to discretize a system of linear SDEs. The author proves that in case of additive noise, the trapezoidal rule (i.e., the semi-implicit scheme with implicitness parameter θ equal to $\frac{1}{2}$) gives the same stationary distribution as the original system, independent of the choice of h. This observation has motivated us to also consider a family of semi-implicit Euler schemes in this paper. Mattingly et al. (2002) provide general conditions under which the ergodicity properties of the original continuous-time process carry over to various discretization schemes. In this paper we discuss one of those, viz. an implicit *split-step* Euler scheme.

To study $\bar{\mu}_h(x)$ under various discretization schemes, we use tools from the theory of random iterated functions, and results on the existence of moments of stationary distributions of Markov chains.

The manuscript is organized as follows. In Section 2 we introduce the discretization schemes that are used throughout the paper, We briefly discuss the existence and uniqueness of the stationary distribution of the continuous-time system (1) and its discretized version, and we specify the term 'shape of the tail'. In Section 3 we introduce our key tools; most notably, we recall a theorem by Goldie (1991) on random iterated functions, which we use to describe the behavior of the tail of a stationary distribution for linear systems of SDEs under the discretization, and a theorem based on Tweedie (1983), which gives conditions for existence and nonexistence of moments. In Section 4 we assess (analytically and numerically) the ratio $\gamma_h(x) := \bar{\mu}_h(x)/\bar{\mu}_0(x)$, using four illustrative examples and various discretization schemes.

2 PRELIMINARIES

In this section we first introduce discretization schemes. Then we present results from the literature on the existence and uniqueness of the invariant measure for SDEs and corresponding discretized versions. Finally, we introduce the notion of the 'shape of the tail'.

2.1 Discretization Schemes for SDEs

Let $(X_t)_{t \geq 0}$ be driven by (1). In order to simulate X_t on a computer we use discretization schemes; for a survey on discretization methods for SDEs we refer to Kloeden and Platen (1992). Let $(W_t)_{t \geq 0}$ be a standard Wiener process and $\Delta W_n := W_{n+1} - W_n$. As motivated in the introduction, we focus on semi-implicit discretization schemes (parametrized by $\theta \in [0, 1]$) and split-step schemes. We distinguish two schemes:

○   The (standard) θ-implicit Euler-Maruyama scheme

$$X_{n+1} := X_n + \big(\theta f(X_{n+1}) + (1 - \theta) f(X_n)\big) h + g(X_n) \sqrt{h} \, \Delta W_n. \tag{2}$$

○   The split-step θ-implicit Euler-Maruyama scheme

$$\begin{cases} X_{n+1}^{\star} := X_n + \big(\theta f(X_{n+1}^{\star}) + (1-\theta)f(X_n)\big)h, \\ X_{n+1} := X_{n+1}^{\star} + g(X_n)\sqrt{h}\,\Delta W_n. \end{cases} \tag{3}$$

We denote by μ_h the stationary measure of the discretization scheme considered. When $\theta = 0$ and $\theta = 1$ we recover the fully explicit and fully implicit Euler schemes, respectively. It is assumed that the schemes are well-defined, in that there exists a unique solution X_{n+1} to the implicit equations (2)-(3). Later in the paper on one occasion we also study a Milstein scheme, but we do not introduce it here, to keep the presentation as clear as possible.

The main difference between a standard scheme (2) and a split-step scheme (3) is the moment the 'noise part' is added: in the former case one first adds the noise and then solves the implicit equation for X_{n+1}, while in the latter case this order is reversed. The main advantage of using the split-step scheme is the simplicity, in the sense that it can be rewritten as, for functions $a : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}$,

$$X_{n+1} = a(X_n) + b(X_n)\Delta W_n. \tag{4}$$

Later in the paper we will observe that the scheme (4) is simpler to study than its standard counterpart (2). In particular, for split-step schemes it is easier to assess the existence of the stationary measure; see e.g. Section 3 of Hansen (2003) for ergodicity conditions of Markov chains with a Gaussian transition kernel. Our examples in Section 4 will reveal that (2) and (3) may lead to entirely different stationary tail distributions, which may also be very different from the tail behavior corresponding to μ_0.

2.2 Existence and Uniqueness of Stationary Distributions

In continuous time when the invariant measure μ_0 exists and has a density π_0, it is given by

$$\pi_0(x) \propto \frac{1}{g^2(x)} \exp\left\{ 2\int^x \frac{f(y)}{g^2(y)}\,\mathrm{d}y \right\}; \tag{5}$$

see e.g. Durrett (1996), p. 210. Hence it is straightforward to derive the asymptotic behavior of $\bar{\mu}_0(x)$. All SDEs considered in this paper have a stationary density proportional to the right-hand side of (5).

We now move to discrete time. We say that the Markov chain $(X_n)_{n \in \mathbb{N}}$ has a stationary distribution with law μ iff $X_0 \sim \mu$ implies $X_1 \sim \mu$ (which implies $X_n \sim \mu$ for all $n \in \mathbb{N}$). Contrary to the continuous-time case, there are no general formulas available for such distribution μ. In this paper, we assess the ergodicity of discrete-time Markov chains using Theorem 2 (introduced later, in Section 3), which is based on results from Tweedie (1983) and Meyn and Tweedie (2012).

2.3 Shape of the Tail

In this paper by the 'shape of the tail' of a random variable X we mean the rate of convergence of the complementary cumulative distribution function $\mathbb{P}(X > x)$ in x as $x \to \infty$. For example we might have *polynomial decay* (i.e., $\mathbb{P}(X > x) \sim Cx^{-\alpha}$ for some constants $C, \alpha > 0$), or *Weibullian* decay (i.e., $\mathbb{P}(X > x) \sim C\exp(-sx^p)$ for some $C, s, p > 0$).

The tails of two random variables can be compared through their moments, as follows. Let X, Y be random variables and $r : \mathbb{R} \to \mathbb{R}_+$ such that $r(x) = 0$ for $x < 0$. If $\int r(x)\mathbb{P}(X \in \mathrm{d}x) < \infty$ and $\int r(y)\mathbb{P}(Y \in \mathrm{d}y) = \infty$, then typically $\mathbb{P}(X > x)/\mathbb{P}(Y > x) \to 0$ as $x \to \infty$; we say that X has a lighter tail than Y.

If $\mathbb{P}(X > x) \sim Cx^{-\alpha}$, then α can be identified from the fractional moments $\mathbb{E}(X^\alpha; X > 0)$. More specifically,

$$\alpha = \sup\left\{ \beta > 0 : \int_0^\infty x^\beta\,\mathbb{P}(X \in \mathrm{d}x) < \infty \right\}. \tag{6}$$

Conversely, if α in (6) is finite, then $\mathbb{P}(X > x) \sim L(x)x^{-\alpha}$ for some 'sub-polynomial' function $L(\cdot)$.

If the random variable X admits all moments, so that $\alpha = \infty$ in (6), then $\mathbb{P}(X > x)$ decays to 0 faster than $x^{-\beta}$ for any $\beta > 0$. In this case, the tail can be described more precisely, via the existence of exponential moments such as $\mathbb{E}\big(\exp(sX^p); X > 0\big)$. In particular, if $\mathbb{P}(X > x) \sim C\exp(-sx^p)$, then p and s obey

$$p := \sup\left\{q > 0 : \exists t > 0 : \int_0^\infty e^{tx^q}\,\mathbb{P}(X \in dx) < \infty\right\}, \; s := \sup\left\{t > 0 : \int_0^\infty e^{tx^p}\,\mathbb{P}(X \in dx) < \infty\right\}. \quad (7)$$

Conversely, if the p, s in (7) are finite, then $\mathbb{P}(X > x) \sim L(x)e^{-sx^p}$ for a 'sub-Weibullian' function $L(\cdot)$.

For example, an exponentially distributed random variable with mean λ^{-1} has $(p,s) = (1,\lambda)$, and a normally distributed random variable with variance σ^2 has $(p,s) = (2,(2\sigma^2)^{-1})$. Both distributions are light-tailed, but the Gaussian is clearly much lighter, as it corresponds to a higher p. One could say that distributions with the same p but different s belong to the same class (but the one with the higher s is lighter).

3 TOOLS FOR THE STUDY OF THE TAILS

The behavior of $\bar{\mu}_0(x)$ (for x large) follows from the density (5). Finding the tail behavior of μ_h is less straightforward. In this section we present two tools.

3.1 Random Iterated Functions

The following result (Goldie 1991, Theorem 4.1) describes the stationary tail distribution $\bar{\mu}$ corresponding to the stochastic recursion $X_{n+1} = M_nX_n + Q_n$, where $(Q_1,M_1),(Q_2,M_2),\ldots$ are i.i.d. It will be useful for analysing the stationary laws corresponding to discretizations of linear SDEs; see Section 4.2.

Theorem 1 Let κ be such that $\mathbb{E}|M_1|^\kappa = 1$, $\mathbb{E}|M_1|^\kappa \log^+|M_1| < \infty$ and $\mathbb{E}|Q_1|^\kappa < \infty$. Assume that the law of $\log|M_1|$ given $M_1 \neq 0$ is non-arithmetic. Then the stochastic recursion $X_{n+1} = M_nX_n + Q_n$ has a unique invariant measure μ satisfying $\bar{\mu}(x) \sim Cx^{-\kappa}$, for some $C > 0$, as $x \to \infty$.

3.2 Existence and Non-Existence of Moments

In Section 2.3 we discussed the connection between existence and nonexistence of moments and the behavior of the tail. Theorem 2 presents sufficient conditions for existence (and nonexistence) of moments of the stationary measure of a Markov chain. Let $P(x,dy)$ be a transition kernel of a Markov chain and $r : \mathbb{R} \to \mathbb{R}$ be such that $r \geq 1$ and $\int r(y)P(x,dy) < \infty$. Define the following parameters associated with the function r:

$$L(r) = \limsup_{|x|\to\infty} \frac{\int r(y)P(x,dy)}{r(x)} \quad \text{and} \quad L^+(r) := \lim_{x\to\infty} \frac{\int_0^\infty r(y)P(x,dy)}{r(x)}. \quad (8)$$

Note that while L is well defined, $L^+(r)$ does not necessarily exist. Below, we use the concepts of aperiodicity, irreducibility and small sets as introduced in Meyn and Tweedie (2012).

Theorem 2 Suppose that $(X_n)_{n\in\mathbb{N}}$ is an aperiodic, μ^{Leb}-irreducible Markov chain with a transition kernel $P(x,dy)$. Suppose that all intervals $[-M,M]$ for $M > 0$ are small. Let $r : \mathbb{R} \to \mathbb{R}$ be such that $r \geq 1$, $\int r(y)P(x,dy) < \infty$ and $L(r),L^+(r)$ are defined in (8). Then

   (i)   If $L(r) < 1$ and $\sup_{x\in[-M,M]}\int r(y)P(x,dy) < \infty$ for all $M > 0$ then there exists a unique stationary probability measure μ with $\int r(x)\mu(dx) < \infty$.

  (ii)   If there exists a unique stationary probability measure μ, and r is non-decreasing for $x \in (0,\infty)$ then $L^+(r) > 2$ implies $\int_0^\infty r(x)\mu(dx) = \infty$.

Proof.   Notice that $L(r) < 1$ implies there exists $\delta, b > 0$ such that for all $x \in \mathbb{R}$,

$$\int r(y)P(x,dy) < (1-\delta)r(x) + b\mathbb{1}\{x \in [-M,M]\}.$$

Thus, part (i) follows from Thm. 14.0.1 of Meyn and Tweedie (2012). Part (ii) can be proven as follows. $L^+(r) > 2$ implies there exists $\delta, x_0 > 0$ such that $\int_0^\infty r(y)P(x, \mathrm{d}y) > (2+\delta)r(x)$ for all $x \geq x_0$. Thus,

$$r(x) + \int_x^\infty r(y)P(x, \mathrm{d}y) \geq \int_0^x r(y)P(x, \mathrm{d}y) + \int_x^\infty r(y)P(x, \mathrm{d}y) = \int_0^\infty r(y)P(x, \mathrm{d}y), \qquad (9)$$

for all $x \geq x_0$, since r is non-decreasing on $(0, \infty)$ and $\int P(x, \mathrm{d}y) = 1$. Combining these elements we find

$$\int_{x_0}^\infty r(y)P(x, \mathrm{d}y) \geq \int_x^\infty r(y)P(x, \mathrm{d}y) > (1+\delta)r(x),$$

for all $x \geq x_0$. Then, with the second equality due to 'Fubini',

$$\int_{x_0}^\infty r(x)\mu(\mathrm{d}x) = \int_{x_0}^\infty r(x)\left(\int \mu(\mathrm{d}y)P(y, \mathrm{d}x)\right) = \int\left(\int_{x_0}^\infty r(x)P(y, \mathrm{d}x)\right)\mu(\mathrm{d}y)$$

$$\geq \int_{x_0}^\infty \left(\int_{x_0}^\infty r(x)P(y, \mathrm{d}x)\right)\mu(\mathrm{d}y) > (1+\delta)\int_{x_0}^\infty r(y)\mu(\mathrm{d}y).$$

This entails $\int_0^\infty r(x)\mu(\mathrm{d}x) = \infty$. $\qquad\square$

Remark 1 Additionally, there is a very simple condition which implies nonexistence of a moment. If there exists a set A such that $\mu(A) > 0$ and $\int r(x)P(x, \mathrm{d}y) = \infty$ for all $x \in A$ then necessarily $\int r(x)\mu(\mathrm{d}x) = \infty$.

In view of the considerations in Section 2.3, it is particularly convenient to work with Thm. 2 for classes of functions such as $r_\alpha(x) = 1 + |x|^\alpha$ and $r_{p,s}(x) = \exp(s|x|^p)$. Note however, that there is a 'moment gap', that is when $L(r) > 1$ and $L^+(r) < 2$, it cannot be inferred from Theorem 2 whether the moment $\int_0^\infty r(x)\mu(\mathrm{d}x)$ exists or not. The gap could be decreased by considering bounds tighter than (9) based on the transition kernel.

4 ASSESSMENT OF THE TAIL IN BENCHMARK MODELS

In this section we study four prototypical SDEs, comparing the tails of the continuous-time processes with the tails of their discrete-time counterparts. Thus, we compare the tails $\bar\mu_0(x)$ and $\bar\mu_h(x)$ for large x. We do so by studying the ratio

$$\gamma_h(x) := \bar\mu_h(x)/\bar\mu_0(x),$$

for large x, both analytically (using the tools presented in Section 3) and numerically. In particular, if $\gamma_h(x) \to 0$, or $\gamma_h(x) \to \infty$, as $x \to \infty$ then μ_h has a tail lighter, or heavier than μ_0 respectively.

The main message from this section is that $\gamma_h(x)$ can differ substantially from 1. In addition, the $\bar\mu_h(x)$ resulting from different discretization schemes can also be orders of magnitude different. This leads to the general conclusion that, when aiming at estimating the SDE's tail distribution, one has to be very careful with using discretization schemes. We focus in this section on one-dimensional SDEs, as for these μ_0 is available in closed-form, but obviously the above warning carries over to multi-dimensional SDEs.

The numerical results in Sections 4.2-4.4 concern the estimation of $\bar\mu_h(x)$ for various values of x and h. In all sections, the estimates were determined by a crude Monte Carlo method. The Markov chain has been divided in blocks of 10^4 time steps (of size h) separated by periods of 10^4 time steps. These blocks of 10^4 time steps are assumed to be independent, and based on that assumption we derive the sample error. The estimation procedure is stopped when the sample relative error falls below 5%.

4.1 Ornstein-Uhlenbeck Process

It was already observed in e.g. Talay (1990) that the stationary distribution of an OU process is affected by discretization (motivating the use of a second-order scheme that has a stationary distribution closer to
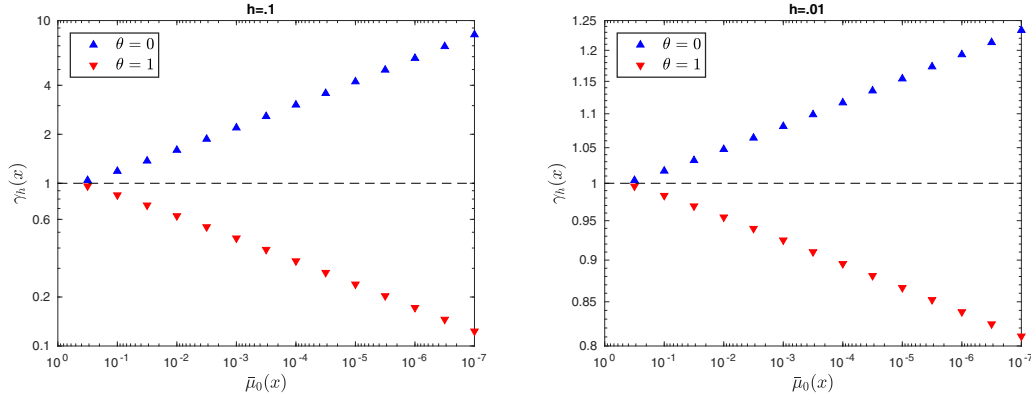
Figure 1: Ornstein-Uhlenbeck process with $\vartheta = 3, \sigma = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.1$ (left) and $h = 0.01$ (right) for the two extreme values of the 'implicitness parameter' θ.

the continuous-time one). Schurz (1999) proposes discretizing multidimensional linear SDEs (covering the OU case) using the implicit Euler methods; he shows that the trapezoidal rule (semi-implicit scheme with $\theta = \frac{1}{2}$) is the only discretization scheme in that family that preserves the correct stationary distribution.

The OU process is defined by the following SDE: for $\vartheta, \sigma > 0$,

$$dX_t = -\vartheta X_t \, dt + \sigma \, dW_t. \tag{10}$$

It is well-known that (10) admits a Gaussian stationary distribution with mean zero and variance $\Sigma := \sigma^2/(2\vartheta)$. The θ-implicit (standard) Euler scheme (2) yields

$$X_{n+1} = X_n \left(1 - \vartheta \Gamma h\right) + \sigma \Gamma \sqrt{h} \Delta W_n,$$

with $\Gamma := (1 + \vartheta \theta h)^{-1}$. This is an AR(1) process, which admits a stationary distribution iff $|1 - \vartheta \Gamma h| < 1$, that is when $h < 2/(\vartheta(1 - 2\theta))$; see e.g. Van der Vaart (2010). More specifically, this stationary distribution is Gaussian with mean zero and variance $\Sigma(1 + \vartheta h(\theta - \frac{1}{2}))^{-1}$. This shows that by taking $\theta = \frac{1}{2}$, the invariant measure of the discretized chain is the same as the invariant measure of the original, continuous-time one, cf. the findings of Schurz (1999).

We take the analysis a bit further to assess what errors one makes for the OU case with other rules than the trapezoid rule. A first, naïve, choice could be the explicit Euler scheme ($\theta = 0$). As seen above, $\mu_0 \sim N(0, \Sigma)$ and $\mu_h \sim N(0, \Sigma/(1 - \frac{1}{2}\vartheta h))$. Now consider the ratio $\gamma_h(x)$. Let $\Phi(\cdot)$ and $\varphi(\cdot)$ denote a standard normal cdf and pdf respectively. For large x, $\Phi(-x) \sim \varphi(x)/x$, so that for large x

$$\gamma_h(x) = \frac{\Phi\left(-\frac{x}{\sqrt{\Sigma}}\sqrt{1 - \frac{1}{2}\vartheta h}\right)}{\Phi\left(-\frac{x}{\sqrt{\Sigma}}\right)} \sim \sqrt{\frac{2}{2 - \vartheta h}} \exp\left\{-\frac{x^2 \vartheta}{2\sigma^2}(2 - \vartheta h - 2)\right\} = \sqrt{\frac{2}{2 - \vartheta h}} \cdot \exp\left\{\frac{\vartheta^2 x^2 h}{2\sigma^2}\right\}.$$

This means that $\gamma_h(x) \to \infty$ for any given $h > 0$, showing that the tail of μ_h is heavier than the tail of μ_0. Similar calculations show that in the fully implicit case ($\theta = 1$) the tail of μ_h is lighter than the one of μ_0. In general, increasing the 'implicitness parameter' θ makes the tail of μ_h lighter; cf. Figure 1.

A pragmatic rule that can be used to control $\gamma_h(x)$ is to decrease the step h proportionally to x^2. More, concretely, let h satisfy $x^2 h \to \varepsilon > 0$, as x grows large. Then $\gamma_h(x) \approx 1 + \varepsilon \vartheta^2/(2\sigma^2)$. So in order to keep the ratio $\gamma_h(x)$ smaller than some given $1 + \delta$, one could choose h in the following way:

$$h = x^{-2}\left(2\vartheta^{-2}\sigma^2 \log(1 + \delta)\right).$$

Conclusions: This example shows that, although μ_h still belongs to the same class of distributions as μ_0 (Weibull-tail with shape parameter 2), the ratio $\gamma_h(x)$ can deviate substantially from 1 (in fact, $\gamma_h(x)$ may even go to 0 or ∞, depending on the choice of the implicitness parameter). This effect can be mitigated by choosing the step-size h sufficiently small. The tail of μ_h becomes lighter when increasing θ.

4.2 Linear Drift and Linear Volatility

We now consider the process $(X_t)_{t \geq 0}$ that solves the SDE

$$\mathrm{d}X_t = \vartheta(m - X_t)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t\,,$$

where $\sigma, \vartheta > 0$ and $m > 0$ (observe that if m would have been 0, then the stationary measure μ_0 is entirely concentrated at 0). Using (5) we see that μ_0 corresponds to an *Inverse-Gamma* distribution with density

$$\pi_0(x) = \beta^{\alpha_0}(\Gamma(\alpha_0))^{-1}\,x^{-(1+\alpha_0)}\,\mathrm{e}^{-\beta/x}, \quad x > 0,$$

with shape parameter $\alpha_0 = 1 + 2\vartheta/\sigma^2$ and rate parameter $\beta = 2\vartheta m/\sigma^2$. As a consequence, for large x, $\pi_0(x) \sim (\beta^{\alpha_0}/\Gamma(\alpha_0))x^{-(1+\alpha_0)}$ and hence $\bar{\mu}_0(x) \sim C_0 x^{-\alpha_0}$ for some $C_0 > 0$.

Now let us focus on the stationary measure μ_h corresponding to the semi-implicit (standard) Euler scheme for various values of θ. The scheme (2) yields

$$X_{n+1} = \frac{\vartheta m h}{1 + \vartheta\theta h} + X_n\left(1 - \frac{\vartheta h}{1 + \vartheta\theta h} + \frac{\sigma\sqrt{h}}{1 + \vartheta\theta h}\Delta W_n\right). \tag{11}$$

Notice that this is exactly the setting of Thm. 1, so it can be applied to derive the tail behavior of μ_h. The following proposition, which is an immediate consequence of Thm. 1, states that $\bar{\mu}_h(x)$ decays polynomially. The corresponding rate α_h (which differs from α_0) solves

$$\mathbb{E}\left|1 - \frac{\vartheta h}{1 + \vartheta\theta h} + \frac{\sigma\sqrt{h}}{1 + \vartheta\theta h}\Delta W_1\right|^{\alpha_h} = 1.$$

Proposition 1 The family of Markov Chains evolving according to (11) admits a stationary distribution μ_h for all $h > 0$ and for each fixed h, which satisfies $\bar{\mu}_h(x) \sim C_h x^{-\alpha_h}$, as $x \to \infty$, for some constant $C_h > 0$.

We now analyze α_h as $h \downarrow 0$, again using Thm. 1. Using Taylor expansions and taking the expectation (recalling that $\Delta W_1 \sim N(0,1)$), we obtain

$$\mathbb{E}\left|1 - \frac{\vartheta h}{1 + \vartheta\theta h} + \frac{\sigma\sqrt{h}}{1 + \vartheta\theta h}\Delta W_1\right|^{\alpha} = 1 + \left(-\alpha\vartheta + \binom{\alpha}{2}\sigma^2\right)h$$
$$+ \left(\alpha\theta\vartheta^2 - 3\binom{\alpha}{3}\sigma^2\vartheta + 3\binom{\alpha}{4} + \binom{\alpha}{2}(\vartheta^2 - 2\theta\sigma^2\vartheta x^2)\right)h^2 + o(h^2),$$

which is to be equated to 1. Now it can be verified that α_h has the following first-order Taylor expansion around $h = 0$: with α_0 the rate corresponding to $\bar{\mu}_0(x)$,

$$\alpha_h = \alpha_0 + \left(\vartheta^2(1 + 2\theta)\sigma^{-2} - \vartheta\right)h + o(h). \tag{12}$$

Remark 2 An expansion similar to (12) can be obtained for other discretization schemes as well, for instance for the semi-implicit *Milstein scheme* given by

$$X_{n+1} = X_n + \left(\theta f(X_{n+1}) + (1 - \theta)f(X_n)\right)h + g(X_n)\sqrt{h}\,\Delta W_n + \frac{1}{2}g(X_n)g'(X_n)h((\Delta W_n)^2 - 1)\,.$$
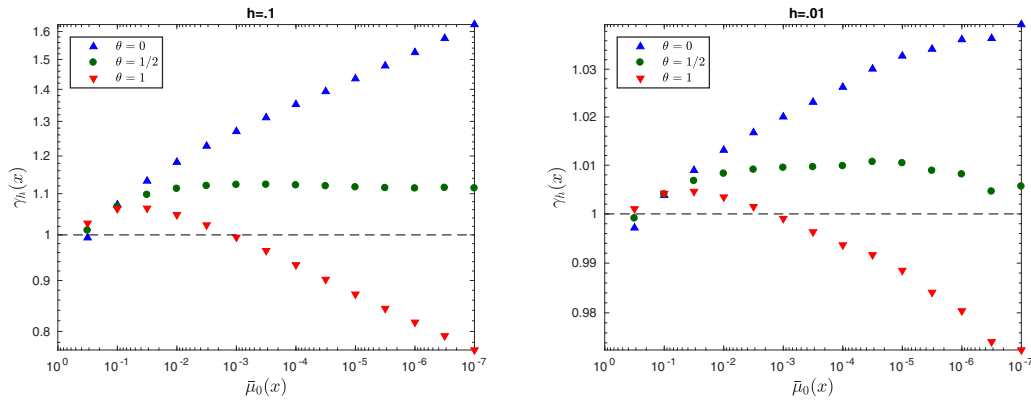
Figure 2: Linear drift and linear volatility process with $\vartheta = 1$, $\sigma = \sqrt{2}$, $m = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.1$ (left) and $h = 0.01$ (right) for $\theta \in \{0, \frac{1}{2}, 1\}$. From (12), the tail of μ_h should be heavier than the tail of μ_0 if $\theta = 0$, and lighter if $\theta = 1$. For $\theta = \frac{1}{2}$ the tails are more similar, as $\alpha_h = \alpha_0 + o(h)$.

This scheme takes the explicit form

$$X_{n+1} = \frac{\vartheta \mu h}{1 + \vartheta \theta h} + X_n \left(1 - \frac{\vartheta + \frac{1}{2}\sigma^2}{1 + \vartheta \theta h} h + \frac{\sigma}{1 + \vartheta \theta h} \sqrt{h} \Delta W_n + \frac{\frac{1}{2}\sigma^2}{1 + \vartheta \theta h} h (\Delta W_n)^2 \right).$$

Without presenting the computations, we claim that the Milstein-discretized stationary distribution is heavy tailed: $\bar{\mu}_h(x) \sim C_h x^{-\alpha_h}$. The power α_h has, as $h \downarrow 0$, the first-order Taylor expansion

$$\alpha_h = \alpha_0 + \left(\vartheta^2 (2\theta - 3)\sigma^{-2} \right) h + o(h).$$

The reasoning is as in the proof of Prop. 1 and the computation of α_h for the schemes (2)-(3). ⬦

The expression (12) for α_h for $\theta = 0$ reveals that if $\vartheta/\sigma^2 > 1$ then the tail of μ_h is lighter than the tail of μ_0, but vice versa if $\vartheta/\sigma^2 < 1$ — a perhaps surprising result for an explicit Euler scheme. The expansion of α_h in (12) advocates choosing $\theta \in [0,1]$ such that the absolute value of $(\vartheta(1+2\theta)/\sigma^2 - 1)$ is minimized. See Figure 2 for illustrations.

Conclusions: In this example, the tail of μ_h and μ_0 are in the same class (power-law decay), and again the ratio $\gamma_h(x)$ can differ substantially from 1. Choosing $\theta = 0$ now does not necessarily lead to the situation of the tail of μ_h being heavier than the tail of μ_0 (cf. the OU process in Section 4.1). The tail of μ_h becomes lighter when increasing θ (also in the Milstein case).

4.3 Cubic Drift and Constant Volatility

We now consider a stochastic process with nonlinear drift, more specifically $(X_t)_{t \geq 0}$ solving

$$\mathrm{d}X_t = -\vartheta X_t^3 \, \mathrm{d}t + \sigma \, \mathrm{d}W_t, \tag{13}$$

where $\vartheta, \sigma > 0$. From the formula for the density of the invariant measure (5), we see that μ_0 has a centered *Generalized Normal Distribution* with the density

$$\pi_0(x) = \beta \left(2\alpha \Gamma(1/\beta) \right)^{-1} \mathrm{e}^{-(|x|/\alpha)^\beta}, \tag{14}$$

with scale parameter $\alpha = \left(2\sigma^2/\vartheta \right)^{1/4}$ and shape parameter $\beta = 4$. It has been shown by Mattingly et al. (2002) that the explicit Euler scheme (2) of the process driven by (13) is transient, i.e., does not converge to a stationary distribution. Denote

$$a(x) := c^{-1} \sinh \left(\operatorname{arcsinh}(3cx)/3 \right), \quad F_\theta(x) := x - \vartheta(1-\theta)x^3 h, \quad \text{and} \quad c := \sqrt{3\vartheta\theta h}/2. \tag{15}$$

We can show that the θ-implicit standard and split-step Euler schemes (2)-(3) read, for $\theta \in (0,1]$,

$$X_{n+1} = a\big(F_\theta(X_n) + \sigma\sqrt{h}\,\Delta W_n\big) \tag{16}$$

$$X_{n+1} = a\big(F_\theta(X_n)\big) + \sigma\sqrt{h}\,\Delta W_n \tag{17}$$

respectively. To study the ergodicity of the Markov chains driven by (16) and (17), it is useful to derive the asymptotics of $a(\cdot)$. The function $a(\cdot)$ is odd, satisfies $|a(x)| \leq x^{1/3}(\vartheta\theta h)^{-1/3}$ for all $x \in \mathbb{R}$, and in addition $a(x)x^{-1/3} = (\vartheta\theta h)^{-1/3}$ as $|x| \to \infty$. Both schemes are ergodic for $\theta > \frac{1}{2}$. For the split-step scheme, ergodicity follows from (Hansen 2003) and the asymptotics of the function $a(\cdot)$ listed earlier. To show the ergodicity of the standard Euler scheme one can use Thm. 2(i). We note that the chain (16) satisfies the conditions of irreducibility, small sets and aperiodicity as stated in Theorem 2. Let $r_{p,s}(x) := \exp(s|x|^p)$ and denote $Z \sim N(0,1)$. We verify the remaining assumptions of Thm. 2(i). Based on the properties of the function $a(\cdot)$, it is straightforward to verify that when $\theta = 1$,

$$\int r_{p,s}(y)P(x,\mathrm{d}y) = \mathbb{E}\exp\big(s\,|a(x+\sigma\sqrt{h}Z)|^p\big) \leq \mathbb{E}\exp\Big(s(\vartheta h)^{-p/3} \cdot \big|x+\sigma\sqrt{h}Z\big|^{p/3}\Big). \tag{18}$$

Now let $p = 6$. When $s < \vartheta^2 h/(2\sigma^2)$, the expression on the rhs of (18) is finite and there exist constants $A, B > 0$ such that it is equal to $A\exp(Bx^2)$. Thus, $L(r_{p,s}) = 0$, where L is defined in (8). Lastly, as

$$\sup_{x\in[-M,M]} \int r_{p,s}(y)P(x,\mathrm{d}y) \leq A\exp(BM^2) < \infty,$$

we conclude that when $\theta = 1$, the standard Euler scheme is ergodic with stationary measure μ_h and moreover, $r_{p,s}(x)\mu_h(\mathrm{d}x) < \infty$ for $p = 6$ and $s < \vartheta^2 h/(2\sigma^2)$. Proving ergodicity in case $\theta \in (1/2, 1)$ can be dealt with similarly using the test function $r(x) = 1 + x^6$.

Our next objective is to compare the tails of μ_h and μ_0 for a fully implicit scheme ($\theta = 1$). From (14), μ_0 has Weibullian decay with $(p,s) = (4, \vartheta/(2\sigma^2))$. For the standard Euler scheme μ_h is characterized by $(p,s) = (6, \vartheta^2 h/(2\sigma^2))$. This follows from the fact that $\int_0^\infty r_{6,s}(y)P(x,\mathrm{d}y) = \infty$ for $s < \vartheta^2 h/(2\sigma^2)$ (cf. Remark 1). The latter can be established using asymptotics of the function $a(\cdot)$, similar to the way we established the existence of moments. For the split-step scheme μ_h has parameters $(p,s) = (2, 1/(2\sigma^2 h))$, as can be shown in an analogous fashion.

The implication is that the standard Euler scheme gives a distribution in a lighter class than μ_0. It becomes heavier as $h \downarrow 0$ (since $s \downarrow 0$), but remains in the class with $p = 6$ (compared to $p = 4$ for μ_0). By contrast, the split-step scheme gives a distribution in a heavier class (viz. with $p = 2$), but its tail becomes lighter as $h \downarrow 0$. This behavior is reflected in our simulation results, see Figure 3. We observe that the ratio $\gamma_h(x)$ goes to 0 as $x \to \infty$ for the standard Euler scheme, whereas $\gamma_h(x)$ explodes for the split-step scheme.

Conclusions: This example shows that μ_0 can be in a different class of distributions than μ_h. In addition, it strongly depends on the discretization scheme in which class μ_h is.

4.4 Cubic Drift and Linear Volatility

In our last example $(X_t)_{t\geq 0}$ solves

$$\mathrm{d}X_t = \vartheta(m - X_t^3)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t,$$

where $\sigma, \vartheta > 0$ and $m > 0$. Like in the previous examples, we can find the density of μ_0:

$$\pi_0(x) \propto x^{-2}\exp\Big\{-\tfrac{\vartheta}{\sigma^2}\big(\tfrac{2m}{x} + x^2\big)\Big\}, \quad x > 0.$$

Let $a(\cdot)$ and $F_\theta(\cdot)$ be as defined in (15). Similar to the example from the previous section, for $\theta \in (0,1]$ the θ-implicit Euler scheme and the θ-implicit split-step scheme evolve according to

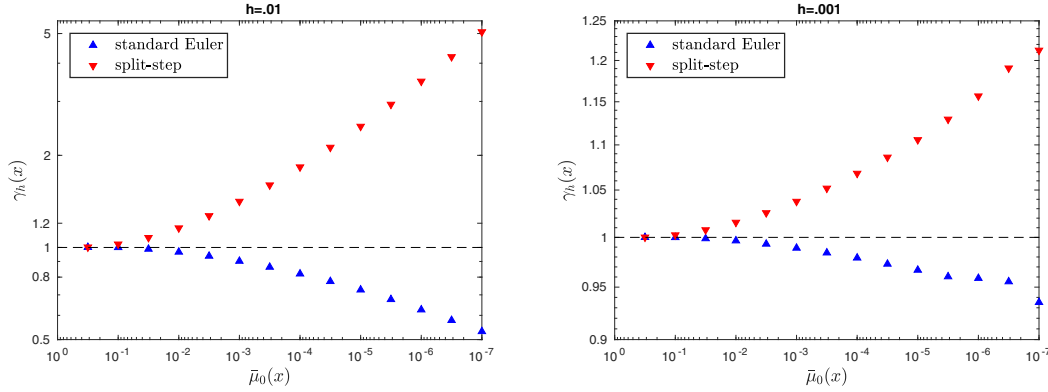$$X_{n+1} = a\big(\vartheta mh + F_\theta(X_n) + \sigma X_n\sqrt{h}\,\Delta W_n\big),$$

Figure 3: Cubic drift and constant volatility process with $\vartheta = 1, \sigma = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.01$ (left) and $h = 0.001$ (right) for fully implicit ($\theta = 1$) standard and split-step Euler schemes.

$$X_{n+1} = a\big(\vartheta mh + F_\theta(X_n)\big) + \sigma X_n \sqrt{h} \Delta W_n,$$

respectively. Both schemes are ergodic for $\theta > 1/2$. We do not prove the regularity conditions (i.e. irreducibility, small sets, aperiodicity) of Thm. 2 here. Furthermore we focus on the fully implicit scheme $\theta = 1$; the case $\theta \in (1/2, 1)$ can be dealt with similarly.

Standard Euler case. We show, using Thm. 2(i), that the fully implicit standard Euler scheme has a stationary measure with Weibullian decay with $p \in [3,6]$. Let $r_{p,s}(x) := \exp(s|x|^p)$. We have

$$\int r_{p,s}(y)P(x, \mathrm{d}y) = \mathbb{E}\exp\left(s|a(\vartheta mh + x + \sigma x\sqrt{h}Z)|^p\right) \leq \mathbb{E}\exp\left(s(\vartheta h)^{-p/3} \cdot \left|\vartheta mh + x + \sigma x\sqrt{h}Z\right|^{p/3}\right),$$

with $Z \sim N(0,1)$. We see that for $p = 3$ the expression on the rhs can be bounded by $A\exp(Bx^2)$ for some $A, B > 0$. Hence $L(r_{3,s}) = 0$ for all $s > 0$. In addition,

$$\sup_{x \in [-M,M]} \int r_{3,s}(y)P(x, \mathrm{d}y) < A\exp(BM^2) < \infty.$$

Thus, based on Thm. 2(i), we can conclude that the chain is ergodic with stationary probability measure μ_h and admits exponential moments $\int r_{3,s}(x)\mu_h(\mathrm{d}x) < \infty$. Furthermore, applying Remark 1 to $r_{6,s}(\cdot)$, we see that $\int_0^\infty r_{6,s}(y)P(x, \mathrm{d}y) = \infty$ for x large enough and thus $\int_0^\infty r_{6,s}(x)\mu_h(\mathrm{d}x) = \infty$.

Split-step Euler case. We show that the fully implicit split-step Euler scheme has a stationary measure with polynomial decay. Let $r_\alpha(x) := 1 \vee |x|^\alpha$ and denote $Z \sim N(0,1)$. We have for $|x| > 1$:

$$\frac{\int r_\alpha(y)P(x, \mathrm{d}y)}{r_\alpha(x)} = \mathbb{E}\left|\frac{a(\vartheta mh + x)}{x} + \sigma\sqrt{h}Z\right|^\alpha \xrightarrow{|x|\to\infty} (\sigma^2 h)^{\alpha/2} \cdot \mathbb{E}|Z|^\alpha = L(r_\alpha),$$

by 'dominated convergence'. Moreover, one can show that $\sup_{x \in [-M,M]} \int r_\alpha(y)P(x, \mathrm{d}y) < \infty$. Using Thm. 2(i) we conclude that the chain is ergodic with stationary probability measure μ_h and has polynomial moments $\int r_\alpha(x)\mu_h(\mathrm{d}x) < \infty$ for $\alpha < \alpha_h$, where α_h solves $(\sigma^2 h)^{\alpha_h/2} \cdot \mathbb{E}|Z|^{\alpha_h} = 1$. The fractional moments are available in closed form: $\mathbb{E}|Z|^\alpha = 2^{\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$ (Winkelbauer 2012). Hence, α_h can be found numerically. For instance, when $\sigma = 1$ (where ϑ and m are irrelevant) $\alpha_1 = 2$ and $\alpha_{1/2} \approx 4.75$. Conversely, as $x \to \infty$

$$\frac{\int_0^\infty r_\alpha(y)P(x, \mathrm{d}y)}{r_\alpha(x)} = \mathbb{E}\left(\left|\frac{a(\vartheta mh + x)}{x} + \sigma\sqrt{h}Z\right|^\alpha; \frac{a(\vartheta mh + x)}{x} + \sigma\sqrt{h}Z > 0\right) \to (\sigma^2 h)^{\alpha/2}\mathbb{E}(Z^\alpha; Z > 0).$$
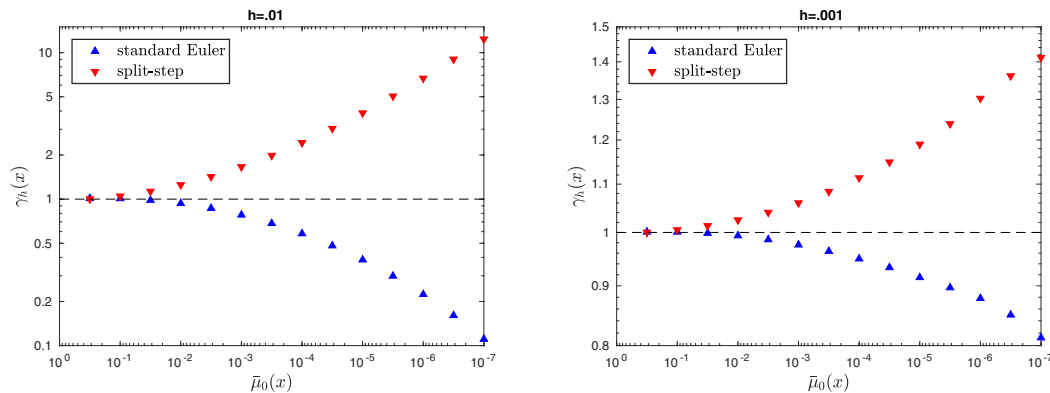
Figure 4: Cubic drift and linear volatility process with $\vartheta = 1, \sigma = 1, m = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.01$ (left) and $h = 0.001$ (right) for fully implicit ($\theta = 1$) standard and split-step Euler schemes.

So $L^+(r_\alpha) = \frac{1}{2}(\sigma^2 h)^{\alpha/2}\mathbb{E}|Z|^\alpha$, which diverges to ∞, as $\alpha \to \infty$. According to Thm. 2(ii), $L^+(r_\alpha) > 2$ implies $\int_0^\infty r(x)\mu_h(\mathrm{d}x) = \infty$. From this we conclude that the tail of μ_h has a power law. Notably, $\alpha_h \to \infty$ as $h \downarrow 0$, so the tail indeed gets lighter as $h \downarrow 0$. Observe however that for all positive h, the invariant measure μ_h has still only finitely many moments, whereas μ_0 decays in a Weibullian way.

Conclusions: In this example μ_0 corresponds to Weibullian decay with $p = 2$. The fully implicit standard and split-step Euler schemes exhibit completely different behavior. The standard Euler scheme gives rise to Weibullian decay with some $p \in [3,6]$, and the split-step scheme (strikingly) results in power-law decay. The tails become lighter in the split-step scheme as $h \downarrow 0$. Experimental results are presented in Figure 4.

5   CONCLUSIONS

In this paper we have studied simulation-based techniques for estimating the stationary distribution of SDEs, with a focus on the tail distribution. More specifically, we have considered how the estimate of the stationary tail distribution is affected by the time-discretization scheme chosen. The main conclusion is that estimation of the stationary tail is a highly delicate issue. Commonly accepted discretization techniques may very well lead to highly inaccurate estimates of the stationary tail distribution of the continuous-time process. For instance, we have identified cases in which the stationary tail of the SDE is Weibullian, while the discretizations lead to power-law tails. In our study we have focused on discretizations of one-dimensional SDEs, as for those the solution of the SDE is available in closed form and can be used as a benchmark; evidently, the above warning carries over to multi-dimensional SDEs.

Given these findings, it is advisable to use different discretization schemes (e.g. θ-implicit standard and split-step Euler with different θ and higher order schemes) for assessing tail behavior, and compare their outcomes. Furthermore, we have shown in the first example that the error can be reduced by adapting the stepsize h to the threshold x. For the OU case, we were able to show how to choose h to control the error. Similar results may be possible for other cases, but require further study.

REFERENCES

Abdulle, A., G. Vilmart, and K. C. Zygalakis. 2014. "High Order Numerical Approximation of the Invariant Measure of Ergodic SDEs". *SIAM Journal on Numerical Analysis* 52(4):1600–1622.

Durrett, R. 1996. *Stochastic Calculus: a Practical Introduction*. CRC Press.

Goldie, C. M. 1991. "Implicit Renewal Theory and Tails of Solutions of Random Equations". *The Annals of Applied Probability* 1(1):126–166.

Hansen, N. R. 2003. "Geometric Ergodicity of Discrete-time Approximations to Multivariate Diffusions". *Bernoulli* 9(4):725–743.

Kloeden, P. E., and E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin.

Majda, A. J., C. Franzke, and D. Crommelin. 2009. "Normal Forms for Reduced Stochastic Climate Models". *Proceedings of the National Academy of Sciences* 106(10):3649–3653.

Mattingly, J. C., A. M. Stuart, and D. J. Higham. 2002. "Ergodicity for SDEs and Approximations: Locally Lipschitz Vector Fields and Degenerate Noise". *Stochastic Processes and their Applications* 101(2):185–232.

Meyn, S. P., and R. L. Tweedie. 2012. *Markov Chains and Stochastic Stability*. Springer, London.

Roberts, G. O., and R. L. Tweedie. 1996. "Exponential Convergence of Langevin Distributions and Their Discrete Approximations". *Bernoulli* 2(4):341–363.

Schurz, H. 1999. "The Invariance of Asymptotic Laws of Stochastic Systems under Discretization". *ZAMM-Zeitschrift fur Angewandte Mathematik und Mechanik* 79(6):375–382.

Stramer, O., and R. Tweedie. 1999. "Langevin-type Models I: Diffusions with Given Stationary Distributions and Their Discretizations". *Methodology and Computing in Applied Probability* 1(3):283–306.

Talay, D. 1990. "Second-order Discretization Schemes of Stochastic Differential Systems for the Computation of the Invariant Law". *Stochastics: An International Journal of Probability and Stochastic Processes* 29(1):13–36.

Tweedie, R. L. 1983. "The Existence of Moments for Stationary Markov Chains". *Journal of Applied Probability* 20(1):191–196.

Van der Vaart, A. W. 2010. "Time Series". http://www.math.leidenuniv.nl/~avdvaart/timeseries/dictaat.pdf.

Winkelbauer, A. 2012. "Moments and Absolute Moments of the Normal Distribution". arXiv:1209.4340.

Yang, S., J. N. Onuchic, and H. Levine. 2006. "Effective Stochastic Dynamics on a Protein Folding Energy Landscape". *The Journal of Chemical Physics* 125(5):054910.

AUTHOR BIOGRAPHIES

KRZYSZTOF BISEWSKI is a PhD student at CWI, Amsterdam affiliated with University of Amsterdam. He obtained a double master degree in mathematics from University of Warsaw and Vrije Universiteit. His interests are in theoretical and applied probability, efficient simulation and much more specifically, in the effect of discretization on the assessment of rare event probabilities. His e-mail address is K.L.Bisewski@cwi.nl.

DAAN CROMMELIN is leader of the Scientific Computing group at CWI, Amsterdam, and a part-time full professor in applied analysis at the University of Amsterdam. Before he was with the Courant Institute of Mathematical Sciences of New York University. His research focuses on stochastic and computational methods for multi-scale dynamical systems, uncertainty quantification, rare event simulation, and applications in climate science and energy networks. His email address is Daan.Crommelin@cwi.nl.

MICHEL MANDJES is a full professor of applied probability at the University of Amsterdam. Before he worked as a member of technical staff at Bell Labs, Murray Hill NJ (US), as a full professor at the University of Twente (Netherlands), and as a group leader at CWI, Amsterdam. In 2008 he was visiting professor at Stanford University, and in 2013-2014 at New York University. Mandjes' research focuses on various aspects of stochastic processes, including efficient simulation techniques, queueing theory, asymptotic techniques, and large deviations. His email address is M.R.H.Mandjes@uva.nl.