Transient dynamics of Pearson diffusions facilitates estimation of rate parameters

Ditlevsen, Susanne; Cencerrado Rubio, Alejandro; Lansky, Petr

Published in:
Communications in Nonlinear Science and Numerical Simulation

DOI:
10.1016/j.cnsns.2019.105034

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Commun Nonlinear Sci Numer Simulat 82 (2020) 105034

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

Research paper

Transient dynamics of Pearson diffusions facilitates estimation of rate parameters

Susanne Ditlevsen[a,*], Alejandro Cencerrado Rubio[a], Petr Lansky[b]

[a] Department of Mathematical Sciences, University of Copenhagen, KøbenhavnØ, Denmark
[b] Institute of Physiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

ARTICLE INFO

Article history:
Available online 11 October 2019

Keywords:
Estimation of rate parameters
Pearson diffusions
Neuronal models
Perturbation

ABSTRACT

Estimation of parameters in stochastic processes has been thoroughly investigated for decades and the asymptotic properties of the estimators are known. However, reaching the regime where the asymptotic properties are valid might require such a long time that the well based theoretical results have no practical value. One example of this situation is at the center of our interest. It concerns determination of the time constant in stochastic Langevin equations with additive or multiplicative white-noise terms. Often the number of observations to achieve the asymptotic conditions is beyond the physical limits. Here we show how to overcome the problem by external perturbation of the system. Furthermore, we show that the perturbation is not at the price of deterioration of the estimates of other parameters unless the observation interval is very short compared to the typical time constant of the system. Three processes from the class of Pearson diffusions are studied. They are frequently used in many applications, in particular, they are examples of leaky integrate-and-fire models, which describe the electrical properties of a neuronal membrane. These neuronal models are often used as examples of systems with excitable dynamics. The most commonly investigated process is the Ornstein-Uhlenbeck process, which has additive noise. Furthermore, the square-root process and the Jacobi process are examples of processes with multiplicative noise. The results are illustrated on computer experiments, which show a striking improvement of the estimates of the rate parameter. It has implications for experimental design, where the information about the parameters can be increased for the same amount and cost of data, which is particularly important when samples are expensive or difficult to obtain.

1. Introduction

Computational and mathematical models of neuronal behavior are frequently appearing in studies of nonlinear dynamics and have attracted the attention of researchers for decades. Among these, the models of single-cell excitability, ranging from complex Hodgkin-Huxley models to the simplest McCulloch-Pitts neurons, are again and again rediscovered and investigated. The large variability of experimental data obtained under identical conditions calls for extending the deterministic models to

* Corresponding author.
E-mail address: susanne@math.ku.dk (S. Ditlevsen).

https://doi.org/10.1016/j.cnsns.2019.105034

their stochastic variants. Furthermore, the necessity of stochastic descriptions of neurons from the principles of coding and decoding of information within the nervous system supports the application of stochastic methods in computational neuroscience. However, not only building and investigating new models are important, but their verification and comparison with experimental data play an equally important role. Therefore, methods of statistical inference have to be carefully developed [17,25,33]. We investigate three variants of a Langevin equation (the leaky integrate-and-fire (LIF) model) frequently used as a simple model of neurons. The variants differ only by the impact of the noise on the dynamics of the membrane potential. This neuronal model is among the most frequent examples of excitable systems. In the review on effects of noise in excitable systems by Lindner et al. [27], the FitzHugh-Nagumo and LIF models are used as prototypes of excitable stochastic dynamics. The same models are used in studies on stochastic and coherence resonance in excitable systems in biology [28,30,39]. For an intuitive overview of the model investigated in this paper, see Ditlevsen and Samson [9].

The methods devised for statistical analysis of experimental data often requires asymptotic conditions such as the number of observations tending to infinity or the sample step tending to zero. In other words, the amount of data has to be large enough for the results to be trusted, but rarely any evaluation of the robustness is done of whether this is fulfilled in a specific data set. Even when applying the best available estimator, such as the maximum likelihood estimator (MLE), it might easily fail to provide trustworthy estimates of the quantity of interest because of insufficient data. We emphasize that it is not the statistical method that fails; in some standard settings, the data simply do not contain enough information about the parameter to be reliably estimated. The key idea presented in this paper is to sample the data more cleverly, such that for the same number of samples, the information about the parameter is increased.

Studying any dynamical system, it is of primary interest to estimate the time constant or relaxation time, or equivalently, the rate of change of the process. This is a particularly important parameter in neuronal models [22], where it represents the conductance, i.e., in LIF models it is the leakage conductance. Let β denote this parameter. As we will show, β is particularly difficult to estimate in real situations since the conditions required for the statistical analysis are often unrealistic. The estimates suffer from large bias and variance. Methods of bias correction have been proposed [29,37,40,41], however, these methods are limited to a specific model, and they only alliviate the problem of the bias, not of the large variance. In this paper we extend the results from [8], and show how an alternative method can solve the problem in cases where the process can be manipulated or the data can be judiciously selected. A related problem was considered in [18,19,38], where an input was adaptively designed and injected into the process to drive the drift such as to optimize the information about the parameters.

The goal is to develop a reliable framework for parameter estimation in a flexible class of stochastic processes called the Pearson diffusions [13]. Several stochastic neuronal models, with additive as well as multiplicative noise, can be seen as special cases of Pearson diffusions in the presence of an absorbing boundary, which serves as a firing threshold and introduces a strong nonlinearity into the model [2,16,32]. The Ornstein-Uhlenbeck model is probably the most common neuronal model with additive noise [15], whereas the Feller, Jacobi or Inhomogeneous Geometric Brownian Motion are models with multiplicative noise [7,12,20,23] and are closer to conductance-based neuronal models. The Pearson diffusions cover all standard distributions as invariant (stationary) distributions, and thus provides a rich tool box of building blocks for statistical modeling. In this model class, the process is attracted to an equilibrium point denoted α, or to be more precise, α is the asymptotic mean in the stochastic dynamical system in absence of external input. It is not only of interest to estimate the rate constant β, but also the mean level α, as well as a parameter scaling the noise level, denoted σ.

The approach taken here is to manipulate the system away from its stationary level α and into a regime of transient behavior, where the dynamics are strongly shaped by the rate parameter. This can be achieved by current stimulation as well as by imposing other non-electrical stimuli. We call this perturbation, and illustrate it by starting the process far away from its mean. A natural generalization would be to perturb the process repeatedly whenever it has reached the proximity of the stationary mean. However, for notational simplicity, we only treat the case where we perturb the initial condition, i.e., we study the properties of the estimator as a function of the initial point where the process is started. It is intuitively clear that the estimation of β benefits from the perturbation if considering the system in absence of noise. The time constant shapes the decay towards the steady state α, but once the system is there, the time constant plays practically no role, and limited information can be extracted. Thus, contrary to the common statistical paradigm that estimation is best conducted on stationary data, we claim that it is the transient behavior that carries information about the time constant. The key variable is the initial condition, which should be far from equilibrium. Then, an obviously important question is if this improved determination of β caused by perturbation of the system, does not happen at the cost of a poorer estimation of the other parameters. Intuitively, the asymptotic mean α is best determined if the system fluctuates around its resting state, where the average of the observations would provide a good estimator. However, this would be a strongly biased estimator if the system is only observed in the transient phase, i.e., before it reaches stationarity. It turns out that this intuition is misleading, and α can still be well estimated in the perturbed system, as long as its estimator is cleverly chosen, at least if the sample size is not very small. Finally, we will also illustrate that the estimator of the variance parameter is in most cases unaffected by the perturbation. The estimators are the MLEs when available, and otherwise estimators approximating the MLE. The estimators possess standard statistical asymptotic properties, ensuring that the parameters can be identified with a prerequired precision. However, this is a theoretical result based on unlimited amounts of data. The main message and result of this paper is that much smaller sample sizes are necessary if the behavior of the system is perturbed.

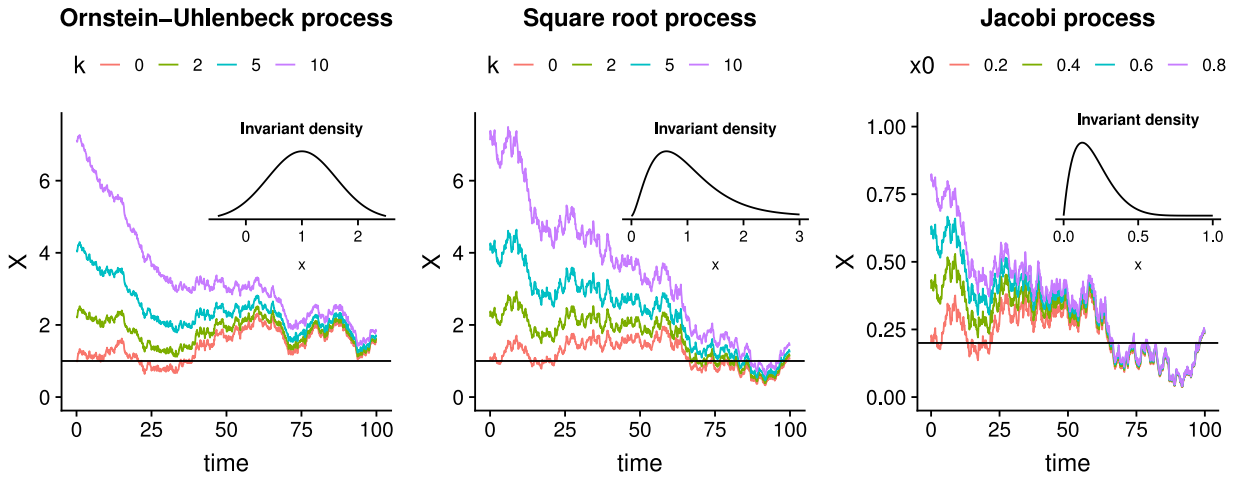

Fig. 1. Example traces of the Ornstein-Uhlenbeck process, the square root process and the Jacobi process, with different starting values. The starting values are indicated by k, the number of standard deviations above the invariant mean for the Ornstein-Uhlenbeck and the square root processes. All traces are simulated with the same sequence of random numbers. The Ornstein-Uhlenbeck and the square root process have the same invariant mean and variance. The black lines are the invariant means. The inlets are the stationary distributions.

2. Models and estimators

A stochastic diffusion with linear drift is a prototype of a noisy relaxation process given by the Langevin equation

$$\mathrm{d}X(t) = -\beta(X(t)-\alpha)\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t); \quad X(0) = x_0 \tag{1}$$

where $W = \{W(t);\, t \geq 0\}$ is a standard Wiener process and $\sigma(\cdot)$ is a real-valued function called the diffusion coefficient, which we assume fulfills conditions such that a solution to (1) exists. The parameter $\beta > 0$ is a scaling of the time, the rate at which the diffusion spontaneously moves towards its resting level α. The inverse, $1/\beta$, is often called the relaxation time constant. The diffusion coefficient function $\sigma(x)$ determines the stationary distribution as well as the transition density of the process, i.e., the distribution of X at time t, given it was at a certain point x at an earlier time s. Assume that $\sigma(x) = \sigma\sqrt{ax^2+bx+c}$, where $\sigma > 0$ is a scale parameter of the diffusion coefficient, and a, b and c are such that the square root is well defined when $X(t)$ is in its state space. Any of a, b and c are allowed to be 0, but not all three at the same time. These are called Pearson diffusions because the class of stationary distributions equals the Pearson system of distributions. Specifically, $\sigma(x)$ can be chosen such that the state space of the process is positive, negative, real valued or bounded, and the distributions can be symmetric or skewed, and both heavy- and light tailed distributions are possible. Examples are normal, gamma or beta distributions, which are the three examples we will treat in more detail. The Ornstein-Uhlenbeck (OU) process is linear and has $a = 0, b = 0, c = 1$ so that $\sigma(x) = \sigma$, it lives on the entire real axis and the invariant distribution is the normal distribution. The other two models are nonlinear. The square-root (SQ) process has $a = 0, b = 1, c = 0$ so that $\sigma(x) = \sigma\sqrt{x}$, it lives on the positive real axis and the invariant distribution is the gamma distribution. The Jacobi (JA) diffusion has $a = -1, b = 1, c = 0$ so that $\sigma(x) = \sigma\sqrt{x(1-x)}$, it lives on the bounded interval (0,1) and the invariant distribution is the beta distribution. In Fig. 1 some example trajectories and stationary densities are plotted. The following results on Pearson diffusions can all be found in Forman and Sørensen [13], Sørensen [35]. The linear drift in (1) determines the mean and the autocorrelation of the process regardless of $\sigma(x)$, as long as $\sigma(x)$ is such that these moments exist. The mean of the stationary distribution is α. The stationary variances in the Pearson diffusions depend on $\sigma(x)$, and are all functions of α and γ^2, defined by the parameter function

$$\gamma^2 := \frac{\sigma^2}{2\beta}. \tag{2}$$

Assume $X(t)$ is observed at equidistant time points $i\Delta$, $i = 0, 1, \ldots, n$, for some $\Delta > 0$, and let x_i denote the observation of $X(t)$ at time $i\Delta$. Likewise, let X_i denote the random variable $X(i\Delta)$. The autocorrelation between X_i and X_{i-1} is defined by $\mathbb{E}[(X_i-\alpha)(X_{i-1}-\alpha)]/\sqrt{\mathrm{V}[X_i]\mathrm{V}[X_{i-1}]}$ and equals $e^{-\beta\Delta}$, denoted by

$$\rho := e^{-\beta\Delta}. \tag{3}$$

The conditional mean is

$$\mathbb{E}[X_{i+j}|X_i = x] = x\rho^j + \alpha(1-\rho^j). \tag{4}$$

Note that the conditional mean converges to α for $j \to \infty$.

If $\Delta\beta \ll 1$ the sampling step is short compared to the time constant, and the autocorrelation given in Eq. (3) is $0 \ll \rho < 1$. A minimal requirement of the sampling is that $\Delta\beta$ is not large, because otherwise $\rho \approx 0$ and the observations are approximately independent. Then only parameters of the stationary distribution can be estimated, i.e., α and the variance of the invariant distribution, which depends on the specific model. However, the rate parameter β and the variance parameter σ^2 can not be desentangled in the parameter function given in (2) from stationary and independent data.

When discretely observed, the Pearson diffusions are statistically tractable, even if the likelihood function is only known in two special cases, namely for the OU and the SQ process. The OU process is a Gaussian process, and the least square estimator obtained by minimizing $\sum(X_i - \mathbb{E}[X_i|X_{i-1}])^2$ coincides with the maximum likelihood estimator. For the SQ and JA models we will use martingale estimating functions [13,35].

If the length of the observation interval $T = \Delta n$ is fixed, any estimator of β in (1) is inconsistent (i.e., the estimate does not concentrate around the true value, no matter how frequent we observe, not even if we observe the complete continuous trajectory), whereas parameters in the diffusion coefficient can be consistently estimated when $n \to \infty$ and $\Delta \to 0$ [35]. If we further require that $T \to \infty$, also β can be consistently estimated, but as we shall see below, the convergence is slow, whereas σ^2 is well estimated even for small sample sizes as long as Δ is not too large. This is because the optimal rate of convergence for β is $\sqrt{T} = \sqrt{\Delta n}$, whereas it is $\sqrt{n}$ for σ, see [35]. Note that $\Delta n\beta = T\beta$ is the length of the observation interval measured in units of the time constant $1/\beta$ of the system.

The initial condition x_0 will play a major role in the following. When $|x_0 - \alpha| \gg 0$ we say that the system is perturbed, whereas if $x_0 \approx \alpha$ we say it is unperturbed. To ease notation we fix the dependence on the initial condition, and all expectations and variances in the following are with respect to $X_0 = x_0$, which is the key variable in the present study. We measure x_0 as distance from the asymptotic mean in units of the asymptotic standard deviation γ, such that

$$x_0 = \alpha + k\gamma \tag{5}$$

for the OU and the SQ processes, and uniformly in the bounded state space for the JA process.

2.1. Ornstein-Uhlenbeck process

The OU process is the simplest and most commonly used Pearson diffusion with applications in numerous fields, such as physics [14], neuroscience [1,6], engineering [4], psychology [36] and finance [37]. We will treat the OU process in detail, and calculate approximate expressions for bias and variance of the estimators, showing how perturbation improves the quality of the estimates. This is possible because relevant quantities are explicitly calculable for a Gaussian process. We will then show through simulations that the same holds true for other Pearson diffusions. The OU process has constant diffusion coefficient, $\sigma(x) = \sigma$, and the transition probability density function of $X(t)$ is normal with conditional mean (4) and conditional variance (see e.g. [6,13])

$$\mathrm{V}[X_i|X_{i-1} = x] = \gamma^2(1 - \rho^2), \tag{6}$$

where ρ and γ^2 are defined in Eqs. (2) and (3). The stationary variance is the limit of (6) for Δ going to infinity, γ^2.

Estimation of α, β and σ^2 can be done by MLE, which generally is the preferred method of choice, and provides both consistent and rate optimal estimators when $T \to \infty$. The MLEs for the OU process of α, β and σ^2 are derived in Section 5.1, see also Lansky [24, Eqs. (24)-(26)], and given as solutions to the equations,

$$\hat{\alpha} = \frac{1}{n}\sum_{j=1}^{n} x_j + \frac{\hat{\rho}}{n(1-\hat{\rho})}(x_n - x_0) \tag{7}$$

$$\approx \frac{1}{n+1}\sum_{j=0}^{n} x_j = \bar{x}, \tag{8}$$

$$\hat{\rho} = \frac{\sum_{j=1}^{n}(x_j - \hat{\alpha})(x_{j-1} - \hat{\alpha})}{\sum_{j=1}^{n}(x_{j-1} - \hat{\alpha})^2}, \tag{9}$$

$$\hat{\sigma}^2 = \frac{2\sum_{j=1}^{n}\left(x_j - x_{j-1}\hat{\rho} - \hat{\alpha}(1-\hat{\rho})\right)^2\hat{\beta}}{n\left(1-\hat{\rho}^2\right)}, \tag{10}$$

where $\hat{\beta} = -\log\hat{\rho}/\Delta$, and the symbol $\hat{}$ indicates it is an estimator. See e.g. [31, p. 152] or [8] for the case when $\alpha = 0$. Note how the estimator (7) of α is approximately equal to the average (8) if the initial and the last observations are close. The last term in (7) is often ignored, but will turn out to be very important for perturbed data sets, where it corrects for the otherwise induced bias, and ensures that estimates of α are still reasonable. When estimating with the true likelihood, i.e., Eqs. (7) and (9), a system with two equations in two unknowns has to be solved, e.g., by the Newton-Raphson method. When (8) is used the estimators are explicit, and there is no need to implement an iterative algorithm. Therefore, this simpler estimator is often used. However, as we shall see, this is not suitable in our case.

The asymptotic variances in the limit $n \to \infty$ of $\hat{\alpha}$ in (7), and $\hat{\beta}$ and $\hat{\sigma}^2$ in (9)–(10) obtained by inverting the Fisher information are $\mathrm{Var}[\hat{\alpha}] \approx \sigma^2/n\Delta\beta^2 = 2\gamma^2/n\Delta\beta$, $\mathrm{Var}[\hat{\beta}] \approx 2\beta/n\Delta$ and $\mathrm{Var}[\hat{\sigma}^2] \approx 2\sigma^4/n$, see Section 5.2. Nevertheless, we are

interested in finite sample size properties, and specifically in the influence of the initial condition, and thus, the asymptotic variances are not relevant here. However, these variances can serve as a benchmark, so if the experiment can be designed such that the variance of an estimator approaches the asymptotic variance, we are close to an optimal situation.

Estimation of the rate parameter In the following we will present the main results of this article. Define the random variables

$$Z = \sum_{i=1}^{n} (X_i - \alpha)(X_{i-1} - \alpha)\,;\; Y = \sum_{i=1}^{n} (X_{i-1} - \alpha)^2. \tag{11}$$

Then approximately, see Eq. (9),

$$\hat{\beta} \approx -\frac{1}{\Delta}\log(Z/Y), \tag{12}$$

where the approximation comes from using the true (unknown) value of α and not the estimated one. We only use expression (12) to calculate approximate theoretical expressions for the mean and variance of the estimator. When estimating we of course use the estimator (7)–(10), since all parameters are unknown. Thus, the variance in the theoretical expressions below will underestimate the true variance of $\hat{\beta}$, since the statistical variability due to α also being estimated will increase the overall variance. Taylor expansions up to second order of $\hat{\beta}$ and $\hat{\beta}^2$ with α fixed at its true value in (12) around the mean values of Z and Y and taking expectations are developed in Section 5.3, and yield

$$\mathbb{E}[\hat{\beta}] \approx \beta + \frac{1}{2\Delta\mathbb{E}[Y]^2}\left(\frac{1}{\rho^2}\mathrm{V}[Z] - \mathrm{V}[Y]\right), \tag{13}$$

$$\mathrm{V}[\hat{\beta}] \approx \frac{1}{\Delta^2\mathbb{E}[Y]^2}\left(\frac{1}{\rho^2}\mathrm{V}[Z] + \mathrm{V}[Y] - \frac{2}{\rho}\mathrm{Cov}[Z,Y]\right). \tag{14}$$

To evaluate (13) and (14) further, moments of (11) have to be inserted. These are evaluated in Section 5.4. The approximate bias and variance of $\hat{\beta}$ obtained from Eqs. (13) and (14) are thus

$$\mathbb{E}[\hat{\beta}] - \beta \approx \frac{1+3\rho^2}{2\Delta\rho^2\left(n + \frac{1-\rho^{2n}}{1-\rho^2}(k^2-1)\right)} - \frac{2n(k^2-1)(1-\rho^2)\rho^{2n-2} + 2(1-\rho^{2n})}{\Delta(1-\rho^2)\left(n + \frac{1-\rho^{2n}}{1-\rho^2}(k^2-1)\right)^2}, \tag{15}$$

$$\mathrm{V}[\hat{\beta}] \approx \frac{(1-\rho^2)}{\Delta^2\rho^2\left(n + \frac{1-\rho^{2n}}{1-\rho^2}(k^2-1)\right)}. \tag{16}$$

Both converges to 0 with $T \to \infty$, the length of the observation interval, as expected. What is more remarkable is that they both scale as $1/k^2$, and thus, because of the square term, the desired convergence to 0 is fast when perturbing the initial condition. The Mean Squared Error (MSE) is the sum of the squared bias and the variance,

$$\mathrm{MSE}\,[\hat{\beta}] = (\mathbb{E}[\hat{\beta}] - \beta)^2 + \mathrm{V}[\hat{\beta}]. \tag{17}$$

In Fig. 2 we illustrate the relative MSE normalized by dividing by β^2, approximated by expressions (15) and (16), as a function of $T = n\Delta$ and k, for high ($\Delta = 0.1$) and low ($\Delta = 1$) frequency sampling. It is clear that to obtain a lower MSE, it is more efficient to increase k than n. It is also clear that what matters is not the number of observations, but the length of the observation interval T.

If we assume $\Delta\beta \ll 1$ we can further approximate the *relative* variance using expression (16), depending on whether n is small or large.

- If n is small, then $\rho^m \approx 1 - m\beta\Delta$ for $m \le 2n$, and the *relative* variance can be further approximated by

$$\frac{\mathrm{V}[\hat{\beta}]}{\beta^2} \approx \frac{2}{\Delta n\beta k^2}. \tag{18}$$

  Thus, if for experimental reasons it is only possible to observe the system over an interval which is short compared to the typical time constant of the system, it is paramount to perturb the system to get reliable estimates, since otherwise the variance explodes.
- If n is large, then $\rho^2 \approx 1 - 2\beta\Delta$ and $\rho^n \approx 0$, and the following approximation is obtained

$$\frac{\mathrm{V}[\hat{\beta}]}{\beta^2} \approx \frac{2}{\Delta n\beta + (k^2-1)/2} \tag{19}$$

  which still emphasizes the positive effect of perturbation.

The approximations above depend on how reliable the Taylor expansions in Eqs. (13) and (14) are. Especially in the boundary situation of n and k small, these expressions probably underestimate the true bias and variance [8].

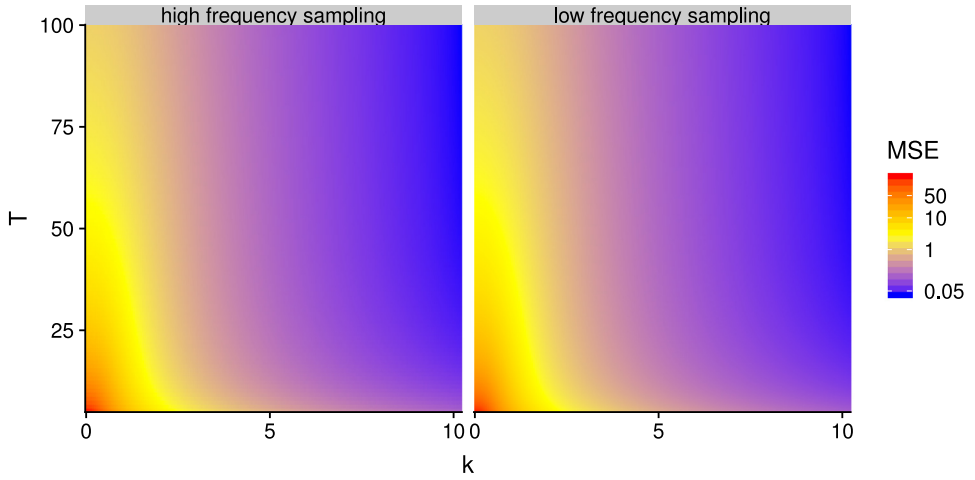

Fig. 2. Relative Mean Squared Error of the estimator of the rate parameter β, approximated by expressions (15) and (16), normalized by dividing by β^2. The MSE is given as a function of the length of the observation interval $T = n\Delta$ on the y axis, and amount of perturbation of x_0 away from the mean α, measured in number k of standard deviations on the x-axis. Parameters are $\beta = 0.05, \alpha = 0, \sigma = 0.01$. Left plot: High frequency, $n = 1000, \Delta = 0.1$. Right plot: Low frequency, $n = 100, \Delta = 1$. The two panels look practically the same, which illustrates that it is not the number of observations but the length of the observation interval that matters for the estimation of the rate parameter.

Estimation of the asymptotic mean The next question is then how the estimation of the other parameters is influenced by perturbation. Expected values of $\hat{\alpha}$ and $\hat{\alpha}^2$ with ρ fixed at its true value in (7), again using expressions (4) and (6), yield

$$\mathbb{E}[\hat{\alpha}] \approx \alpha, \qquad \mathrm{V}[\hat{\alpha}] \approx \frac{\gamma^2(1+\rho)}{n(1-\rho)}. \tag{20}$$

The only approximation used here is assuming ρ known, implying a slight underestimation of the variance, but the moments are otherwise exact. Note how the value of k, or equivalently x_0, does not enter the expressions, and the estimator is approximately unbiased even under perturbation. This is not true for the average estimator (8), where the bias is $k\gamma(1-\rho^{n+1})/(n+1)(1-\rho)$, which is increasing with larger perturbation, but decreasing in n. It is therefore important to use the slightly more involved maximum likelihood estimator (7) when perturbing. However, if the observation interval is too small compared to the typical time constant of the system, $1/\beta$, then the estimator of α will deteriorate for large perturbation, as we will see in the simulation study in the next section. The variance in (20) is equal to the asymptotic variance given by the inverse of the Fisher information.

Estimation of the variance parameter The approximate mean and variance of $\hat{\sigma}^2$ are

$$\mathbb{E}[\hat{\sigma}^2] \approx \sigma^2, \qquad \mathrm{V}[\hat{\sigma}^2] \approx \frac{2\sigma^4}{n} \tag{21}$$

which are not affected by the perturbation. In general, the scale parameter of the diffusion coefficient, σ^2, is not difficult to estimate. Note how only n enters, and thus, decreasing Δ and increasing n for fixed observation interval T will improve the estimation of σ. This is not the case for the drift parameters, where more observations for fixed T does not improve the estimation.

2.2. Square root process

The SQ process has been used in finance under the name of the Cox-Ingersoll-Ross process and in neuroscience under the name of the Feller process, but also finds application in other fields such as physics and biology [3,7,34]. It has diffusion coefficient, $\sigma(x) = \sigma\sqrt{x}$. The transition probability density function of $X(t)$ is a non-central chisquare distribution with conditional mean (4) and variance

$$\mathrm{V}[X_i|X_{i-1} = x] = \alpha\gamma^2(1-\rho)^2 + 2x\gamma^2(1-\rho)\rho, \tag{22}$$

see [7]. It is a process restricted to the positive real line with a right skewed distribution, and is thus a good alternative to the OU process when modeling strictly positive processes. Following Feller's classification of boundaries [21], the boundary 0 is entrance if $\alpha \geq \gamma^2$, meaning that it can start from 0, but will have probability zero of ever hitting 0 again. In this case it admits a stationary distribution, which is a gamma distribution with mean α and variance $\alpha\gamma^2$, that is, with shape parameter α/γ^2 and scale parameter γ^2 [7].

The likelihood of the SQ process involves special functions of the Bessel kind, which can be numerically unstable to evaluate. Instead we will use approximate MLEs, based on martingale estimating functions. The estimator of α coincides

with Eq. (7), and the estimators for $\beta(\rho)$ and σ^2 are [35, p. 21]

$$\hat{\rho} = \frac{n\sum_{j=1}^n \frac{x_j}{x_{j-1}} - \left(\sum_{j=1}^n x_j\right)\left(\sum_{j=1}^n \frac{1}{x_{j-1}}\right)}{n^2 - \left(\sum_{j=1}^n x_{j-1}\right)\left(\sum_{j=1}^n \frac{1}{x_{j-1}}\right)}, \tag{23}$$

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n \frac{1}{x_{j-1}}\left(x_j - x_{j-1}\hat{\rho} - \hat{\alpha}(1-\hat{\rho})\right)^2 \hat{\beta}}{\sum_{j=1}^n \frac{1}{x_{j-1}}\left((\frac{\hat{\alpha}}{2} - x_{j-1})\hat{\rho}^2 - (\hat{\alpha} - x_{j-1})\hat{\rho} + \frac{\hat{\alpha}}{2}\right)}. \tag{24}$$

It is difficult to find theoretical results on moments of the estimators for small samples and as a function of a perturbation for the SQ, but we conjecture that the same pattern emerges as for the OU process, in particular, that the estimate of β can be hugely improved by perturbation. Since the process is bounded below by 0, it is most natural to perturb it to values above the mean, whereas for the OU process this is irrelevant since OU is symmetric around its mean. We study the estimators through simulations in the next section.

2.3. Jacobi diffusion

The Jacobi diffusion is the only bounded Pearson diffusion, and has for example been applied in finance [26], genetics [21, chap 15, sec 2.F], neuroscience [11,12] and for ion channel dynamics [5,10]. It has diffusion coefficient $\sigma(x) = \sigma\sqrt{x(1-x)}$, and lives on the bounded interval (0,1). A unique ergodic solution exists if $\alpha \in (0, 1)$, $\alpha \geq \gamma^2$ and $(1-\alpha) \geq \gamma^2$. In that case, it has stationary distribution the beta distribution, $\text{Beta}(\alpha/\gamma^2, (1-\alpha)/\gamma^2)$, with mean α and variance $\alpha(1-\alpha)\gamma^2/(1+\gamma^2)$. A special case is obtained for $\alpha = \gamma^2 = 1/2$, in which case it is standard uniformly distributed. It follows that for any strictly increasing and twice differentiable distribution function F, another diffusion is obtained by $F^{-1}(X(t))$ with invariant density given by $f = F'$. It is thus an important and flexible building block for modeling purposes. In particular, an affine transformation can transfer it to any desired bounded interval.

The transition probability density function of $X(t)$ is unknown, but all moments are known. It has conditional mean (4) and variance

$$\begin{aligned}\mathrm{V}[X_i|X_{i-1} = x] = {} & \rho^2 e^{-\sigma^2\Delta}\left(x^2 - \frac{\alpha+\gamma^2}{\frac{1}{2}+\gamma^2}\left(x - \frac{\frac{1}{2}\alpha}{1+\gamma^2}\right)\right) \\ & - \rho^2(x-\alpha)^2 + \rho(x-\alpha)\frac{(1-2\alpha)\gamma^2}{\frac{1}{2}+\gamma^2} + \frac{\alpha(1-\alpha)\gamma^2}{1+\gamma^2},\end{aligned} \tag{25}$$

see Section 5.5. Since the transition probability densities are unknown, also the likelihood function is unknown. Instead we will use approximate MLEs, based on martingale estimating functions [35]. They are given as solutions to the equations,

$$\hat{\alpha} = \frac{\sum_{j=1}^n \frac{x_j - x_{j-1}\hat{\rho}}{x_{j-1}(1-x_{j-1})}}{(1-\hat{\rho})\sum_{j=1}^n \frac{1}{x_{j-1}(1-x_{j-1})}}, \tag{26}$$

$$\hat{\rho} = \frac{\sum_{j=1}^n \frac{(x_j - \hat{\alpha})(x_{j-1} - \hat{\alpha})}{x_{j-1}(1-x_{j-1})}}{\sum_{j=1}^n \frac{(x_{j-1} - \hat{\alpha})^2}{x_{j-1}(1-x_{j-1})}}, \tag{27}$$

$$\hat{\sigma}^2 = \frac{1}{n\Delta}\sum_{j=1}^n \frac{\left(x_j - x_{j-1}\hat{\rho} - \hat{\alpha}(1-\hat{\rho})\right)^2}{x_{j-1}(1-x_{j-1})}, \tag{28}$$

which are derived in Section 5.6. Also for this model, it is difficult to find theoretical results on moments of the estimators for small samples and as a function of a perturbation, but we conjecture that the same pattern emerges, and that the estimate of β can be hugely improved by perturbation. Note that it is not possible to perturb very much, since the process is bounded between 0 and 1, and it does not make sense to measure the perturbation in k, as was done for the OU and the SQ processes. Instead, we use different starting conditions in the interval (0,1). We study the estimators through simulations in the next section.

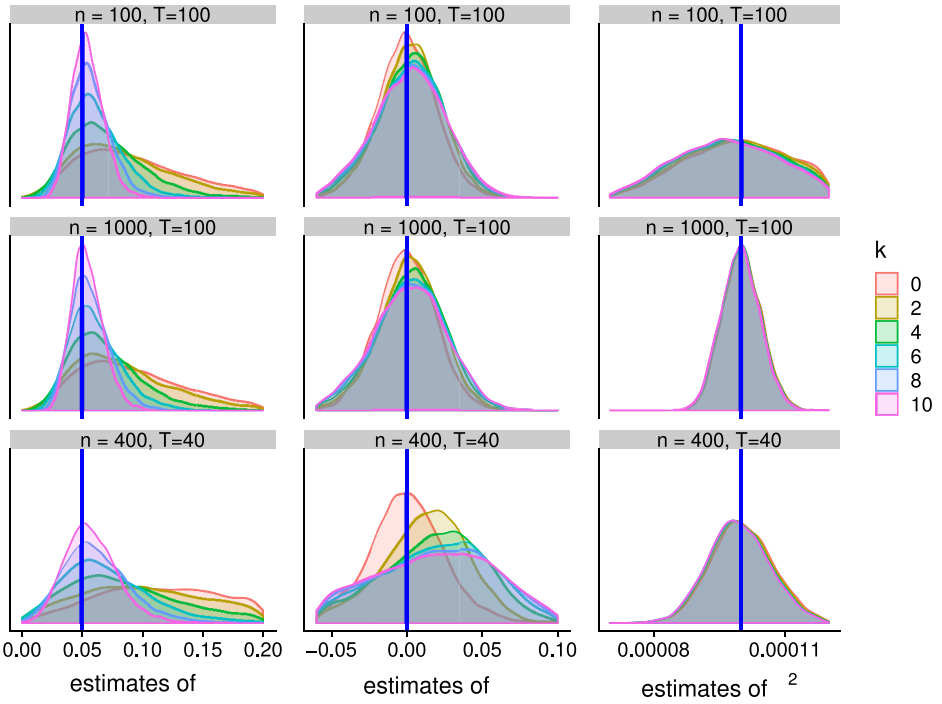

Fig. 3. Simulation study for the Ornstein-Uhlenbeck process. For each starting condition indicated by the number k of standard deviations above the stationary mean, 10,000 data sets were simulated, and on each data set, parameters were estimated. The panels show density plots of the 10,000 estimated parameters. Three situations were considered. Upper plots: Low frequency, long observation interval, $n = 100, \Delta = 1$. Middle plots: High frequency, long observation interval, $n = 1000, \Delta = 0.1$. Lower plots: High frequency, short observation interval, $n = 400, \Delta = 0.1$. The long observation interval corresponds to 5 times the typical time constant of the system, $1/\beta$, the short observation interval corresponds to 2 times. The parameter values used in the simulations are indicated by the blue straight lines, they are $\beta = 0.05, \alpha = 0, \sigma = 0.01$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Simulation studies

In this section we test the estimators of the three processes through simulation studies, in particular, we vary the amount of perturbation through changing the initial condition x_0. When running a new set of simulations with a new value of x_0, we always use the same set of random numbers, such that changes in estimates are due to the perturbation and not due to random variation.

3.1. Ornstein-Uhlenbeck process

Trajectories from the OU process were simulated according to the exact distribution, the normal distribution with mean given in Eq. (4) and variance given in Eq. (6) with parameter values $\beta = 0.05, \alpha = 0, \sigma = 0.01$. Thus, the typical time constant of the system is $1/\beta = 20$ time units. We denote the sampling to be high frequency when $\Delta = 0.1$, and low frequency when $\Delta = 1$. This corresponds to sampling 200 and 20 times per typical time constant, respectively. We denote the observation interval to be long when $T = 100$ and short when $T = 40$. This corresponds to sampling over an interval that has length 5 and 2 times the typical time constant, respectively. Three sets of simulations were run: 1) High frequency sampling, long observation interval, $n = 1000, \Delta = 0.1$. 2) High frequency, short observation interval, $n = 400, \Delta = 0.1$. 3) Low frequency, long observation interval, $n = 100, \Delta = 1$. Within each set of simulations, the perturbation was varied as $k \in \{0, 2, 4, 6, 8, 10\}$. For each value of k, 10,000 data sets were simulated, and on each data set, the three parameters β, α and σ^2 were estimated.

Density plots of the 10,000 estimates for each condition can be found in Fig. 3. It is clearly seen how the estimate of β improves with increasing perturbation, and that for these relatively short sampling intervals, it is crucial to have at least some perturbation, preferably larger than $k = 2$ standard deviations away from the stationary mean. The estimate of α is for long sampling intervals not affected by the perturbation, however, for short sampling intervals, it does deteriorate for increasing perturbation. Both drift parameters are unaffected by the frequency of the sampling, but improve with increasing length T. The estimate of the variance parameter σ is not at all affected by the perturbation, but improves with increasing sampling frequency and increasing number of observations. Note that the estimate of σ improves even when the sampling interval is shorter, as long as n is larger.

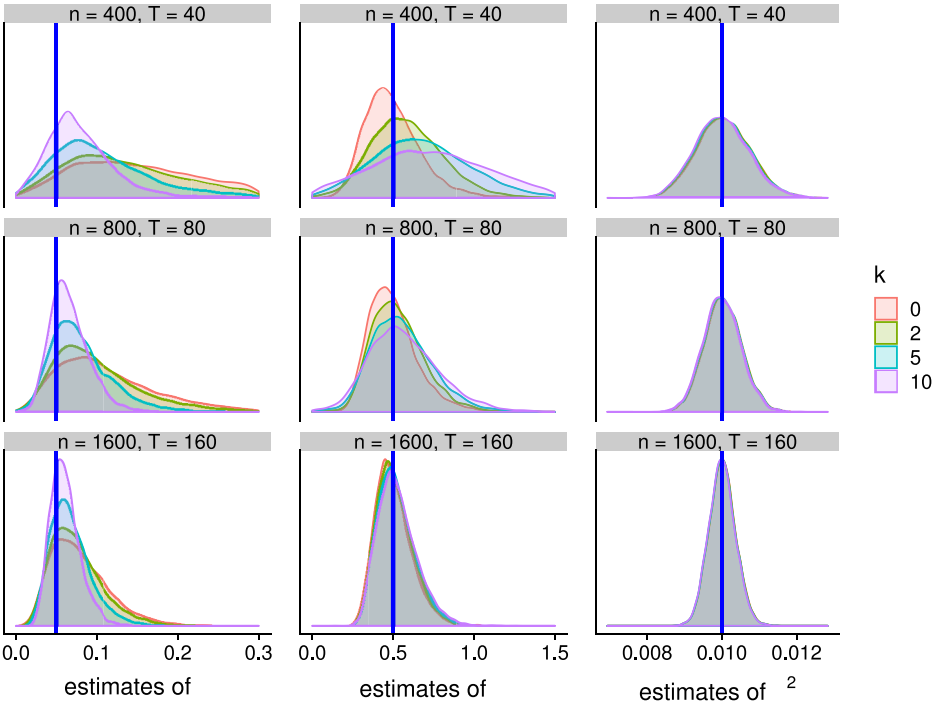

Fig. 4. Simulation study for the square-root process. For each starting condition indicated by the number k of standard deviations above the stationary mean, 10,000 data sets were simulated, and on each data set, parameters were estimated. The panels show density plots of the 10,000 estimated parameters. Three situations were considered. Upper plots: Short observation interval, $n = 400, \Delta = 0.1$. Middle plots: Medium observation interval, $n = 800, \Delta = 0.1$. Lower plots: Long observation interval, $n = 1600, \Delta = 0.1$. The long observation interval corresponds to 8 times the typical time constant of the system, $1/\beta$, the medium observation interval corresponds to 4 times, and the short observation interval corresponds to 2 times. The parameter values used in the simulations are indicated by the blue straight lines, they are $\beta = 0.05, \alpha = 0, \sigma = 0.1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Square-root process

Trajectories from the SQ process were simulated according to the Euler-Maruyama scheme, with a time step of 0.01 time units, which was then subsampled to a time step of $\Delta = 0.1$, to minimize discretization bias. Parameter values were $\beta = 0.05, \alpha = 0.5, \sigma = 0.1$. Thus, as for the OU process, the typical time constant of the system is $1/\beta = 20$ time units. We only considered high frequency sampling, since the main interest is the quality of the estimation of the drift parameters, which depends on the length of the sampling interval, as was shown for the OU process. We denote the observation interval to be long when $T = 160$, medium when $T = 80$ and short when $T = 40$ time units. This corresponds to sampling over an interval that has length 8, 4 and 2 times the typical time constant, respectively. Three sets of simulations were run: 1) Long observation interval, $n = 1600, \Delta = 0.1$. 2) Medium observation interval, $n = 800, \Delta = 0.1$. 3) Short observation interval, $n = 400, \Delta = 0.1$. Within each set of simulations, the perturbation was varied as $k \in \{0, 2, 5, 10\}$. For each value of k, 10,000 data sets were simulated, and on each data set, the three parameters β, α and σ^2 were estimated.

Density plots of the 10,000 estimates for each condition can be found in Fig. 4. It is clearly seen that also for the SQ process, the estimate of β improves with increasing perturbation, and that for these relatively short sampling intervals, it is crucial to have at least some perturbation, preferably larger than $k = 2$ standard deviations above the stationary mean. The estimate of α is for long sampling intervals not affected by the perturbation, however, for short sampling intervals, it does deteriorate for increasing perturbation. Sometimes for short observation intervals, the estimate of α becomes negative, in particular for large perturbation. This occurs because the last term on the right hand side in equation (7) dominates, and thus, in these cases, the estimate of α does not exist, even if the estimate of β is reasonable. Both drift parameters improve with increasing length T. The estimate of the variance parameter σ is not at all affected by the perturbation, but improves with increasing number of observations.

It sometimes occurs that the estimator for a specific data set does not exist. This happens if the statistic in the right hand side of Eq. (23), which is the estimate of the autocorrelation, is not in the interval (0,1). The probability that the estimator does not exist decreases with increasing length of the sampling interval. A relevant question is then how the perturbation affects this probability. In Table 1, the percentages of the simulated samples where the estimator did not exist are listed. It is seen that the estimator nearly always exists, and the perturbation improves (decreases) this probability for medium length observation intervals, however, for short sampling intervals, it slightly increases this probability.

Table 1
The percentages of simulated samples where the estimator of the square-root process did not exist.

	$k = 0$	$k = 2$	$k = 5$	$k = 10$
$n = 400$	0.69%	0.74%	1.06%	0.88%
$n = 800$	0.11%	0.04%	0.05%	0.01%
$n = 1600$	0%	0%	0%	0%

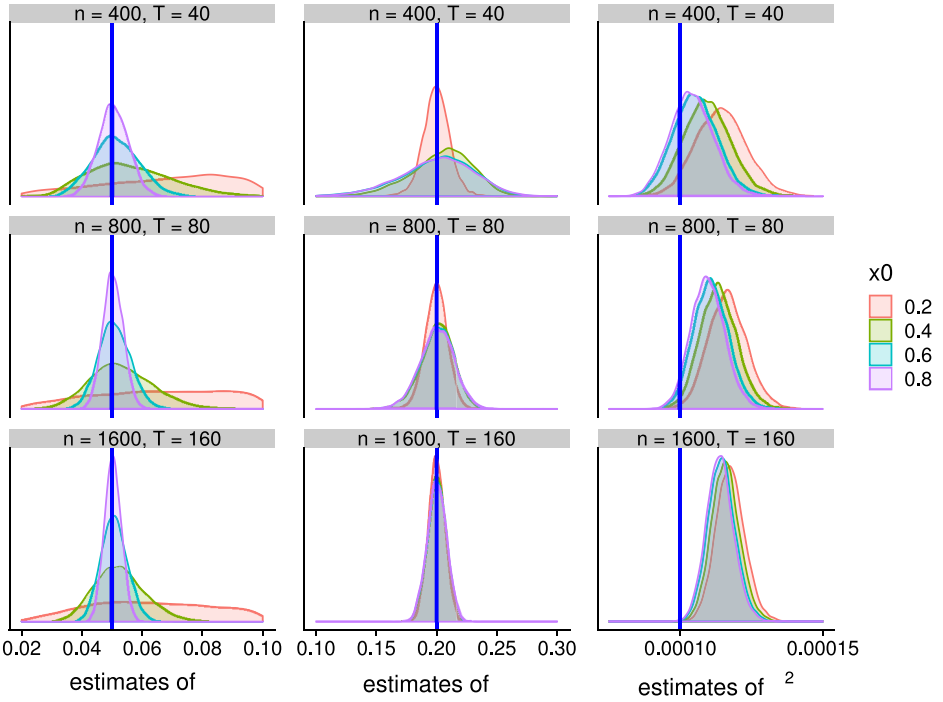

Fig. 5. Simulation study for the Jacobi process. For each starting condition x_0, 10,000 data sets were simulated, and on each data set, parameters were estimated. The panels show density plots of the 10,000 estimated parameters. Three situations were considered. Upper plots: Short observation interval, $n = 400, \Delta = 0.1$. Middle plots: Medium observation interval, $n = 800, \Delta = 0.1$. Lower plots: Long observation interval, $n = 1600, \Delta = 0.1$. The long observation interval corresponds to 8 times the typical time constant of the system, $1/\beta$, the medium observation interval corresponds to 4 times, and the short observation interval corresponds to 2 times. The parameter values used in the simulations are indicated by the blue straight lines, they are $\beta = 0.05, \alpha = 0.2, \sigma = 0.01$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Jacobi process

Trajectories from the JA process were simulated according to the Euler-Maruyama scheme, with a time step of 0.01 time units, which was then subsampled to a time step of $\Delta = 0.1$, to minimize discretization bias. Parameter values were $\beta = 0.05, \alpha = 0.2, \sigma = 0.01$. Otherwise, the setup is as for the SQ process, except for the initial conditions. Within each set of simulations, the perturbation was varied as $x_0 \in \{0.2, 0.4, 0.6, 0.8\} = \alpha \times \{1, 2, 3, 4\}$. For each value of x_0, 10,000 data sets were simulated, and on each data set, the three parameters β, α and σ^2 were estimated.

Density plots of the 10,000 estimates for each condition can be found in Fig. 5. It is clearly seen that also for the JA process, the estimate of β improves with increasing perturbation. The estimate of α is for long sampling interval not affected by the perturbation, however, for short sampling intervals, it does slightly deteriorate for increasing perturbation. Both drift parameters improve with increasing length T. The estimate of the variance parameter σ is slightly biased, however, note the scale, so the bias is small.

The estimator for a specific data set does not exist if the numerator in the right hand side of Eq. (27) is negative. In Table 2, the percentages of the simulated samples where the estimator did not exist are listed. It is seen that the estimator nearly always exists.

4. Concluding remarks

We have shown how perturbation can remedy the statistical problems related to the estimation of time scaling parameters in stochastic processes in a wide class of one dimensional diffusion models. Probably also for multidimensional

Table 2
The percentages of simulated samples where the estimator of the Jacobi process did not exist.

	$x_0 = 0.2$	$x_0 = 0.4$	$x_0 = 0.6$	$x_0 = 0.8$
$n = 400$	1.61%	5.20%	0.46%	0%
$n = 800$	0.37%	0%	0%	0%
$n = 1600$	0.01%	0%	0%	0%

stochastic processes with different time scales, the statistical inference will benefit from data obtained from the transient parts of the state space.

In many experiments only limited amount of data can be obtained, for example because samples are expensive, or the experimental setting only allows for a finite number of samples. Therefore it is important to design the experiment such that maximal information can be extracted from each data point.

In some physical systems an unlimited perturbation is not possible, such as injecting current into a neuronal cell that might die. An alternative could be repeated smaller perturbations to improve the estimate of the rate constant, which would furthermore have the advantage that the estimate of the mean parameter would not be deteriorated.

5. Mathematical details

5.1. Maximum likelihood estimators of the Ornstein-Uhlenbeck process

The likelihood function of the OU process is the product of the transition densities

$$L_n(\theta) = \prod_{i=1}^{n} p(\triangle, X_{i-1}, X_i; \theta)$$

where the transition density is normal with mean (4) and variance (6),

$$p(\triangle, X_{i-1}, X_i; \theta) = \frac{1}{\sqrt{2\pi\gamma^2(1-\rho^2)}} \exp\left(-\frac{(X_i - X_{i-1}\rho - \alpha(1-\rho))^2}{2\gamma^2(1-\rho^2)}\right).$$

The score function is the vector of derivatives of the log-likelihood function with respect to the parameters, and the MLE is given as solution to the likelihood equations $\partial_\theta \log L_n(\theta) = 0$, where θ is either β, α or σ^2. The elements of the score function are thus

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\log L_n(\theta) &= \frac{(1-\rho)}{\gamma^2(1-\rho^2)}\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho)),\\
\frac{\partial}{\partial\beta}\log L_n(\theta) &= \frac{n}{2\beta} - \frac{n\Delta\rho^2}{1-\rho^2} - \frac{\Delta\rho\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))(X_{i-1}-\alpha)}{\gamma^2(1-\rho^2)}\\
&\quad - \frac{\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2}{\sigma^2(1-\rho^2)} + \frac{\Delta\rho^2\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2}{\gamma^2(1-\rho^2)^2},\\
\frac{\partial}{\partial\sigma^2}\log L_n(\theta) &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2}{2\sigma^2\gamma^2(1-\rho^2)},
\end{aligned}$$

whose zeros provide the MLEs in Eqs. (7)–(10), by observing that setting $\frac{\partial}{\partial\sigma^2}\log L_n(\theta)$ to zero implies that the first and fourth terms in $\frac{\partial}{\partial\beta}\log L_n(\theta)$ cancel out. It requires that $\sum_{i=1}^{n}(X_i - \hat\alpha)(X_{i-1} - \hat\alpha) > 0$, otherwise the MLE does not exist.

5.2. Asymptotic variances of the MLEs in the Ornstein-Uhlenbeck process

The Fisher Information $\mathcal{I}$ of the MLEs equals minus the expectation of the Hessian $\mathcal{H}$ of the log-likelihood function. For the OU log-likelihood, the elements of $\mathcal{H}$ are given by

$$\begin{aligned}
\frac{\partial^2}{\partial\alpha^2}\log L_n(\theta) &= -\frac{n(1-\rho)}{\gamma^2(1+\rho)},\\
\frac{\partial^2}{\partial\alpha\beta}\log L_n(\theta) &= \sum_{i=1}^{n}\left(C_1(X_{i-1}-\alpha) + C_2(X_i - X_{i-1}\rho - \alpha(1-\rho))\right)\\
\frac{\partial^2}{\partial\alpha\sigma^2}\log L_n(\theta) &= C_3\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho)),
\end{aligned}$$

$$\frac{\partial^2}{\partial\beta^2}\log L_n(\theta) = -\frac{n}{2\beta^2} + \frac{2n\Delta^2\rho^2}{(1-\rho^2)^2} + C_4\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))(X_{i-1}-\alpha) - \frac{\Delta^2\rho^2}{\gamma^2(1-\rho^2)}\sum_{i=1}^{n}(X_{i-1}-\alpha)^2$$
$$+\frac{2\Delta\rho^2}{\gamma^2(1-\rho^2)^2}\left(\frac{1}{\beta} - \frac{\Delta(1+\rho^2)}{1-\rho^2}\right)\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2$$
$$\frac{\partial^2}{\partial\beta\sigma^2}\log L_n(\theta) = C_5\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))(X_{i-1}-\alpha)$$
$$+\frac{1}{\sigma^2\gamma^2(1-\rho^2)}\left(\frac{1}{2\beta} - \frac{\Delta\rho^2}{1-\rho^2}\right)\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2,$$
$$\frac{\partial^2}{\partial(\sigma^2)^2}\log L_n(\theta) = \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(X_i - X_{i-1}\rho - \alpha(1-\rho))^2}{\sigma^4\gamma^2(1-\rho^2)},$$

where $C_i, i = 1,\dots,5$, are deterministic constants that will disappear when taking expectations. Thus, the Fisher Information is

$$\mathcal{I} = -\mathbb{E}\mathcal{H} = n\begin{bmatrix} \frac{(1-\rho)}{\gamma^2(1+\rho)} & 0 & 0 \\ 0 & \frac{1}{2\beta^2} - \frac{2\Delta\rho^2}{\beta(1-\rho^2)} + \frac{\Delta^2\rho^2(1+\rho^2)^2}{(1-\rho^2)^2} & \frac{1}{\sigma^2}\left(\frac{1}{2\beta} - \frac{\Delta\rho^2}{1-\rho^2}\right) \\ 0 & \frac{1}{\sigma^2}\left(\frac{1}{2\beta} - \frac{\Delta\rho^2}{1-\rho^2}\right) & \frac{1}{2\sigma^4} \end{bmatrix}.$$

The inverse of the Fisher Information provides the asymptotic covariance matrix,

$$\frac{1}{n}\begin{bmatrix} \frac{\gamma^2(1+\rho)}{(1-\rho)} & 0 & 0 \\ 0 & \frac{1-\rho^2}{\Delta^2\rho^2} & \frac{\sigma^2}{\Delta}\left(\frac{1-\rho^2}{\Delta\beta\rho^2} - 2\right) \\ 0 & \frac{\sigma^2}{\Delta}\left(\frac{1-\rho^2}{\Delta\beta\rho^2} - 2\right) & \sigma^4\left(\frac{1-\rho^2}{\beta^2\Delta^2\rho^2} - \frac{4}{\beta\Delta} + \frac{2(1+\rho^2)}{1-\rho^2}\right) \end{bmatrix}.$$

The diagonal elements provide the asymptotic variances of α, β and σ^2, respectively. The asymptotic covariance matrix is for $\Delta\beta \ll 1$ approximately equal to

$$\frac{1}{n}\begin{bmatrix} \frac{\sigma^2}{\Delta\beta^2} & 0 & 0 \\ 0 & \frac{2\beta}{\Delta} & 2\beta\sigma^2 \\ 0 & 2\beta\sigma^2 & 2\sigma^4 \end{bmatrix}.$$

Note that the variances of $\hat{\alpha}$ and $\hat{\beta}$ are on the order of $1/n\Delta$, whereas the variance of $\hat{\sigma}^2$ is on the order of $1/n$ and thus converges faster, and that $\hat{\alpha}$ is asymptotically independent of the other two.

5.3. Taylor expansion of the estimator of β in the Ornstein-Uhlenbeck process

Here, we develop Taylor expansions up to second order of $\hat{\beta}$ and $\hat{\beta}^2$ in (12) with α fixed at its true value around the mean values of Z and Y in (11). First, we note that by the tower property of conditional expectations and using (4),

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{i=1}^{n}\mathbb{E}(X_i-\alpha)(X_{i-1}-\alpha)\\ &= \sum_{i=1}^{n}\mathbb{E}(\mathbb{E}(X_i-\alpha)(X_{i-1}-\alpha)|X_{i-1})\\ &= \sum_{i=1}^{n}\mathbb{E}((\mathbb{E}(X_i|X_{i-1})-\alpha)(X_{i-1}-\alpha))\\ &= \sum_{i=1}^{n}\mathbb{E}((X_{i-1}\rho+\alpha(1-\rho)-\alpha)(X_{i-1}-\alpha))\\ &= \rho\sum_{i=1}^{n}\mathbb{E}(X_{i-1}-\alpha)^2\\ &= \rho\mathbb{E}[Y]. \end{aligned} \tag{29}$$

Define $f(z,y) = -\log(z/y)/\Delta$, then $\hat{\beta} \approx f(Z,Y)$. A Taylor expansion up to second order of $f(z, y)$ around $\mathbb{E}[Z]$ and $\mathbb{E}[Y]$ yields

$$f(z,y) \approx -\frac{1}{\Delta}\left(\log\rho + \frac{(z-\mathbb{E}[Z])}{\mathbb{E}[Z]} - \frac{(y-\mathbb{E}[Y])}{\mathbb{E}[Y]} + \frac{1}{2}\left(-\frac{(z-\mathbb{E}[Z])^2}{\mathbb{E}[Z]^2} + \frac{(y-\mathbb{E}[Y])^2}{\mathbb{E}[Y]^2} + 0\right)\right).$$

Taking expectations and using (29) yields

$$\mathbb{E}f(Z,Y) \approx \beta + \frac{1}{2\Delta}\left(\frac{\mathrm{V}[Z]}{\rho^2\mathbb{E}[Y]^2} - \frac{\mathrm{V}[Y]}{\mathbb{E}[Y]^2}\right), \tag{30}$$

which is Eq. (13). Now define $g(z,y) = (\log(z/y)/\Delta)^2$. A Taylor expansion up to second order of $g(z, y)$ around $\mathbb{E}[Z]$ and $\mathbb{E}[Y]$ yields

$$g(z,y) \approx \frac{1}{\Delta^2}\left((\log\rho)^2 + \frac{2\log\rho}{\mathbb{E}[Z]}(z-\mathbb{E}[Z]) - \frac{2\log\rho}{\mathbb{E}[Y]}(y-\mathbb{E}[Y]) + \right.$$
$$\left.\frac{1}{2}\left(\frac{2-2\log\rho}{\mathbb{E}[Z]^2}(z-\mathbb{E}[Z])^2 + \frac{2+2\log\rho}{\mathbb{E}[Y]^2}(y-\mathbb{E}[Y])^2 - \frac{2}{\mathbb{E}[Z]\mathbb{E}[Y]}(z-\mathbb{E}[Z])(y-\mathbb{E}[Y])\right)\right).$$

Taking expectations and using (29) yields

$$\mathbb{E}g(Z,Y) \approx \beta^2 + \frac{1}{\Delta^2\mathbb{E}[Y]^2}\left(\frac{1}{\rho^2}\mathrm{V}[Z] + \mathrm{V}[Y] - \frac{2}{\rho}\mathrm{Cov}[Z,Y]\right) + \frac{\log\rho}{\Delta^2\mathbb{E}[Y]^2}\left(\mathrm{V}[Y] - \frac{\mathrm{V}[Z]}{\rho^2}\right).$$

Taking the square of (30) and retaining terms of up to order two yields

$$(\mathbb{E}\hat{\beta})^2 \approx (\mathbb{E}f(Z,Y))^2 \approx \beta^2 - \frac{\beta}{\Delta\mathbb{E}[Y]^2}\left(\mathrm{V}[Y] - \frac{\mathrm{V}[Z]}{\rho^2}\right)$$

so we obtain Eq. (14)

$$\mathrm{V}[\hat{\beta}] = \mathbb{E}(\hat{\beta})^2 - (\mathbb{E}\hat{\beta})^2 \approx \frac{1}{\Delta^2\mathbb{E}[Y]^2}\left(\frac{1}{\rho^2}\mathrm{V}[Z] + \mathrm{V}[Y] - \frac{2}{\rho}\mathrm{Cov}[Z,Y]\right),$$

where we have used that $\log\rho = -\beta\Delta$.

5.4. Moments of equation (11)

To evaluate (13) and (14), moments of Eq. (11) have to be inserted. After a considerable amount of algebra, using the conditional moments (4) and (6) in (11), one obtains

$$\frac{\mathbb{E}[Y]}{\gamma^2} = n + \frac{1-\rho^{2n}}{1-\rho^2}(k^2-1), \quad \mathbb{E}[Z] = \rho\mathbb{E}[Y],$$
$$\frac{\mathrm{V}[Y]}{\gamma^4} = \frac{n\left(2(1+\rho^2) - 8(k^2-1)\rho^{2n}\right)}{1-\rho^2} + \frac{4(k^2-1)(\rho^2+\rho^{2n})(1-\rho^{2n})}{(1-\rho^2)^2} - \frac{4\rho^2(1-\rho^{2n}) + 2(1-\rho^{2n})^2}{(1-\rho^2)^2},$$
$$\frac{\mathrm{V}[Z]}{\gamma^4} = \frac{n\left(1+4\rho^2-\rho^4 - 4(k^2-1)\rho^{2n}(1+\rho^2)\right)}{1-\rho^2} + \frac{(k^2-1)\left((1+\rho^2)^2 - \rho^{2n}(1-\rho^2)^2 - 4\rho^{4n+2}\right)}{(1-\rho^2)^2} - \frac{4\rho^2(1-\rho^{2n}) + 2\rho^2(1-\rho^{2n})^2}{(1-\rho^2)^2},$$
$$\frac{\mathrm{Cov}[Z,Y]}{2\rho\gamma^4} = \frac{n\left(2-(k^2-1)(1+3\rho^2)\rho^{2(n-1)}\right)}{1-\rho^2} + \frac{(k^2-1)(1+\rho^2+\rho^{2n}(1-\rho^2) - 2\rho^{4n})}{(1-\rho^2)^2} - \frac{(1+\rho^2)(1-\rho^{2n}) + (1-\rho^{2n})^2}{(1-\rho^2)^2}.$$

5.5. Conditional variance of the Jacobi diffusion

The generator of the Jacobi diffusion is given by

$$L(\cdot) = -\beta(x-\alpha)\frac{d}{dx}(\cdot) + \frac{\sigma^2}{2}x(1-x)\frac{d^2}{dx^2}(\cdot). \tag{31}$$

Eigenfunctions $\phi_j(x)$ and eigenvalues λ_j of generator (31) are defined by

$$L\phi_j(x) = -\lambda_j\phi_j(x). \tag{32}$$

The eigenfunctions of (31) are the Jacobi polynomials [35], from where the diffusion has its name. The eigenfunctions and the corresponding eigenvalues can be used to find conditional moments and define explicit martingale estimating functions using the following result [35],

$$\mathbb{E}\phi_j(X_t|X_0 = x) = e^{-\lambda_j t}\phi_j(x). \tag{33}$$

To find the conditional variance, we need the eigenfunctions being the first two Jacobi polynomials,

$$\phi_1(x) = x - \alpha, \tag{34}$$

$$\phi_2(x) = x^2 - \frac{2\alpha\beta+\sigma^2}{\beta+\sigma^2}x + \frac{\alpha\beta(2\alpha\beta+\sigma^2)}{(\beta+\sigma^2)(2\beta+\sigma^2)} \tag{35}$$

with corresponding eigenvalues $\lambda_1 = \beta$ and $\lambda_2 = 2\beta + \sigma^2$, which can be checked by inserting in Eq. (32). Using (33) we obtain

$$\mathbb{E}(X_i - \alpha|X_{i-1}) = e^{-\beta\Delta}(X_{i-1} - \alpha) = \rho(X_{i-1} - \alpha), \tag{36}$$

which is Eq. (4), and

$$\mathbb{E}(X_i^2 - \frac{2\alpha\beta+\sigma^2}{\beta+\sigma^2}X_i + \frac{\alpha\beta(2\alpha\beta+\sigma^2)}{(\beta+\sigma^2)(2\beta+\sigma^2)}|X_{i-1}) = e^{-(2\beta+\sigma^2)t}\left(X_{i-1}^2 - \frac{2\alpha\beta+\sigma^2}{\beta+\sigma^2}X_{i-1} + \frac{\alpha\beta(2\alpha\beta+\sigma^2)}{(2\beta+\sigma^2)+(\beta+\sigma^2)}\right).$$

Re-arranging and using Eqs. (2), (3) and (36), we obtain

$$\mathbb{E}(X_i^2|X_{i-1}) = \rho^2 e^{-\sigma^2\Delta}\left(X_{i-1}^2 - \frac{\alpha+\gamma^2}{\frac{1}{2}+\gamma^2}\left(X_{i-1} - \frac{\frac{1}{2}\alpha}{1+\gamma^2}\right)\right) + \frac{\alpha+\gamma^2}{\frac{1}{2}+\gamma^2}\left(\rho(X_{i-1}-\alpha) + \alpha\frac{\frac{1}{2}+\gamma^2}{1+\gamma^2}\right). \tag{37}$$

Finally, the conditional variance can be found by $\mathrm{V}(X_i|X_{i-1}) = \mathbb{E}(X_i^2|X_{i-1}) - (\mathbb{E}(X_i|X_{i-1}))^2$ using Eqs. (36) and (37), which is Eq. (22).

5.6. Martingale estimating function for the Jacobi diffusion

To estimate the two drift parameters α and β (or equivalently ρ), we will use an approximate optimal linear martingale estimating function [35], given by the zeros of the vector function

$$G(\alpha,\beta) = \sum_{i=1}^{n}\frac{\partial\mu(X_{i-1})}{\sigma^2 X_{i-1}(1-X_{i-1})}(X_i - \mathbb{E}(X_i|X_{i-1}))$$

where $\mu(x) = -\beta(x-\alpha)$ is the drift function, and the partial derivative is with respect to α and β. We obtain

$$G_1(\alpha,\beta) = \sum_{i=1}^{n}\frac{\beta}{\sigma^2 X_{i-1}(1-X_{i-1})}(X_i - \alpha - \rho(X_{i-1}-\alpha)),$$

$$G_2(\alpha,\beta) = \sum_{i=1}^{n}\frac{-(X_{i-1}-\alpha)}{\sigma^2 X_{i-1}(1-X_{i-1})}(X_i - \alpha - \rho(X_{i-1}-\alpha)).$$

The estimators are given by $G(\hat{\alpha},\hat{\rho}) = 0$, which are Eqs. (26) and (27). To estimate σ^2, we use an approximate moment estimator, where the infinitesimal variance (the diffusion function) is estimated, Eq. (28). This will be exact in the limit $\Delta \to 0$, but will introduce a minor discretization bias for finite samples, as seen in Fig. 5.

Acknowledgments

The work was part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. PL is supported by the Institute of Physiology RVO:67985823 and by the Czech Science Foundation project 17-06943S.

References

[1] Berg RW, Ditlevsen S. Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations. J Neurophysiol 2013;110(4):1021–34. doi:10.1152/jn.00006.2013.
[2] Bhattacharjee A, Das MK. Emergent dynamics of spiking neurons with fluctuating threshold. Commun Nonlinear Sci NumerSimul 2017;46:126–34. doi:10.1016/j.cnsns.2016.10.022.
[3] Cox J, Ingersoll J, Ross S. A theory of the term structure of interest-rates. Econometrica 1985;53(2):385–407. doi:10.2307/1911242.
[4] Ditlevsen S, Ditlevsen O. Parameter estimation from observations of first-passage times of the Ornstein-Uhlenbeck process and the Feller process. Probabilistic Engineering Mechanics 2008;23(2–3):170–9. doi:10.1016/j.probengmech.2007.12.024. 5th International Conference on Computational Stochastic Mechanics, Rhodes, GREECE, MAY 21–23, 2006

[5] Ditlevsen S, Greenwood P. The Morris-Lecar neuron model embeds a leaky integrate-and-fire model. J Math Biol 2013;67(2):239–59. doi:10.1007/s00285-012-0552-7.
[6] Ditlevsen S, Lansky P. Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model. Phys Rev E 2005;71:011907.
[7] Ditlevsen S, Lansky P. Estimation of the input parameters in the Feller neuronal model. Phys Rev E 2006;73:061910.
[8] Ditlevsen S, Lansky P. Only through perturbation can relaxation times be estimated. Phys Rev E 2012;86(5, 1):050102.
[9] Ditlevsen S, Samson A. Introduction to stochastic models in biology. In: Stochastic biomathematical models: with applications to neuronal modeling. In: Lecture Notes in Mathematics, 2058. Springer Berlin Heidelberg; 2013. p. 3–35. doi:10.1007/978-3-642-32157-3_1. 978-3-642-32157-3; 978-3-642-32156-6.
[10] Ditlevsen S, Samson A. Estimation in the partially observed stochastic Morris-Lecar neuronal model with particle filter and stochastic approximation methods. Ann Appl Stat 2014;8(2):674–702. doi:10.1214/14-AOAS729.
[11] D'Onofrio G, Lansky P, Tamborrino M. Inhibition enhances the coherence in the Jacobi neuronal model. Chaos, Soliton Fractals 2019;128:108–13.
[12] D'Onofrio G, Tamborrino M, Lansky P. The Jacobi diffusion process as a neuronal model. Chaos 2018;28(10). doi:10.1063/1.5051494.
[13] Forman JL, Sørensen M. The Pearson diffusions: a class of statistically tractable diffusion processes. Scand J Stat 2008;35(3):438–65.
[14] Gardiner C. Handbook of stochastic methods for physics, chemistry and the natural sciences. 2nd edn. Springer, Berlin; 1990.
[15] Gerstner W, Kistler WM. Spiking neuron models - single neurons, populations, plasticity. Cambridge University Press, Cambridge UK; 2002.
[16] Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal dynamics - from single neurons to networks and models of cognition. Cambridge University Press, Cambridge UK; 2014.
[17] Harrison L, David O, Friston K. Stochastic models of neuronal dynamics. Philos Trans R Soc Lond B Biol Sci 2005;360:1075–91.
[18] Hooker G, Lin KK, Rogers B. Control theory and experimental design in diffusion processes. SIAM-ASA Journal on Uncertainty Quantification 2015;3(1):234–64. doi:10.1137/140962280.
[19] Iolov A, Ditlevsen S, Longtin A. Optimal design for estimation in diffusion processes from first hitting times. SIAM-ASA J Uncertain Quantificat 2017;5(1):88–110. doi:10.1137/16M1060376.
[20] Jahn P, Berg RW, Hounsgaard J, Ditlevsen S. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. J Comput Neurosci 2011;31(3):563–79. doi:10.1007/s10827-011-0326-z.
[21] Karlin S, Taylor H. A second course in stochastic processes. Academic Press, San Diego; 1981.
[22] Koch C, Rapp M, Segev I. A brief history of time (constants). Cerebral Cortex 1996;6(2):93–101. doi:10.1093/cercor/6.2.93.
[23] Lanska V, Lansky P. Input parameters in a one-dimensional neuronal model with reversal potentials. Biosystems 1998;48:123–9.
[24] Lansky P. Inference for the Diffusion-Models of Neuronal-Activity. Mathematical Biosciences 1983;67(2):247–60. doi:10.1016/0025-5564(83)90103-7.
[25] Lansky P, Ditlevsen S. A review of the methods for signal estimation in stochastic diffusion leaky integrate-and-fire neuronal models. Biological Cybernetics 2008;99(4–5):253–62. doi:10.1007/s00422-008-0237-x.
[26] Larsen KS, Sørensen M. Diffusion models for exchange rates in a target zone. Math Finance 2007;17(2):285–306. doi:10.1111/j.1467-9965.2006.00304.x.
[27] Lindner B, Garcia-Ojalvo J, Neiman A, Schimansky-Geier L. Effects of noise in excitable systems. Phys Report-Rev Sec Phys Lett 2004;392(6):321–424. doi:10.1016/j.physrep.2003.10.015.
[28] Longtin A. Stochastic resonance in neuron models. J Stat Phys 1993;70(1–2):309–27. doi:10.1007/BF01053970. NATO Advanced Research Workshop : Stochastic Resonance in Physics and Biology, San Diego, CA, MAR 30-APR 03, 1992
[29] Mullooly J. Maximum likelihood estimation for stochastic first order reactions. Bull Math Biophys 1971;33(1):83.
[30] Pikovsky A, Kurths J. Coherence resonance in a noise-driven excitable system. Phys Rev Lett 1997;78(5):775–8. doi:10.1103/PhysRevLett.78.775.
[31] Prakasa Rao B. Statistical inference for diffusion type processes. Arnold; 1999.
[32] Sacerdote L, Giraudo MT. Stochastic integrate and fire models: a review on mathematical methods and their applications. In: Stochastic biomathematical models: with applications to neuronal modeling. In: Lecture Notes in Mathematics, 2058. Springer Berlin Heidelberg; 2013. p. 99–148. doi:10.1007/978-3-642-32157-3_1. 978-3-642-32157-3; 978-3-642-32156-6
[33] Shekofteh Y, Jafari S, Sprott JC, Golpayegani SMRH, Almasganj F. A gaussian mixture model based cost function for parameter estimation of chaotic biological systems. Commun Nonlinear Sci Numer Simul 2015;20(2):469–81. doi:10.1016/j.cnsns.2014.05.028.
[34] Shoji I. Detecting the sampling rate through observations. Commun Nonlinear Sci NumerSimul 2018;62:445–53. doi:10.1016/j.cnsns.2018.02.031.
[35] Sørensen M. Statistical methods for stochastic differential equations. Chapman & Hall; 2012. p. 1–99.
[36] Tamborrino M, Ditlevsen S, Markussen B, Kyllingsbaek S. Gaussian counter models for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. J Math Psychol 2017;79:85–103. doi:10.1016/j.jmp.2017.02.003.
[37] Tang CY, Chen SX. Parameter estimation and bias correction for diffusion processes. J Econometr 2009;149(1):65–81.
[38] Thorbergsson L, Hooker G. Experimental design for partially observed Markov decision processes. SIAM-ASA J Uncertain Quantificat 2018;6(2):549–67. doi:10.1137/16M1084924.
[39] Wang Q, Zhang H, Perc M, Chen G. Multiple firing coherence resonances in excitatory and inhibitory coupled neurons. Commun Nonlinear Sci NumerSimul 2012;17(10):3979–88. doi:10.1016/j.cnsns.2012.02.019.
[40] Wang X, Phillips PCB, Yu J. Bias in estimating multivariate and univariate diffusions. J Econometr 2011;161(2):228–45.
[41] Yu J. Bias in the estimation of the mean reversion parameter in continuous time models. J Econometr 2012;169(1):114–22.