COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2018, VOL. 47, NO. 3, 721–737
https://doi.org/10.1080/03610918.2017.1291960

Parameter identification for a stochastic logistic growth model with extinction

Fabien Campillo[a], Marc Joannides[b], and Irène Larramendy-Valverde[b]

[a]Inria, Lemon Project, rue Saint-Priest, Montpellier, France; [b]Université Montpellier, IMAG, UMR, Place Eugène Bataillon, Case courrier, Montpellier, France

ABSTRACT

We consider a stochastic logistic growth model given by a stochastic differential equation, for which extinction can occur. We first propose appropriate adaptation of some standard inference methods when the process is observed at discrete time. Second, we show that the individual birth and death can be identified separately to some extent.

ARTICLE HISTORY
Received 1 March 2016
Accepted 31 January 2017

KEYWORDS
Diffusion processes; Extinction; Fokker–Planck equation; Estimation; Logistic model; Monte Carlo

MATHEMATICS SUBJECT CLASSIFICATION
Primary 62M05; Secondary 62P10

1. Introduction

During the last 20 years, stochastic differential equations (SDEs) have gained a high interest in population dynamics, biology, ecology, or environmental sciences. Ordinary differential equations (ODEs) are still widely used, but are not suited to capture the random fluctuations caused by the demographic and environmental noises. SDEs can be derived from ODEs by just adding a diffusion term to model the effect of the noise on the dynamics. Alternatively, they can also be derived as diffusion approximations of pure jump Markov processes. No matter which way it was obtained, the stochastic process can behave differently depending on the diffusion coefficient, particularly as it pertains to extinction in population dynamics. Beside, both deterministic and stochastic models are known, up to a set of parameters, to be estimated from observed data. From this perspective, the diffusion term (i.e., the noise) should not be considered as harmful but as source of some valuable information.

Consider the classical deterministic logistic population growth model:

$$\dot{x}(t) = r\,x(t)\left(1 - \frac{x(t)}{K}\right), \quad x(0) = x_0 > 0,\ t \geq 0 \tag{1}$$

where $x(t) \geq 0$ is the density of some population, $r > 0$ the net growth rate, and $K > x_0$ the carrying capacity of the environment. When they are not known, the parameters r and K in (1) are usually identified using least-squares methods, based on a dataset of discrete observations.

CONTACT Marc Joannides marc.joannides@umontpellier.fr Université Montpellier, IMAG, UMR 5149, Place Eugène Bataillon, Case courrier 051, Montpellier, 34095 Cedex 5 France.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

This parameterization fails to render the complexity of the growth mechanism. Indeed, the per capita growth rate results from the balance between a birth rate $B(x)$ and a death rate $D(x)$:

$$\dot{x}(t) = (B(x(t)) - D(x(t)))\, x(t).$$

A possible way to arrive at (1) is to postulate the particular form:

$$B(x) = \lambda \quad \text{and} \quad D(x) = \mu + \alpha\, x$$

where λ, μ, and α are the meaningful non-negative parameters. The competition for some critical resource is encoded in the death rate only, through the logistic coefficient α, but there are other choices. This yields (1) with $r = \lambda - \mu$ and $K = \frac{\lambda-\mu}{\alpha}$. It is easily seen that the model (1) does not allow extinction of the population since the known analytic solution cannot vanish. One would have to discard datasets where extinction actually occurs. Second, λ and μ appear only through their difference. Consequently, even if the coefficient α is known, only the net growth rate $\lambda - \mu$ can be identified. A stochastic version of this model, obtained by diffusion approximation, has been proposed in Campillo et al. (2016) as:

$$\mathrm{d}X_t = (\lambda - \mu - \alpha\, X_t)\, X_t\, \mathrm{d}t + \rho \sqrt{(\lambda + \mu + \alpha\, X_t)\, X_t}\, \mathrm{d}B_t \tag{2}$$

where $\rho > 0$ is the noise intensity that relates to the order of magnitude of the underlying population; B_t is a standard Brownian motion; the law of the initial condition X_0 is supported by $\mathbb{R}_+$; B_t and X_0 are supposed independent. Note that (2) can be reparameterized as:

$$\mathrm{d}X_t = r\left(1 - \frac{X_t}{K}\right) X_t\, \mathrm{d}t + \rho \sqrt{r'\left(1 + \frac{X_t}{K'}\right) X_t}\, \mathrm{d}B_t$$

with $r = \lambda - \mu$, $K = \frac{r}{\alpha}$, $r' = \lambda + \mu$ and $K' = \frac{r'}{\alpha}$, so that its instantaneous mean is the same as in (1). The analysis carried over in Campillo et al. (2016) established that, for this model, extinction occurs almost surely in finite time. More precisely, if $\tau_0 \overset{\text{def}}{=} \inf\{t \geq 0\,;\ X_t = 0\}$ is the extinction time, then for all $x \geq 0$, $\mathbb{P}_x(\tau_0 < \infty) = 1$ where $\mathbb{P}_x$ is the probability measure such that $X_0 = x$ and for any values of the parameters. Other stochastic models derived from or leading to the same deterministic model (1), but with different qualitative behaviors, have been proposed in, e.g., Heydari et al. (2014) or Schurz (2007). Depending on the diffusion coefficient chosen, the ultimate extinction may or may not occur. Notice that in (2), the stochastic integral is understood in the Ito sense, since this equation arises from a diffusion approximation of a pure jump Markov process.

Two separate questions are addressed in this work. First, we provide a technique in order to adapt the classical inference methods to face the extinction problem. Indeed, these methods are designed to approximate probability distributions that are absolutely continuous with respect to the Lebesgue measure. This assumption is not satisfied by the process (2). Second, we show that, for this model, the death and growth rates can be identified separately with some degree of precision, thanks to the specific form of the diffusion coefficient.

Section 2 briefly reviews the essential results of Campillo et al. (2016) on the main properties of the model. We recall the stochastic differential equation defining our model and we introduce the *complete* Fokker–Planck equation (CFPE) governing the evolution of the associated diffusion process. We then derive the expression of a likelihood function, when discrete time observations of the process are available. The numerical methods specifically designed to handle extinction are presented in Section 3. Finally, all of these methods are evaluated through numerical experiments in Section 4. In particular, we investigate numerically the

second objective of this article, that is, the ability of the methods to discriminate between pairs $(\lambda,\ \mu)$ for a given difference r.

2. Statistical model

Let us denote by $\theta \in \Theta = (0, \infty)^p$, the unknown parameter to be estimated from a discrete sample of one trajectory of model (2). The parameter θ may include some or all parameters $(\lambda, \mu, \alpha, \rho)$ and may also appear in the initial distribution law. Hence, (2) can be rewritten:

$$\mathrm{d}X_t = b^\theta(X_t)\,\mathrm{d}t + \sigma^\theta(X_t)\,\mathrm{d}B_t, \quad 0 \le t \le T, \quad X_0 \sim \pi_0^\theta(\mathrm{d}x), \tag{3}$$

where $b^\theta(x) = (\lambda - \mu - \alpha\, x)\, x$ and $\sigma^\theta(x) = \rho\,\sqrt{(\lambda + \mu + \alpha\, x)\, x}$; π_0^θ is the initial distribution. We denote by $\mathbb{P}^\theta$ the underlying distribution of the process $\{X_t\}_{0\le t\le T}$. Observations from the SDE (3) are available as:

$$\xi_k^{\mathrm{o}} = X_{t_k^{\mathrm{o}}}, \quad k = 0, \ldots, M,$$

where, for sake of simplicity, the observation instants are equally spaced, i.e., $t_k^{\mathrm{o}} = k\,\Delta$ with $\Delta \overset{\text{def}}{=} T/M$. By the Markov property, the distribution of the measurements vector is:

$$\mathbb{P}^\theta(\xi_0^{\mathrm{o}} \in \mathrm{d}\xi_0, \ldots, \xi_M^{\mathrm{o}} \in \mathrm{d}\xi_M) = \pi_0^\theta(\mathrm{d}\xi_0) \prod_{k=0}^{M-1} Q_\Delta^\theta(\mathrm{d}\xi_{k+1} \mid \xi_k).$$

All likelihood based procedures are grounded on the fact that the transition kernel $Q_\Delta^\theta(\mathrm{d}x \mid y)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^+$. This is obviously not the case for model (2), since it gives positive probability to the boundary point 0 corresponding to extinction. Instead, the following decomposition holds:

$$Q_\Delta^\theta(\mathrm{d}y \mid x) = E_\Delta^\theta(x)\,\delta_0(\mathrm{d}y) + p_\Delta^\theta(y \mid x)\,\mathrm{d}y \tag{4}$$

where $E_t^\theta(x)$ is the *extinction probability* starting from x. Since $Q_t^\theta(\mathrm{d}y \mid x)$ is a probability measure for any $x \ge 0$, $E_t^\theta(x)$ and $p_t^\theta(\cdot \mid x)$ are linked together by

$$E_t^\theta(x) = 1 - \int_0^\infty p_t^\theta(y \mid x)\,\mathrm{d}y.$$

Defining the reference measure on $\mathbb{R}_+$

$$m(\mathrm{d}y) \overset{\text{def}}{=} \delta_0(\mathrm{d}y) + \mathrm{d}y$$

we get that $Q_t^\theta(\mathrm{d}y \mid x)$ is absolutely continuous with respect to $m(\mathrm{d}y)$ with density:

$$q_t^\theta(y \mid x) \overset{\text{def}}{=} \begin{cases} E_t^\theta(x), & \text{if } y = 0, \\ p_t^\theta(y \mid x), & \text{otherwise.} \end{cases} \tag{5a}$$

We also suppose that the initial distribution π_0^θ is absolutely continuous with respect to $m(\mathrm{d}y)$, and we let:

$$q_0^\theta(y) \overset{\text{def}}{=} \frac{\pi_0^\theta(\mathrm{d}y)}{m(\mathrm{d}y)} = \begin{cases} E_0^\theta, & \text{if } y = 0, \\ p_0^\theta(y), & \text{otherwise.} \end{cases} \tag{6}$$

Our statistical model is therefore dominated by the product measure $m(\mathrm{d}\xi_0)\dots m(\mathrm{d}\xi_M)$, and a likelihood function is given by

$$\mathcal{L}(\theta) = q_0^\theta(\xi_0)\prod_{k=0}^{M-1} q_\Delta^\theta(\xi_{k+1}\,|\,\xi_k). \tag{7}$$

There is no explicit expression for the transition density $q_\Delta^\theta(x\,|\,y)$. Throughout the following, we will rely on the following characterization, proved in Campillo et al. (2016):

Theorem 2.1. *Let* $a^\theta(y) \overset{\text{def}}{=} [\sigma^\theta(y)]^2$ *and* $\mathcal{A}$ *the infinitesimal generator of the Markov process (3):*

$$\mathcal{A}f(y) \overset{\text{def}}{=} b^\theta(y)\,f'(y) + \frac{1}{2}\,a^\theta(y)\,f''(y)\,, \quad \forall f \in \mathcal{C}_K^\infty(\mathbb{R}_+) \tag{8}$$

where $\mathcal{C}_K^\infty(\mathbb{R}_+)$ *is the set of functions differentiable for all degrees of differentiation and with compact support included in* $[0, +\infty)$. *Then for any fixed* $x > 0$, *the transition density* $q_t^\theta(y\,|\,x)$ *satisfies the CFPE:*

$$\frac{\partial p_t^\theta(y\,|\,x)}{\partial t} = \mathcal{A}^* p_t^\theta(y\,|\,x)\,, \quad \lim_{t\downarrow 0} p_t^\theta(y\,|\,x)\,\mathrm{d}y = \delta_x(\mathrm{d}y), \tag{9a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}E_t^\theta(x) = \frac{1}{2}\,\frac{\partial\,(a(y)\,p_t^\theta(y\,|\,x))}{\partial y}\bigg|_{y=0} p_t^\theta(0\,|\,x), \quad E_0^\theta(x) = 0. \tag{9b}$$

As seen from (9), the law of the process is described by a set of two equations of different kind. Equation (9a) is a PDE in a classical sense describing the evolution of the process *before extinction*. It follows that $y \mapsto p_t^\theta(y\,|\,x)$ is the density of a *defective distribution*. Equation (9b) links the rate of extinction to the defective density. This equation has been extensively studied by Feller (1952) in a general setting.

Remark 2.1. We already know that $E_t^\theta(x)$ increases to 1, so that $Q_t^\theta(\cdot\,|\,x)$ will eventually degenerate to the Dirac mass at 0. We note that this convergence may be slow, i.e., that the contribution of the Dirac mass in (4) may not be significant for the time scale at which the system is observed. This feature is discussed in Grasman and van Herwaarden (1999).

3. Numerical approximations

Estimation of the parameter θ by a likelihood-based procedure requires the computation of the transition density for which we have no analytic expression. The exact maximum likelihood is therefore out of reach and one has to employ approximation methods. A number of such methods can be found in the most considered regular case, where an absolutely continuous density exists, which satisfies the (simple) Fokker–Planck equation. Solving this equation by finite difference methods appears in Lo (1988) or Jensen and Poulsen (2002). Aït-Sahalia (2002) used Hermite expansion. Among other options, simulated maximum likelihood estimation (SMLE) using Monte Carlo samples has proved to be an efficient alternative to the discretized PDE approach. For a detailed review, see Hurn et al. (2007) and references therein or Fearnhead (2008) which includes many application examples. The importance sampling framework appeared in the SDE context with the work of Pedersen (1995) which was later improved by Durham and Gallant (2002). Pastorello and Rossi (2010) introduced the efficient importance sampling variant. The Bayesian context was also considered, see, e.g., Golightly and Wilkinson (2008).

None of the above procedures can be applied straightforwardly, due to the singular part of the transition density. In this section, we design modifications of some of these methods enabling them to estimate jointly both absolutely continuous and singular parts. Approximating the solution of (5) by the use of a specific difference scheme has been considered in Campillo et al. (2016) and is recalled in the Appendix. We now focus on the Monte Carlo approach.

The numerical approximations presented hereafter involve simulation of a N-samples with common law $Q_\Delta^\theta(\mathrm{d}y \,|\, x)$, for many different initial conditions x. In our case, we will not be able to draw random variates from $Q_\Delta^\theta(\mathrm{d}y \,|\, x)$ exactly, but only from approximating distributions. The simplest algorithm for simulating trajectories of (2) is the Euler–Maruyama scheme, restricted to non-negative values, that is:

$$\bar{X}_{t_{k+1}} = \max(0, \bar{X}_{t_k} + \delta\, b(\bar{X}_{t_k}) + \sqrt{\delta}\, \sigma(\bar{X}_{t_k})\, w_k), \quad k = 0, \ldots, n-1 \tag{10}$$

with $\bar{X}_0 = x$ and where w_k are i.i.d. $\mathcal{N}(0, 1)$. One iteration of the approximation scheme (10) amounts to drawing from the transition kernel $K_\delta(\mathrm{d}z \,|\, x)$ instead of $Q_\delta(\mathrm{d}z \,|\, x)$ where $K_\delta(\mathrm{d}z \,|\, x)$ is defined by:

$$K_\delta(\mathrm{d}z \,|\, x) \stackrel{\text{def}}{=} \begin{cases} e_\delta(x)\, \delta_0(\mathrm{d}z) + g_\delta(z \,|\, x)\, \mathrm{d}z, & \text{if } x > 0, \\ \delta_0(\mathrm{d}z), & \text{if } x = 0, \end{cases} \tag{11}$$

with

$$e_\delta(x) \stackrel{\text{def}}{=} 1 - \int_0^\infty g_\delta(z \,|\, x)\, \mathrm{d}z,$$

$$g_\delta(z \,|\, x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi\, \delta\, \sigma(x)}} \exp\left\{ -\frac{(z - x - \delta\, b(x))^2}{2\, \delta\, \sigma(x)} \right\} \mathbf{1}_{\mathbb{R}_+}(z).$$

According to Risken (1996, Section 4.4.1), this approximation is sound if δ is sufficiently small. The numerical scheme (10) produces samples of common law $Q_\Delta^{\mathrm{E}}(\mathrm{d}y \,|\, x)$:

$$Q_\Delta^{\mathrm{E}}(\mathrm{d}y \,|\, x) = \int_{\{y_1 \geq 0\}} \cdots \int_{\{y_{n-1} \geq 0\}} K_\delta(\mathrm{d}y_1 \,|\, x)\, K_\delta(\mathrm{d}y_2 \,|\, y_1) \cdots K_\delta(\mathrm{d}y \,|\, y_{n-1}), \tag{12}$$

which is an approximation of the true transition kernel $Q_\Delta(\mathrm{d}y \,|\, x)$. Notice that, using the semi-group and the Markov properties, we also have a similar decomposition:

$$Q_\Delta(\mathrm{d}y \,|\, x) = \int_{\{y_1 \geq 0\}} \cdots \int_{\{y_{n-1} \geq 0\}} Q_\delta(\mathrm{d}y_1 \,|\, x)\, Q_\delta(\mathrm{d}y_2 \,|\, y_1) \cdots Q_\delta(\mathrm{d}y \,|\, y_{n-1}).$$

Remark 3.1. The recent works Beskos and Roberts (2005) about the *Exact Algorithm* (EA) seem promising for drawing exactly from $Q_\Delta(\mathrm{d}y \,|\, x)$. To our knowledge, this algorithm cannot be applied directly to our specific case, due to the almost sure extinction. There are also other alternatives to (10), such as the Milstein scheme, see, e.g., Kloeden et al. (2003), or Euler–Maruyama scheme for killed diffusion, see Gobet (2001) and Casella and Roberts (2008). Modifications of these algorithms might be necessary to cope with the extinction problem.

Remark 3.2. If Δ is itself small enough, there would be no need to simulate the solution of (2) at intermediate time between 0 and Δ, since K_Δ has an explicit density with respect to m. However, for most applications, the observations are not available at such a high sampling rate.

3.1. Nonparametric estimation

A simple usage of the approximation scheme (10) is to produce a N–sample $\bar{X}_\Delta^{(1)}, \dots \bar{X}_\Delta^{(N)}$ of $Q_\Delta^{\mathrm{E}}(\mathrm{d}y \,|\, x)$. These simulated observations are then fed into a non-parametric estimate of the density $p_\Delta(y \,|\, x)$ at y, denoted by $\hat{p}_\Delta(y \,|\, x)$. Again, the case of extinction should be cared for by first discarding the values $\bar{X}_\Delta^{(i)} = 0$, if there are any, from the estimation. The resulting approximation reads:

$$q_\Delta(y \,|\, x) \simeq \begin{cases} \frac{N-N_s}{N}, & \text{if } y = 0, \\ \frac{N_s}{N} \hat{p}_\Delta(y \,|\, x), & \text{otherwise,} \end{cases}$$

where N_s is the random number of trajectories still alive at time Δ, that is:

$$N_s \stackrel{\text{def}}{=} \#\big\{i = 1, \dots, N;\ \bar{X}_\Delta^{(i)} \neq 0\big\}.$$

This approach, without extinction, is presented in Hurn et al. (2003). Its efficiency relies on that of the non-parametric estimation method used and is therefore subject to the classical problems of choice of bandwidth and leakage of mass in the inaccessible region ($\mathbb{R}_-$ in our case).

3.2. Pedersen method

It is possible to avoid the nonparametric estimation stage. Indeed from the Markov property:

$$Q_\Delta(\mathrm{d}y \,|\, x) = (Q_{\Delta-\delta} Q_\delta)(\mathrm{d}y \,|\, x) = \int_{\{y_{n-1} \geq 0\}} Q_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x)\, Q_\delta(\mathrm{d}y \,|\, y_{n-1}), \tag{13}$$

first approximate $Q_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x)$ by $Q_{\Delta-\delta}^{\mathrm{E}}(\mathrm{d}y_{n-1} \,|\, x)$, hence:

$$Q_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x) \simeq \frac{1}{N} \sum_{i=1}^{N} \delta_{\bar{X}_{t_{n-1}}^{(i)}}(\mathrm{d}y_{n-1})$$

where $\bar{X}_{t_{n-1}}^{(i)} \stackrel{\text{iid}}{\sim} Q_{\Delta-\delta}^{\mathrm{E}}(\mathrm{d}y_{n-1} \,|\, x)$, $i = 1, \dots, N$; then approximating $Q_\delta(\mathrm{d}y \,|\, y_{n-1})$ by $K_\delta(\mathrm{d}y \,|\, y_{n-1})$, leads to the following approximation of the kernel $Q_\Delta(\mathrm{d}y \,|\, x)$:

$$Q_\Delta^{\mathrm{P}}(\mathrm{d}y \,|\, x) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} K_\delta\big(\mathrm{d}y \,|\, \bar{X}_{t_{n-1}}^{(i)}\big) \tag{14}$$

see Fig. 1. Let us re-number the sampled trajectories so that the surviving ones correspond to $i = 1, \dots, N_s$, according to (11) we get:

$$\begin{aligned} Q_\Delta^{\mathrm{P}}(\mathrm{d}y \,|\, x) &= \frac{1}{N} \sum_{i=N_s+1}^{N} K_\delta(\mathrm{d}y \,|\, 0) + \frac{1}{N} \sum_{i=1}^{N_s} K_\delta\big(\mathrm{d}y \,|\, \bar{X}_{t_{n-1}}^{(i)}\big) \\ &= \frac{N-N_s}{N}\, \delta_0(\mathrm{d}y) + \frac{1}{N} \sum_{i=1}^{N_s} \Big[e_\delta\big(\bar{X}_{t_{n-1}}^{(i)}\big)\, \delta_0(\mathrm{d}y) + g_\delta\big(y \,|\, \bar{X}_{t_{n-1}}^{(i)}\big)\, \mathrm{d}y\Big] \\ &= \left[\frac{N-N_s}{N} + \frac{1}{N} \sum_{i=1}^{N_s} e_\delta\big(\bar{X}_{t_{n-1}}^{(i)}\big)\right] \delta_0(\mathrm{d}y) + \frac{1}{N} \sum_{i=1}^{N_s} g_\delta\big(y \,|\, \bar{X}_{t_{n-1}}^{(i)}\big)\, \mathrm{d}y \end{aligned}$$

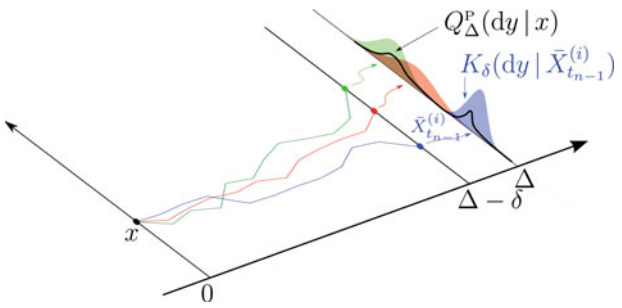

Figure 1. The Perdersen approximation Q^{P}_{Δ} of the kernel Q_{Δ} is obtained by sampling N independent trajectories $\bar{X}^{(i)}_{t_{0:n-1}}$ from (10), then the approximation is given by (14). See text for the precise treatment of extinct trajectories.

so that $Q^{\mathrm{P}}_{\Delta}(\mathrm{d}y \mid x)$ admits the following density with respect to the measure $m(\mathrm{d}y)$:

$$q^{\mathrm{P}}_{\Delta}(y \mid x) \stackrel{\mathrm{def}}{=} \begin{cases} \dfrac{N - N_s}{N} + \dfrac{1}{N} \displaystyle\sum_{i=1}^{N_s} e_\delta\big(\bar{X}^{(i)}_{t_{n-1}}\big), & \text{if } y = 0, \\ \dfrac{1}{N} \displaystyle\sum_{i=1}^{N_s} g_\delta\big(y \mid \bar{X}^{(i)}_{t_{n-1}}\big), & \text{otherwise.} \end{cases}$$

This approach is presented in Pedersen (1995) for diffusion on $\mathbb{R}^d$ having an absolutely continuous density. Even with our adaptation allowing for the extinction, it is very easy to implement and does not involve heavy computations. It suffers however from a well-known problem of Monte Carlo methods: if few trajectories terminate around the observation y at which the density is to be evaluated, then only a few terms will significantly contribute to the approximation of $p_{\Delta}(\mathrm{d}y \mid x)$. In this case, the approximation will be of poor quality. Beside, a number of trajectories will have been generated uselessly.

3.3. Importance sampling with a modified Brownian bridge

An improvement of the Pedersen methods was proposed by Durham and Gallant (2002) to circumvent the above-mentioned problem. The idea, natural in the importance sampling context, is to use a weighted sample of a suitably chosen law for which most trajectories will contribute to the estimation. Once again, we have to modify the method to account for the

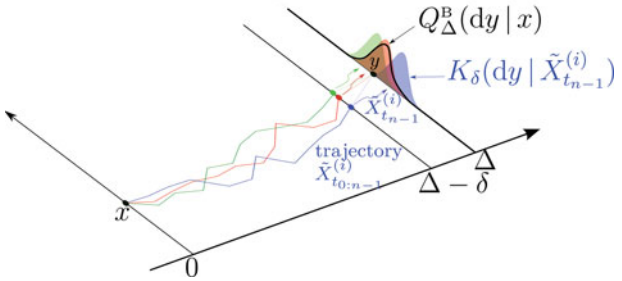

Figure 2. The approximation Q^{B}_{Δ} by importance sampling with a Brownian bridge of the kernel Q_{Δ} is obtained by sampling N independent trajectories $\tilde{X}^{(i)}_{t_{0:n-1}}$ of the approximation of the Brownian bridge (15); then the approximation is given by (17). See the text for the precise treatment of extinct trajectories.

possible extinction. Our choice is to generate the trajectories according to

$$\tilde{X}_{t_{k+1}} = \max\left(0, \tilde{X}_{t_k} + \delta\,\frac{y - \tilde{X}_{t_k}}{\Delta - t_k} + \sqrt{\delta}\,\sigma(\tilde{X}_{t_k})\,w_k\right), \quad k = 0, \ldots, n-1 \tag{15}$$

with $X_0 = x$ and $w_k \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ which is nothing but a (modified for extinction) Euler–Maruyama scheme for the SDE:

$$\mathrm{d}X_t = \frac{y - X_t}{\Delta - t}\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad 0 \le t < \Delta. \tag{16}$$

The drift term is designed so as to force the trajectories towards the given final value y, see Fig. 2. Note that solution of (16) would be a "true" Brownian bridge if σ were constant. The transition kernel associated with (15) depends on $t < \Delta$ and reads:

$$\tilde{K}_{t,\delta}(\mathrm{d}z \,|\, x) = \begin{cases} \tilde{e}_\delta(x)\,\delta_0(\mathrm{d}z) + \tilde{g}_{t,\delta}(z \,|\, x)\,\mathrm{d}z\,, & \text{if } x > 0, \\ \delta_0(\mathrm{d}z)\,, & \text{if } x = 0, \end{cases}$$

where

$$\tilde{e}_\delta(x) \overset{\text{def}}{=} \left(1 - \int_0^\infty \tilde{g}_{t,\delta}(z \,|\, x)\,\mathrm{d}z\right),$$

$$\tilde{g}_{t,\delta}(z \,|\, x) \overset{\text{def}}{=} \frac{1}{\sqrt{2\pi\,\delta\,\sigma(x)}}\,\exp\left\{-\frac{(z - x - \delta\,\frac{y-x}{\Delta-t})^2}{2\,\delta\,\sigma(x)}\right\}\mathbf{1}_{\mathbb{R}_+}(z).$$

This transition kernel is absolutely continuous with respect to $K_\delta(\mathrm{d}z \,|\, x)$:

$$K_\delta(\mathrm{d}z \,|\, x) = \psi_{t,\delta}(z \,|\, x)\,\tilde{K}_{t,\delta}(\mathrm{d}z \,|\, x)$$

with

$$\psi_{t,\delta}(z \,|\, x) = \begin{cases} \dfrac{1 - \int_0^\infty g_\delta(z \,|\, x)\,\mathrm{d}z}{1 - \int_0^\infty \tilde{g}_{t,\delta}(z \,|\, x)\,\mathrm{d}z}\,, & \text{if } z = 0, \\ \dfrac{\int_0^\infty g_\delta(z \,|\, x)\,\mathrm{d}z}{\int_0^\infty \tilde{g}_{t,\delta}(z \,|\, x)\,\mathrm{d}z}\,, & \text{otherwise,} \end{cases}$$

for $x > 0$ and $\psi_{t,\delta}(z \,|\, 0) = \mathbf{1}_{\mathbb{R}_+}(z)$. We now have another expression for $Q^{\mathrm{E}}_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x)$ as:

$$Q^{\mathrm{E}}_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x) = \int_0^\infty \cdots \int_0^\infty \Psi(y_1, \ldots, y_{n-1} \,|\, x)\,\tilde{K}_{0,\delta}(\mathrm{d}y_1 \,|\, x) \ldots \tilde{K}_{\Delta-2\,\delta,\delta}(\mathrm{d}y_{n-1} \,|\, y_{n-2})$$

where

$$\Psi(y_1, \ldots, y_{n-1} \,|\, x) = \psi_{\delta,\delta}(y_1 \,|\, x) \ldots \psi_{\Delta-2\,\delta,\delta}(y_{n-1} \,|\, y_{n-2}).$$

Hereafter, we will denote by

$$\tilde{X}_{t_{1:n-1}} = (\tilde{X}_{t_1}, \ldots, \tilde{X}_{t_{n-1}}) \in \mathbb{R}^n$$

a trajectory generated by (15) up to time $\Delta - \delta$, with initial value $\tilde{X}_0 = x$.

With this setting, and like in the previous section, in (13) first approximate $Q_{\Delta-\delta}$ by a weighted sample:

$$Q_{\Delta-\delta}(\mathrm{d}y_{n-1} \,|\, x) \simeq \frac{1}{N}\sum_{i=1}^N \Psi\left(\tilde{X}^{(i)}_{t_{1:n-1}} \,|\, x\right)\delta_{\tilde{X}^{(i)}_{t_{n-1}}}(\mathrm{d}y_{n-1})$$

then, as before, approximate $Q_\delta(\mathrm{d}y \,|\, y_{n-1})$ by $K_\delta(\mathrm{d}y \,|\, y_{n-1})$; finally, the kernel $Q_\Delta(\mathrm{d}y \,|\, x)$ defined by (13) is approximated by:

$$Q_\Delta^{\mathrm{B}}(\mathrm{d}y \,|\, x) \stackrel{\mathrm{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, K_\delta\big(\mathrm{d}y \,|\, \tilde{X}_{t_{n-1}}^{(i)}\big). \tag{17}$$

As before, let us re-number the sampled trajectories so that the surviving ones correspond to $i = 1, \ldots, N_s$:

$$\begin{aligned}
Q_\Delta^{\mathrm{B}}(\mathrm{d}y \,|\, x) &= \frac{1}{N} \sum_{i=N_s+1}^{N} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, K_\delta(\mathrm{d}y \,|\, 0) + \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, K_\delta\big(\mathrm{d}y \,|\, \tilde{X}_{t_{n-1}}^{(i)}\big) \\
&= \frac{1}{N} \sum_{i=N_s+1}^{N} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, \delta_0(\mathrm{d}y) \\
&\quad + \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big) \Big[e_\delta\big(\tilde{X}_{t_{n-1}}^{(i)}\big)\, \delta_0(\mathrm{d}y) + g_\delta\big(y \,|\, \tilde{X}_{t_{n-1}}^{(i)}\big)\, \mathrm{d}y\Big] \\
&= \left[\frac{1}{N} \sum_{i=N_s+1}^{N} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big) + \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, e_\delta\big(\tilde{X}_{t_{n-1}}^{(i)}\big)\right] \delta_0(\mathrm{d}y) \\
&\quad + \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, g_\delta\big(y \,|\, \tilde{X}_{t_{n-1}}^{(i)}\big)\, \mathrm{d}y
\end{aligned}$$

hence $Q_\Delta^{\mathrm{B}}(\mathrm{d}y \,|\, x)$ admits the following density with respect to $m(\mathrm{d}y)$:

$$q_\Delta^{\mathrm{B}}(y \,|\, x) \stackrel{\mathrm{def}}{=} \begin{cases} \frac{1}{N} \sum_{i=N_s+1}^{N} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big) + \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, e_\delta\big(\tilde{X}_{t_{n-1}}^{(i)}\big), & \text{if } y = 0, \\ \frac{1}{N} \sum_{i=1}^{N_s} \Psi\big(\tilde{X}_{t_{1:n-1}}^{(i)} \,|\, x\big)\, g_\delta\big(y \,|\, \tilde{X}_{t_{n-1}}^{(i)}\big), & \text{otherwise.} \end{cases}$$

Remark 3.3. It is possible to use other importance sampler than (15). Durham and Gallant (2002) used another modified Brownian bridge for which they claim that it reduces further the variance of the resulting approximation. The generating algorithm reads:

$$\tilde{X}_{t_{k+1}} = \tilde{X}_{t_k} + \delta\, \frac{y - \tilde{X}_{t_k}}{\Delta - t} + \sqrt{\delta}\, \sigma(\tilde{X}_{t_k}) \left(1 - \frac{\delta}{\Delta - t}\right) w_k.$$

The drift term is as in (15) but the variance is progressively damped up to final time $t = \Delta - 2h$ at which it equals $\frac{1}{2}\sigma(Y_t)$. For this sampler also an adaptation to our singular case is needed. We also mention the recent work by Bladt and Sørensen (2014).

4. Numerical experiments

We consider a scenario defined by the model parameter given in Table 1 already used in Campillo et al. (2016) for producing Figs. 4 and 7. For such values, the deterministic model (dashed line in Figs. 3 and 4) has a classical increasing "S" shape. For our stochastic model, we will distinguish three types of qualitatively different datasets as illustrated in Fig. 3:

- Type 1 is close to the deterministic case. We observe a transient phase of quick growth, after which the process enters an apparent stationary regime. The population persists at the final observation time. Recall that it will definitely become extinct in the future.

Table 1. Parameters

Model						
λ	μ	α	ρ	x_0	T	N
20	18	1	10^{-1}	0.25	10	200

Simulation		
h	δ	N
10^{-3}	10^{-3}	200

- Type 2 corresponds to early extinction, due to the low initial population. Datasets of this sort will induce a poor estimation of parameters.
- Type 3 shares some of the other attributes. The population collapses within the observation interval but completes the transient phase before.

A hint on the probability of occurrence of the three types is given in Fig. 4, which shows the finite difference approximation of the (defective) density. During the transient phase, both mean and variance of the density increases. During this phase, the high values of the density in the neighborhood of absorbing point 0 induce an important loss of mass. This corresponds to trajectories of Type 2. When reaching the (almost) stationary regime, the loss of mass through boundary 0 goes on imperceptibly (Type 2 trajectories). Type 1 trajectories are associated with the remaining probability. We clearly see that the deterministic trajectory does not follow the first moment of the density. These three types of trajectories correspond to the timescales of interest. Beside, if real data are available, only one trajectory, of some type, is given which is not sufficient to approximate the law of the extinction time.

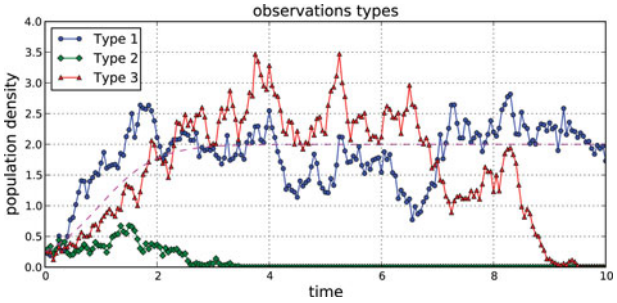

Figure 3. Three datasets qualitatively different. Population persists on the time horizon for Type 1. Extinction occurs early for Type 2 and later for Type 3. The dashed line is the classical deterministic solution.

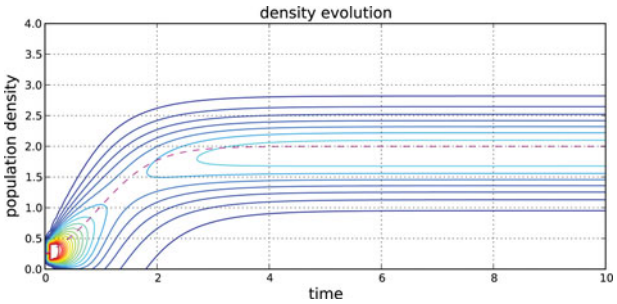

Figure 4. Contour plot of the density approximated by the finite difference scheme. The dashed line is the classical deterministic solution.

The simulation parameters of the approximation methods are also given in Table 1. In particular, the same time step δ is used for the Euler–Maruyama and for the finite difference scheme.

Growth rate estimation

First, note that there could not be any comparison with the standard algorithms since they produce probability densities with respect to the Lebesgue measure, which is not appropriate in our case.

The results presented hereafter have been obtained with a C++ code, computing the likelihood function by the deterministic FD (finite difference) methods and the two Monte Carlo methods (Pedersen and modified Brownian bridge). Optimization was performed with the help of the library `NLopt nonlinear-optimization package` of Steven G. Johnson, freely available at http://ab-initio.mit.edu/nlopt++ which provides the implementation of various algorithms. We have chosen to use a variant of the Nelder and Mead simplex algorithm described in Rowan (1990). Dependence on the initial condition was studied by varying the starting point of the optimization algorithm. The evaluation of the Monte Carlo methods uses 200 replicas.

For a given observed trajectory, the FD method always gives the same estimated density, whereas the randomness of the generated sample persists in the estimation for the two Monte Carlo methods. Even worse, we need two independent samples for estimation of the density at two different locations for the MBB, while a single sample is sufficient for the Pedersen method. To have a sensible comparison, we will first compare the Monte Carlo methods to the FD methods for a given observed trajectory. We will also use independent random sample for both Perdersen and MBB methods.

We first check that the estimation methods presented above can correctly handle datasets of all types, with or without extinction. Since we do not want this question to interfere with the identifiability of λ and μ separately, we will focus on the estimation of $\lambda - \mu$, with known values of α and κ. Figure 5 shows the results obtained with the three methods with the datasets of Fig. 3. Although we do not know if the maximum uniquely exists, the results for the FD

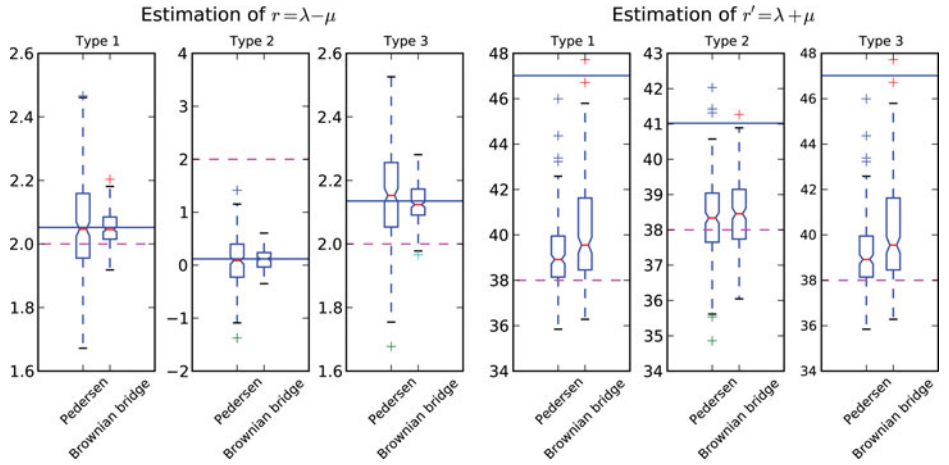

Figure 5. Comparison of the finite difference method and the two Monte Carlo methods, for parameters r (left) and r' (right). The horizontal dashed line is the true value. The solid line is the estimation given by the finite difference method.

method are stable, independently of the initial condition. As we can expect, datasets of Type 2 lead to a poor estimation of the growth rate $r = \lambda - \mu$. We also check that the quality of estimations with datasets of Type 1 and 3 do not differ significantly. For Type 3, all methods seem to exploit the information contained in the portion of trajectory preceding extinction. Beyond the quality of estimation, the Monte Carlo methods agree in mean with the deterministic FD method. We finally recover the well-known difference between Pedersen and modified Brownian bridge methods: the variance of the latter is much less, but with a computational cost much higher. Experiments with Type 1 trajectories sampled with $\Delta = 0.5$ have been carried out, without significantly affecting the picture.

Figure 5 also shows the results of the estimation of $r' = \lambda + \mu$ but this graphic is misleading for the two Monte Carlo methods. Indeed, monitoring the optimization algorithm with these methods reveals that the procedure does not stop because an optimum value has been found, but because there is no consistent progress from one step to another. The exit value of the optimization algorithm is then meaningless. This fact is a consequence of the poor identifiability of the model and is studied in the next paragraph.

Identifying λ and μ separately

We will note $\ell(\theta) = -\log(\mathcal{L}(\theta))$ the function to minimize with a fixed dataset of Type 1. Figure 6 shows the graph of the approximations of $\ell(\theta)$ for the three methods. We clearly distinguish two orthogonal directions of variations. Along the axis $\lambda + \mu = r'$ for a given r', there exists a unique minimum that is likely to be detected by all methods. The last plot shows that this is absolutely not the case for the axis $\lambda - \mu = r$. Indeed, the randomness of the Monte Carlo methods induces fluctuations greater than the order of magnitude of the variations of $\ell(\theta)$. For these methods, it is therefore predictable that a minimization algorithm will rapidly

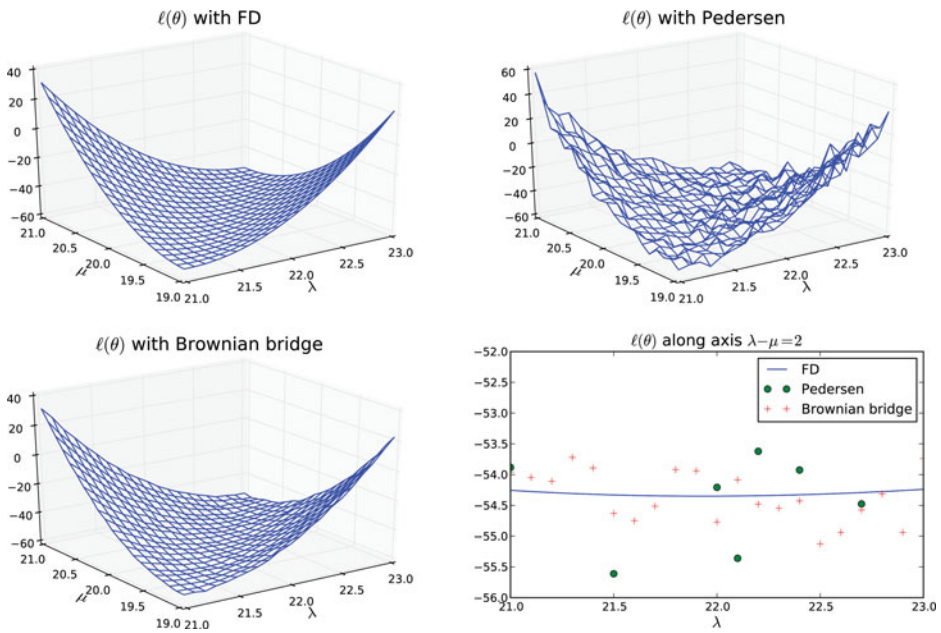

Figure 6. Approximations of $\ell(\theta)$ on a square grid. The characteristic shape is preserved by all methods. Last plot is a cross-section of 3D plots by the vertical plane $\lambda - \mu = 2$. Monte Carlo methods exhibit fluctuations greater than the variation of $\ell(\theta)$ itself along this axis.

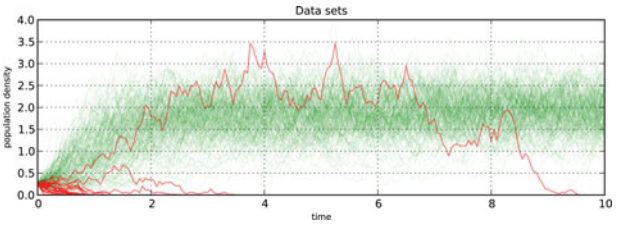

Figure 7. Datasets used for estimation by FD approximation. Extinction occurs for 20 datasets. For one of these (of Type 3), the transient phase is completed.

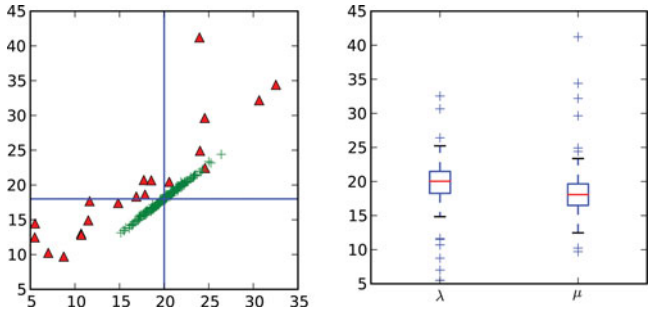

Figure 8. Empirical distribution of the MLE, using the finite difference approximation and parameters of Table 1. The right plot shows the marginal distribution. On the left plot, the red triangles correspond to trajectories becoming extinct before final time, see Fig. 7. Late extinction (the red triangle among green crosses) does not affect the estimation.

progress along the first axis, to be finally misled by the random fluctuations along the second one. As a result, we can rely only on the FD approximation since it is not subject to this shortcoming.

In Fig. 8, we plot the estimated values for the 200 datasets taken from Campillo et al. (2016) showed in Fig. 7. We again notice the asymmetry of the empirical distribution of the MLE. The variance is much smaller along the second axis, which means that information on $(\lambda - \mu)$ in the data is better understood than information on $(\lambda + \mu)$. The drift coefficient depends on $(\lambda - \mu)$, whereas the diffusion coefficient depends on $(\lambda + \mu)$. In our case, inferring on the diffusion appears to be the hardest task.

Finally, we observe the bad quality of the estimation when the dataset is of Type 2. Indeed, even if we have taken good care of the extinction, the information given by the data is not sufficient to allow a useful inference on the parameter. On the other hand, the estimation is not noticeably different for Type 1 and Type 3 datasets.

Concluding remarks

In most practical situations, the observed trajectory does not reach extinction. Indeed, if $\lambda > \mu$, ρ is small and X_0 is sufficiently large, the mass absorbed at 0 is negligible, so that the transition kernel is essentially $p^{\theta}_{\Delta}(y \mid x)\,\mathrm{d}y$. It can then be argued that the interest in the extinction question is irrelevant. Indeed, since the probability flux through the boundary expressed in (9b) is almost 0, one can impose a zero-flux boundary condition to the classical FPE. Likewise, the classical Pedersen and MBB methods would not differ significantly from the variants proposed above. Nonetheless, it makes sense to take care of the extinction probability $E_{\Delta}(x)$ since the two parts are strongly related. Any maximization algorithm will evaluate this

likelihood for many values of the parameter. For some of these values, neglecting the extinction could lead to an abnormally high value of the likelihood and thus mislay the maximization algorithm. Coupling extinction and non-extinction in the CFPE results in a more robust and accurate estimation procedure. This approach is not specific to the particular application under consideration.

A natural extension would be to consider noisy measurements, thus making the state process $\{X_t\}_{0\le t\le T}$ a *hidden diffusion*. Despite its obvious practical relevance, the literature on this problem is far from being abundant, even for the regular case. We mention Gloter and Jacod (2001) or Favetto (2014), where minimum contrast estimators are constructed. These constructions should again be adapted to face extinction.

Acknowledgments

This work was partially supported by the Laboratory of Excellence (Labex) NUMEV (Digital and Hardware Solutions, Modelling for the Environment and Life Sciences) coordinated by University of Montpellier, France.

Appendix

Finite difference approximation

The numerical scheme presented hereafter was introduced and discussed in the previously cited work Campillo et al. (2016). Its design comes straight from a probabilistic viewpoint, since the solution looked for is a probability density. We just describe it briefly, focusing on the key features and referring to Campillo et al. (2016) for details.

The space is first discretized as a regular grid:

$$x_\ell = \ell\, h\,, \quad \ell = 0, \dots, L$$

for h and L given, with an additional *cemetery point* Υ at location 0, see Fig. A.1. Indeed, notice that the boundary point 0 has a twofold status as seen from (9): it is the only point appearing in the continuous part $p_t(y \mid x)$ and in the discrete part E_t. We distinguish these status by using the node x_0 of the grid to compute the approximation of the continuous component $p_t(0 \mid x)$, and the cemetery point Υ to carry the approximation of the extinction probability E_t. It should be noted that the continuous component p_t can be approximated independently from E_t, whereas the latter depends directly on p_t. Following the probabilistic approach presented in Kushner and Dupuis (2001), we replace the operator $\mathcal{A}$ by a discrete approximation $A = (A_{\ell,k})_{\ell,\, k\in\{\Upsilon,0,\dots,L\}}$, in such a way that A is the generator of a pure jump Markov process on the grid. Denote by $P_t(\ell)$ the probability that this process occupies site ℓ at time t,

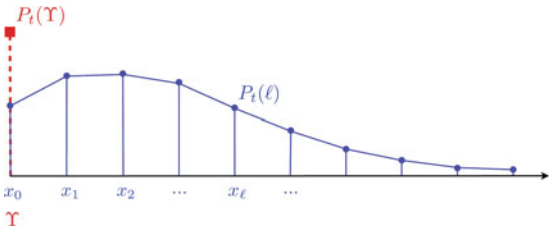

Figure A.1. Discretization of the state space as a regular finite grid. Value $y = 0$ is either the node x_0 at which the value of p_t is evaluated and the cemetery point Υ.

we will consider the approximations (see Fig. A.1):

$$P_t(\ell) \simeq h\, p_t(x_\ell \,|\, x), \qquad \text{for } 0 < \ell < L$$
$$P_t(0) \simeq \frac{h}{2}\, p_t(0 \,|\, x), \quad P_t(L) \simeq \frac{h}{2}\, p_t(x_L \,|\, x), \quad P_t(\Upsilon) \simeq E_t(x).$$

The generator A obtained with an upwind scheme is a tridiagonal matrix, whose line ℓ is given for any interior point $\ell = 1, \ldots, L-1$ by:

$$A_{\ell,\ell-1} = \frac{b^-(x_\ell)}{h} + \frac{a(x_\ell)}{2\,h^2}, \quad A_{\ell,\ell} = -\frac{|b(x_\ell)|}{h} - \frac{a(x_\ell)}{h^2}, \quad A_{\ell,\ell+1} = \frac{b^+(x_\ell)}{h} + \frac{a(x_\ell)}{2\,h^2}.$$

The boundary condition at node 0 gives the coefficients of line $\ell = 0$:

$$A_{0,\Upsilon} = -A_{0,0} \quad A_{0,0} = -|b'(0)| - \frac{1}{h}\, a'(0)$$

as an approximation of Eq. (9b). All other coefficients are set to 0. Finally, an artificial reflecting condition at node x_L completes the matrix:

$$A_{L,L-1} = \frac{|b(x_L)|}{h} + \frac{a(x_L)}{h^2}, \qquad A_{L,L} = -\frac{|b(x_L)|}{h} - \frac{a(x_L)}{h^2}.$$

As a remarkable feature of this scheme, we notice Υ is an absorbing point, but not $x_0 = 0$. Indeed, the process can jump to x_0 and stay there for a while before jumping to cemetery point Υ. The occupation probability of site x_0 yields an approximation of $p_t(0|x)$, while the occupation probability of site Υ is an approximation of extinction probability E_t.

Time discretization. We introduce the notation $\mathbf{P}_t = (P_t(\ell))_{\ell=\Upsilon,0,\ldots,L}$ for the law of the jump process at time t starting from x. This probability distribution solves the Kolmogorov forward equation for jump processes that reads:

$$\dot{\mathbf{P}}_t = A^* \mathbf{P}_t. \tag{18}$$

Equation (18) is discretized in time using the Euler implicit scheme:

$$[I - \delta\, A]^*\, \tilde{\mathbf{P}}_{t_{k+1}} = \tilde{\mathbf{P}}_{t_k}, \quad k = 0, \ldots, n-1$$

where $t_k \overset{\text{def}}{=} k\,\delta$ with $\delta = \Delta/n$, n given; see Fig. A.2. The initial condition is approximated by

$$\tilde{P}_{t_0}(l) = \begin{cases} 1, & \text{if } \ell = \ell_0, \\ 0, & \text{otherwise,} \end{cases}$$

where x_{ℓ_0} is the nearest neighbor in the grid of the initial condition x. The numerical solution $\tilde{\mathbf{P}}_\Delta$ yields a numerical approximation $\tilde{p}_\Delta(x_\ell \,|\, x)$ for the density at a grid point, that can be linearly interpolated to obtain an approximation $\tilde{p}_\Delta(y \,|\, x)$ for $0 \le y \le x_L$. The density q_Δ is

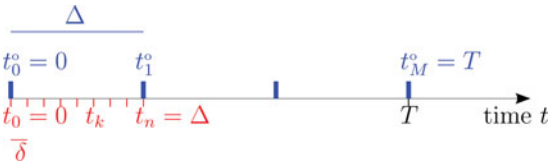

Figure A.2. The observation instants are $t_k^o = k\,\Delta$ with $\Delta = \frac{T}{M}$; the time-discretization instants are $t_k = k\,\delta$ with $\delta = \frac{\Delta}{n}$ for n given.

finally approximated by

$$q_\Delta(x,y) \simeq \begin{cases} P_\Delta(\Upsilon)\,, & \text{if } y = 0\,, \\ \tilde{p}_\Delta(y\,|\,x)\,, & \text{if } y \in]0, x_L]\,. \end{cases}$$

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* 70(1):223–262.

Beskos, A., Roberts, G. O. (2005). Exact simulation of diffusions. *The Annals of Applied Probability* 15(4):2422–2444.

Bladt, M., Sørensen, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli* 20(2):645–675.

Campillo, F., Joannides, M., Larramendy-Valverde, I. (2016). Analysis and approximation of a stochastic growth model with extinction. *Methodology and Computing in Applied Probability* 18(2):499–515.

Casella, D., Roberts, G. (2008). Exact Monte Carlo simulation of killed diffusions. *Advances in Applied Probability* 40(1):273–291.

Durham, G. B., Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics* 20(3): 297–338.

Favetto, B. (2014). Parameter estimation by contrast minimization for noisy observations of a diffusion process. *Statistics* 48(6):1344–1370.

Fearnhead, P. (2008). Computational methods for complex stochastic systems: A review of some alternatives to MCMC. *Statistics and Computing* 18(2):151–171.

Feller, W. (1952). The parabolic differential equations and the associated semi-groups of transformations. *The Annals of Mathematics* 55(3):468–519.

Gloter, A., Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimator. *ESAIM PS* 5:243–260.

Gobet, E. (2001). Euler schemes and half-space approximations for the simulation of diffusion in a domain. *ESAIM PS* 5:261–293.

Golightly, A., Wilkinson, D. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis* 52(3): 1674–1693.

Grasman, J., van Herwaarden, O. A. (1999). *Asymptotic Methods for the Fokker-Planck Equation and the Exit Problem in Applications*. New York: Springer-Verlag.

Heydari, J., Lawless, C., Lydall, D. A., Wilkinson, D. J. (2014). Fast Bayesian parameter estimation for stochastic logistic growth models. *Biosystems* 122(0):55–72.

Hurn, A. S., Jeisman, J., Lindsay, K. A. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics* 5(3):390–455.

Hurn, A. S., Lindsay, K. A., Martin, V. L. (2003). On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations. *Journal of Time Series Analysis* 24(1):45–63.

Jensen, B. and Poulsen, R. (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *The Journal of Derivatives* 9(4):18–32.

Kloeden, P. E., Platen, E., Schurz, H. (2003). *Numerical Solution of SDE Through Computer Experiment*. New York: Springer–Verlag.

Kushner, H. J., Dupuis, P. (2001). *Numerical methods for stochastic control problems in continuous time, Vol. 24 of Applications of Mathematics*. 2nd ed. New York: Springer–Verlag.

Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* 4(2):231–247.

Pastorello, S., Rossi, E. (2010). Efficient importance sampling maximum likelihood estimation of stochastic differential equations. *Computational Statistics & Data Analysis* 54(11):2753–2762. The Fifth Special Issue on Computational Econometrics.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22(1):55–71.
Risken, H. (1996). *The Fokker–Planck Equation: Methods of Solutions and Applications*. 2nd ed. Springer Series in Synergetics. New York: Springer.
Rowan, T. H. (1990). *Functional Stability Analysis of Numerical Algorithms*. PhD thesis, Department of Computer Sciences, University of Texas at Austin.
Schurz, H. (2007). Modeling, analysis and discretization of stochastic logistic equations. *International Journal of Numerical Analysis and Modeling* 4(2):178–197.