Population explosion suppressed by noise: stationary distributions and how to simulate them

P F Góra

Marian Smoluchowski Institute of Physics and Mark Kac Complex Systems Research Centre, Jagellonian University, Reymonta 4, 30–059 Kraków, Poland
E-mail: gora@if.uj.edu.pl

Abstract. We show that two dynamical systems exhibiting very different deterministic behaviours possess very similar stationary distributions when stabilized by a multiplicative Gaussian white noise. We also discuss practical aspects of numerically simulating these systems. We show that there exists a noise level that is optimal in the sense that the interval during which discrete-time versions of the systems remain physical is maximized. Analytical results are illustrated by numerical examples.

Contents

1. Introduction

Population dynamics is one of the natural areas of application of the theory of stochastic processes: parameters of deterministic models are perturbed by random fluctuations in order to model the influence of an ever-changing environment consisting of a multitude of individual

'agents'. Analytical solutions are known for only a handful of such systems. A majority of physically important systems is accessible only via numerical simulations. It is, therefore, particularly important to thoroughly understand those few whose analytical properties are known and test how well the existing numerical methods can approximate them. When a rigorous mathematical statement can be made about a system described by a stochastic differential equation (SDE), one expects that it will be corroborated by numerical simulations or by an actual physical experiment, if the latter is possible: after all, an SDE is a mathematical idealization of a physical reality that is discrete by its very nature. If the results of the two approaches, continuous time SDE versus discrete time numerical simulations, do not agree, there are causes for serious concern.

One of the most surprising results, reported recently by Mao *et al* in [1], is the fact that a multiplicative noise can stabilize a system whose deterministic counterpart is divergent. The class of systems discussed in that paper may comprise several interacting species but its simplest form, to which we will restrict ourselves within this paper, is described by a scalar Langevin equation

$$\dot{x} = rx(1 \pm x) + \sigma x^2 \xi(t), \tag{1}$$

where $r > 0$ and $\xi(t)$ is a Gaussian white noise (GWN), interpreted in the sense of Ito, with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t_1)\xi(t_2) \rangle = \delta(t_1 - t_2)$. If there is no noise, $\sigma = 0$, the '$-$' sign corresponds to the well-known logistic equation, while the '$+$' sign describes a 'population' that diverges in a finite time. This latter case defies one of the ecological laws that a population should be self-limited [2], although equation (1) perhaps can be used to model explosive phenomena observed in some chemical, nuclear or stellar systems where a divergence is interpreted as a destruction of the original system. However, our primary purpose of discussing the '+' sign is to present the surprising effects that noise can have on a deterministically divergent system.

It is rigorously shown in [1] that for any non-zero noise, $\sigma \neq 0$, if the system (1) starts from a positive initial value, it will for all times remain positive and bounded almost surely, regardless of the choice of the sign inside the brackets. This means that even a tiny addition of the noise can stabilize the system. We note that a proof of this statement runs in two stages: first it is proven that the system remains positive, and later this fact is used to prove that the system remains bounded.

Apart from the formal proof, [1] provides several numerical examples. These examples, however, are much less convincing than the fine mathematical points that they are supposed to illustrate. Firstly, it is not clear whether the numerical trajectories presented are typical or handpicked 'best' ones. Secondly, the noise has been approximated by an uncorrelated Bernoulli noise, while it is believed that one should either use an exponentially correlated dichotomic noise and then take a certain double limit to simulate a GWN [3], or directly use good generators of a GWN. We will see that the problem of simulating the system described by equation (1) is unexpectedly tricky: even though this system almost surely never diverges, we will show that its discretizations realized by several popular and well-established methods do diverge.

This paper is organized as follows: in section 2 we construct stationary probability distributions corresponding to equation (1). Several interacting species may not converge to a stable fixed point even if their trajectories are bounded. They may display oscillations instead and may not have a stationary distribution. However, if there is only one species, finding its stationary distribution is a straightforward task. We will show that systems whose deterministic dynamics are completely different may possess similar stationary distribution when perturbed by

the noise. Then in sections 3 and 4 we will discuss how to simulate the system (1). In particular, we will show that there is an 'optimal' noise level that maximizes the average lifespan of a discretized system. Conclusions are given in section 5.

2. The stationary distribution

Equation (1) leads to the following Fokker–Planck equation:

$$\frac{\partial P(x,t)}{\partial t} = -r\frac{\partial}{\partial x}[x(1 \pm x)P(x,t)] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}[x^4 P(x,t)].\tag{2}$$

A stationary solution to this equation can be easily found by standard methods [5, 6]. It is, however, instructive to take a different approach. We make a substitution

$$y = \frac{1}{x}.\tag{3}$$

Note that because x is non-negative almost surely, the substitution (3) is always valid, i.e. we do not risk a division by zero almost surely. As a result of this substitution, the Fokker–Planck equation (2) in the Ito interpretation is transformed into [4]

$$\frac{\partial P(y,t)}{\partial t} = \frac{\partial}{\partial y}\left[\left(ry \pm r - \frac{\sigma^2}{y}\right)P(y,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 P(y,t)}{\partial y^2},\tag{4}$$

which, in turn, corresponds to the following Langevin equation:

$$\dot{y} = -\left(ry \pm r - \frac{\sigma^2}{y}\right) + \sigma\xi(t).\tag{5}$$

The deterministic part of this equation describes an overdamped motion in a 'potential'

$$U(y) = \tfrac{1}{2}ry^2 \pm ry - \sigma^2 \ln y.\tag{6}$$

We can see that the noise interpreted in the sense of Ito introduces a barrier preventing y from crossing zero, or preventing the population x from diverging, as predicted in [1]. Note that this barrier is absent if equation (1) is interpreted in the sense of Stratonovich, and indeed there is no proof that the population x does not explode in this case. Finding the stationary solution to equation (4) is now a matter of a simple calculation:

$$P_{st}(y) = \mathcal{N}y^2 \exp\left[-\frac{r}{\sigma^2}y(y \pm 2)\right],\tag{7}$$

or after transforming back to the original variable,

$$P_{st}(x) = \frac{\mathcal{N}}{x^4}\exp\left[-\tilde{r}\frac{1 \pm 2x}{x^2}\right],\tag{8}$$

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

where $\tilde{r} = r/\sigma^2$ and the normalization constant equals

$$\mathcal{N} = \frac{4\tilde{r}^{3/2}}{\sqrt{\pi}\exp(\tilde{r})(1 + 2\tilde{r})(1 \mp \text{Erf}(\sqrt{\tilde{r}})) \mp 2\sqrt{\tilde{r}}} \tag{9}$$

and $\text{Erf}(\cdot)$ is the error function. It is easy to verify that for any positive value of $\tilde{r}$, and for both choices of the sign, the normalization constant $\mathcal{N}$ is positive and finite. The distribution (8) is the stationary probability distribution corresponding to equation (2) with the constraint $x > 0$. It reaches a maximum at $x = (\sqrt{\tilde{r}(\tilde{r} + 8)} \pm \tilde{r})/4$.

Observe that the '−' sign in (1) and the subsequent equations corresponds to a noisy logistic equation; the corresponding deterministic system is certainly stable and never diverges. The sign '+' corresponds to a system whose deterministic counterpart diverges very rapidly. However, the choice of this sign has very little effect on the shape and properties of the distribution (8). It is indeed very surprising that two systems whose deterministic behaviours differ so dramatically have very similar stationary distributions when driven by a GWN.

3. Numerical simulations

When one deals with a Langevin equation of the form

$$\dot{x} = f(x) + g(x)\xi(t) \tag{10}$$

and does not know a stationary distribution corresponding to it, the usual way to proceed is to use an appropriate discretization in time, solve such a system numerically, let it 'equilibrate' for a certain time and try to reconstruct the (unknown) stationary distribution from the numerical trajectories. If $\xi(t)$ is a GWN, the well-known Euler–Maruyama scheme is most frequently used for this purpose. In this scheme, the numerical trajectory is generated by

$$x_{n+1} = x_n + hf(x_n) + \sqrt{h}g(x_n)\eta_n, \tag{11}$$

where h is the time step and η_n is a discrete GWN: $\langle\eta_n\rangle = 0$, $\langle\eta_n\eta_m\rangle = \delta_{nm}$. The Euler–Maruyama scheme is consistent with equation (10) in the Ito interpretation in the limit $h \to 0^+$. This scheme is numerically very cheap but its strong order of convergence is rather small, $\gamma = 1/2$ [7].

Simulating the noisy logistic equation, corresponding to a '−' in equation (1), does not pose any particular problem. We will, therefore, focus on simulating the system whose deterministic counterpart diverges. If we apply the Euler–Maruyama scheme to equation (1), we can see that the population is prevented from exploding by sudden jumping down. The amplitude of these jumps is scaled by the square of the current value of the population: the higher the system climbs, the lower it can drop. Large jumps up are equally possible, but a long series of jumps up is unlikely. If x_n is large, even a small (and positive) value of η_n introduces a significant increase in the population size. In the next step, however, the probabilities of jumping up and down are equal and even smaller to the absolute value but negative $\eta_{n+1}(\eta_{n+1} < 0, |\eta_{n+1}| < \eta_n)$ can undo the cumulative effect of several jumps up as it is scaled by a much larger factor. We know that the ideal, continuous time system neither diverges, nor becomes negative, but the discretization scheme (11) applied to the equation (1) may not obey these principles. If the current magnitude of the population becomes too large or negative, the model (1) clearly loses any physical meaning.

We say that the numerical trajectory becomes unphysical either if $x_{n+1} < 0$ or if $x_{n+1} > X \gg 0$ and stop the simulation when either of these situations occurs. Observe that the sequence $\{x_n\}$ does not converge to zero without becoming negative: the probability

$$\text{Prob}(0 < x_{n+1} < x_n | x_n > 0) = \text{Prob}(\eta_n < -(\sqrt{h}r/\sigma)(1 + 1/x_n)), \quad (12)$$

becomes negligible for $0 < x_n \ll 1$. On the other hand, the probability that x_{n+1} becomes negative equals

$$\text{Prob}(x_{n+1} < 0 | x_n > 0) = \text{Prob}\left(\eta_n < -\frac{1 + hr + hrx_n}{\sqrt{h}\sigma x_n}\right)$$
$$= \frac{1}{2}\left[1 - \text{Erf}\left(\frac{1 + hr + hrx_n}{\sqrt{h}\sigma x_n}\right)\right]. \quad (13)$$

This probability goes to zero in the limit $h \to 0^+$, which is nothing more than a consistency check of the Euler–Maruyama scheme, but is non-zero for any $h > 0$ and any $x_n > 0$. This means that after a certain, perhaps very large but finite, number of iterations, the Euler–Maruyama scheme will produce a negative, i.e. unphysical, value. Thus, the discretized system has a finite lifespan, i.e. the time after which the system, when started from a positive initial value, either becomes negative or exceedingly large.

Observe that if $0 < x_n \ll 1$, probability (13) is very small and increases with an increasing value of x_n. If $x_n \gg 1$, probability (13) reaches a constant value, $\text{Prob}(x_{n+1} < 0 | x_n > 0) \simeq (1 - \text{Erf}(\sqrt{h}\,r/\sigma))/2$, which may be quite large; for example, for $h = 2^{-16}$ and $r/\sigma = 1$, $\text{Prob}(x_{n+1} < 0 | x_n \gg 1) \simeq 0.498$. We can see that it is easier for a discretized version of system (1) to cross zero if the current value of the process, or the current population, is large than when it is small. This conclusion is well confirmed by numerical simulations.

Events in which x_n becomes larger than the threshold value $X \gg 0$ are very rare. If x_n becomes large, it is more likely for it to decrease in a sudden jump and start its way up again, or even become negative when the simulation stops than gradually climb up towards the threshold.

If the noise intensity, σ, is very small, the 'stopping effect' of the noise is limited: the population can build up beyond the threshold or, more likely, to such a value where the probability of the next iterate becoming negative is significant despite the small value of σ. On the other hand, in the limit $\sigma \to \infty$, the probability (13) goes to 1/2 and the system becomes unphysical very rapidly. Therefore, we expect the lifespan of a system with a very small or very large noise intensity to be short. There should be an 'optimal' range of the intensity of the noise, in which the noise effectively stops the system from climbing too high and at the same time the probability (13) has a modest value. Lifespans of the system with such a noise intensity should be larger than for either small or large noise intensities.

We have performed numerical simulations to verify this. We have generated trajectories of equation (1) according to Euler–Maruyama scheme with a time step of $h = 2^{-16} \simeq 1.5 \times 10^{-5}$. The noise has been generated by Marsaglia algorithm [8] with Mersenne Twister as the underlying uniform generator [9]. The simulations were stopped when x_{n+1} became negative or larger than $X = 10^{20}$; the latter events were very rare. For each value of σ, the lifespans were averaged over 1024 realizations of the process. Results are presented in figure 1. As we can see, the average lifespan increases rapidly as σ increases, reaches a plateau and then gradually decreases. These results show clearly that there is a range of 'optimal' noise intensities that maximize the lifespan
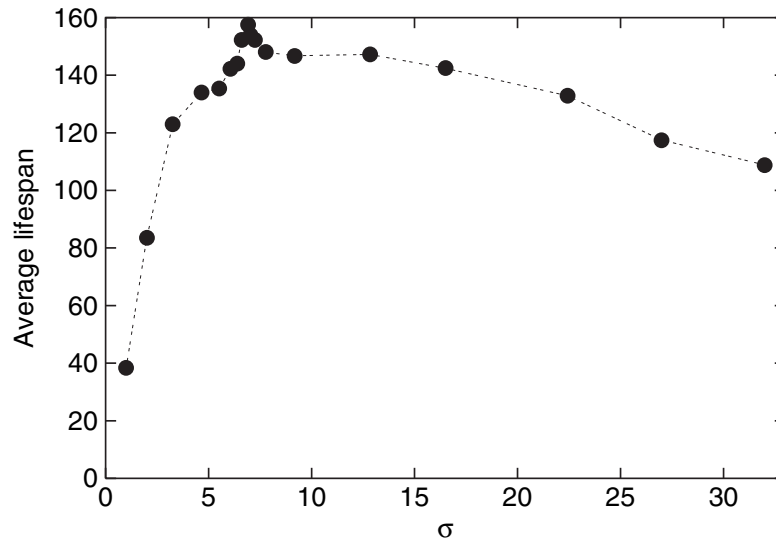
Figure 1. Average lifespan of system (1) discretized by the Euler–Maruyama scheme (11) as a function of the noise amplitude σ. The lifespan and σ are dimensionless. The line connecting the points is meant as a guide for the eye only.

of the discretized process. A precise location of the maximum is not particularly important as it may depend on realizations of the noise.

The average lifespan displays a strong dependence on the time step used to perform the simulations: the average lifespans increase (decrease) with a decreasing (increasing) value of the time step, h, and this increase (decrease) can be quite significant. However, with very short time steps, the time needed to perform any realistic simulations gets prohibitively large.

Data gathered during the simulations were also used to draw 'experimental' histograms of the distribution (8). Specifically, from all realizations with $h = 2^{-16}$ and lifespans equal to or greater than 64, values of $x(t)$ for $32 \leqslant t \leqslant 64$ were collected, binned in intervals of the size 2^{-10}, averaged over the respective number of realizations and compared with histograms obtained from the theoretical distribution. Selected results are presented in figure 2. The agreement between theory and numerical experiment is very good for $1/2 \lesssim \sigma \lesssim 1024$, meaning that the Euler–Maruyama scheme reproduces properties of the equation (1) well. Results outside this interval are worse because few sufficiently long trajectories were available and the statistics was poor; for large σ one should also use a smaller bin size.

4. Higher-order schemes

Sometimes there is a need to use a higher-order discretization scheme. Several such methods are known in the literature and in this section we will discuss the performance of two of them. One is the Heun method [10] which many people, including the present author, have applied on several occasions with a great success. However, when this method is used to discretize the equation (1), it leads to negative values of the population very rapidly. The reason for this apparent failure is the fact that the Heun method is consistent with the equation (10) in the Stratonovich, not Ito, interpretation.
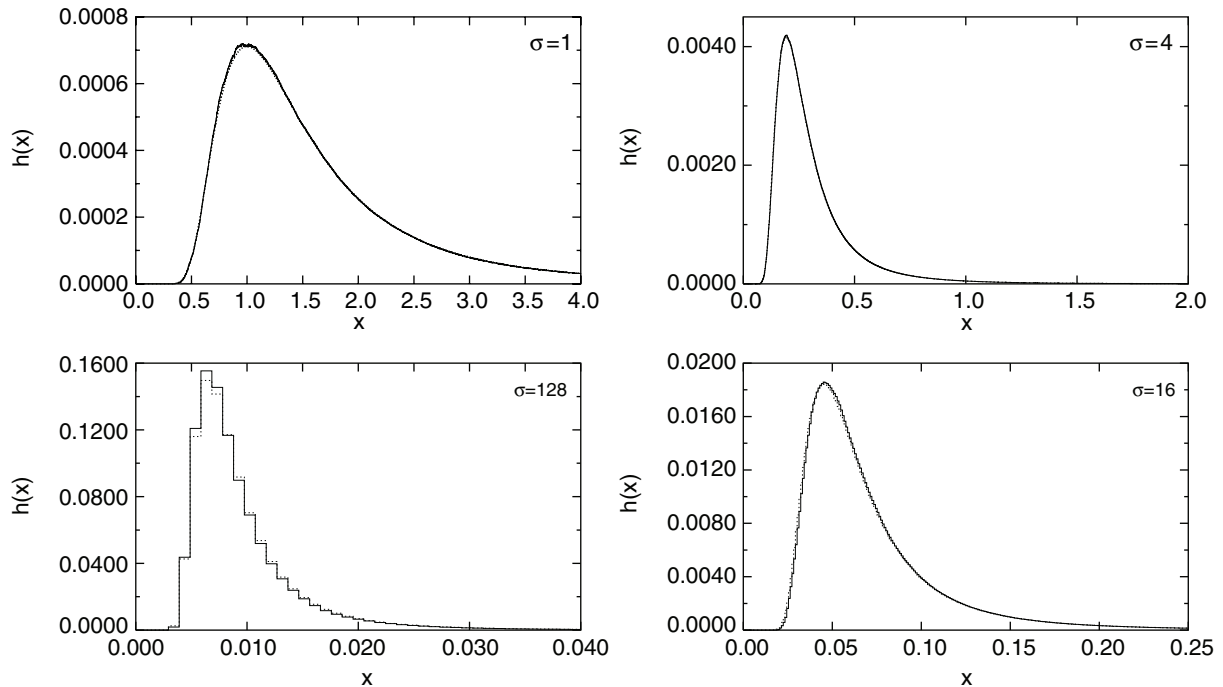
Figure 2. Numerical (——) and theoretical (- - - -) histograms of distribution (8) corresponding to the deterministically divergent case (the ' + ' sign). Clockwise from top left $\sigma = 1, 2, 16$ and 128, respectively. The agreement between theoretical and numerical histograms is very good.

The other method, perhaps less popular among physicists, is the Milstein scheme [11]:

$$x_{n+1} = x_n + hf(x_n) + \sqrt{h}g(x_n)\eta_n + \frac{h}{2}g(x_n)g'(x_n)\left(\eta_n^2 - 1\right), \tag{14}$$

where h and η_n are as in equation (11) and $g'(x_n) = (\mathrm{d}g/\mathrm{d}x)|_{x=x_n}$.

The Milstein scheme differs from Euler–Maruyama only in the last term and this term vanishes if the noise is purely additive ($g = $ const). Its computational burden depends on how difficult it is to evaluate the derivative. In case of a simple polynomial, the overall cost does not exceed twice that for the Euler–Maruyama method.

When this method is applied to equation (1), we find for the probability of the next iterate becoming negative

$$\mathrm{Prob}(x_{n+1} < 0 | x_n > 0) = \frac{1}{2}\left[\mathrm{Erf}\left(\frac{\sqrt{\Delta} - 1}{2\sqrt{h}\sigma x_n}\right) + \mathrm{Erf}\left(\frac{\sqrt{\Delta} + 1}{2\sqrt{h}\sigma x_n}\right)\right], \tag{15a}$$

where

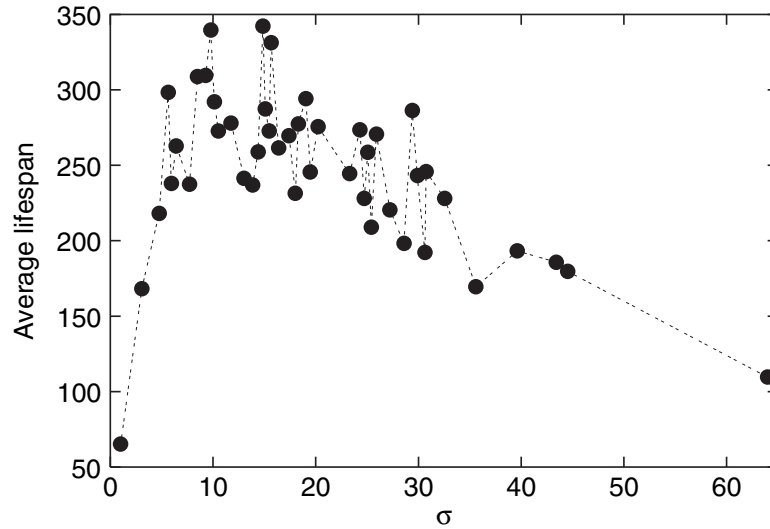$$\Delta = 4h\left(\sigma^2 x_n^2 - rx_n - x_n\right) \tag{15b}$$

Figure 3. Same as figure 1 but for the Milstein scheme (14). The lifespan and σ are in the same units as in figure 1.

provided that $\Delta > 0$; otherwise x_{n+1} cannot become negative. This last condition means that if

$$x_n \leqslant x_{\min} = \frac{r\sqrt{h} + \sqrt{3\sigma^2 + hr(r + 4\sigma^2)}}{2\sqrt{h}\sigma^2}, \tag{16}$$

the next iterate must be positive. For a small time step or for a small noise intensity, $x_{\min}$ can be large. This is good news as this prevents the discretized system from becoming unphysical. On the other hand, in the limit of a very strong noise, $\text{Prob}(x_{n+1}<0 \,|\, x_n>0) \rightarrow \text{Erf}(1) \simeq 0.843$, which is much larger than the corresponding value for the Euler–Maruyama scheme. The same happens if x_n becomes large, which is likely if the noise intensity is small. There is a trade-off between preventing the system from becoming unphysical for small values of the population and a rapid divergence once the populations builds up sufficiently large. As values of x_n are moderate for most of the time, we expect that the Milstein scheme would lead to larger lifespans than Euler–Maruyama, but the qualitative picture is much the same as in the latter: the lifespans for both small and large noise intensities are small and there is a range of noise intensities that maximize the average lifespan. These predictions have been confirmed numerically. Simulations were run under the same conditions as in case of the Euler–Maruyama scheme. Results are presented in figure 3. Indeed, the average lifespans first increase, then reach a range of elevated values and eventually decrease. Note that the maximal lifespans are more than twice as large as those for the Euler–Maruyama scheme. The structure seen in figure 3 (there are apparently two maxima visible) does not necessarily correspond to any realistic properties of the iteration (14) applied to equation (1). Rather than that, this structure may reflect a very pronounced stochastic variability of the system: with the Milstein scheme, an average lifespan of the order of 2.5×10^2 corresponds to a pool of runs with lifespans ranging from very short ($\sim10^{-3}$) to very large ($>10^4$) values.

5. Conclusions

It is well known that a discretization can dramatically change properties of a dynamical system. The deterministic logistic equation is one classic example: its continuous version is perfectly regular but after the discretization, the logistic equation may display dynamical chaos. Here we have a system—incidentally, closely related to the logistic equation—which, when modelled via the continuous time SDE, is mathematically proven to remain positive and bounded almost surely, but after a discretization it has a finite lifespan after which it becomes negative or diverges. We have argued that, for a given time step, there exists a noise level that is 'optimal' in that it maximizes the average lifespan of the discretized system. Our numerical results strongly support this argument.

We stress that the stabilization effect is observed only in the Ito interpretation. It is usually assumed that it is the physical interpretation of the stochastic force that should decide which interpretation of the noise, Ito or Stratonovich, should be used [4]. One may argue, however, that a robust noise-induced stabilization should be oblivious to the noise interpretation. In this respect, [1] and the present paper provide an example of a system in which the consequences of the two approaches differ.

In this paper we are lucky to numerically simulate a system whose analytical properties are fully known. In practice, it is systems whose analytical properties are not known that are examined by numerical simulations. If the simulations produce divergent results, one usually concludes that the lifespan of the system under consideration is finite. We have shown that such divergences can be merely an artefact of a discretization of an SDE.

Strangely, the above argument can also be reversed: a continuous time SDE is a mathematical idealization of a system that is discrete by its very nature. Firstly, populations consist of discrete individuals. A fractional x can be interpreted only as a fraction of a certain reference population. With $x \ll 1$, this interpretation breaks and so may break a model based on a continuous SDE. Secondly, it is assumed that the stochastic force acts continually. This idealization is fully justified when the characteristic timescale of the elementary random events is many orders of magnitude smaller than the characteristic timescale of a macroscopic process, which is the case for example in the classical diffusion. However, when an SDE is used to describe a biological or ecological process, the separation of timescales is less perfect. For instance, in a biological process the characteristic time of 'random' changes in environmental conditions can be of the order of seconds, and the characteristic time of the macroscopic process, like changes in a population, can be of the order of days. In such a case the characteristic timescales are separated by only five orders of magnitude, much as in the numerical examples reported above. As a result, a process that is deterministically divergent may remain so even after a stochastic perturbation is applied, even though solutions to a corresponding continuous time SDE remain bounded almost surely.

References

[1] Mao X, Marion G and Renshaw E 2002 *Stochastic Process. Appl.* **97** 95
[2] Turchin P 2001 *Oikos* **94** 17
[3] Van Den Broeck C 1982 *J. Stat. Phys.* **31** 467
[4] Van Kampen N G 1987 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland) ch VIII
[5] Risken H 1984 *The Fokker–Planck Equation* (Berlin: Springer)
[6] Gardiner C W 1993 *Handbook of Stochastic Methods* (Berlin: Springer)

Institute *of* **Physics** ◆ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

[7] Kloeden P E and Platen E 1992 *Numerical Solution of Stochastic Differential Equations* (Berlin: Springer)
   Kloeden P E 2002 *Milan J. Math.* **70** 187
[8] Marsaglia G and Bray T A 1964 *SIAM Rev.* **6** 260
   Kinderman A J and Ramage J G 1976 *J. Am. Stat. Assoc.* **71** 893
   Wieczorkowski R and Zieliński R 1997 *Komputerowe generatory liczb losowych* (Warszawa: WNT) (in Polish)
[9] Matsumoto M and Nishimura T 1998 *ACM Trans. on Modeling and Computer Simulation* **8** 3
[10] Mannella R 2002 *Int. J. Mod. Phys.* C **13** 1177 and references quoted therein
[11] Milstein G N 1974 *Theory Probab. Appl.* **19** 557