



PII: S0010-4825(96)00033-9

ON THE SIMULATION OF BIOLOGICAL DIFFUSION PROCESSES

HENRY C. TUCKWELL*^a and PETR LÁNSKÝ^b

^a Centre de Physique Théorique, CNRS Marseille F 13288 Cedex 9, France

^b Institute of Physiology, Academy of Sciences of the Czech Republic Videnska 1083, 142 20 Prague 4, Czech Republic

Abstract—Many phenomena of interest in biology can be modeled using diffusion processes satisfying a stochastic differential equation. We consider first the stochastic differential equation $dX = \mu X dt + \sigma X dW$, where W is a standard Wiener process and representing a population growth process. This is simulated using both a strong Euler scheme involving normal pseudorandom numbers and a weak Euler scheme using Bernoulli pseudorandom numbers. Results are given for the mean of $X(1)$ and its 95% confidence intervals for various numbers of simulations. It is found that there are no significant differences between the results obtained by these two schemes at a particular value of the time step, but that the weak scheme takes less computer time than the strong scheme. We also consider the process satisfying $dX = (-\gamma_1 X + \gamma_2(1-X))dt + \sqrt{X(1-X)}dW$, representing a gene frequency under the influence of random mating and mutation. It is similarly found that the results of simulation by the two schemes are not significantly different. It is concluded that in simulations of many biological diffusion processes it is often advantageous to employ a scheme involving Bernoulli rather than Gaussian random variates not only because it involves fewer machine arithmetic operations but also because problems with large jumps that sometimes occur with extreme values of normal variates are less likely, thus enabling one to employ a larger time step with a concomitant saving in machine time. © 1997 Elsevier Science Ltd

Stochastic processes Simulation Biological diffusions

INTRODUCTION

The study of the effects of chance elements in biological systems has a considerable literature. Some of the first such works involving random processes of the diffusion type arose in population genetics— see Kimura [1], Crow and Kimura [2] and Roughgarden [3] for reviews. In that field of research attention was mainly focused on the fixation of various genes under the influences of random mating, selection (both deterministic and random), migration and mutation. Not long afterwards a great deal of interest arose in random effects in the processes of population growth themselves. An early study in this area was that of Lewontin and Cohen [4], who studied Malthusian growth in a random environment. May [5] considered stochastic effects in logistic growth and a general class of growth processes satisfying a stochastic differential equation was analyzed by Tuckwell [6].

In neurophysiology, Gerstein and Mandelbrot [7] introduced a Wiener process model as an approximation for nerve membrane potential. Among the statistical properties sought was the time at which the potential first reached a threshold value. More realistic diffusion process models were introduced later, including the Ornstein–Uhlenbeck process [8,9] and a process which incorporated the synaptic reversal potentials [10,11].

In most of the work mentioned above, the random process of interest is one-dimensional diffusion process. If $X = \{ X(t), t \geq 0 \}$ is such a process, then it satisfies an (Ito) stochastic differential equation of the form,

$$dX = \alpha(X,t)dt + \beta(X,t)dW, \quad (1)$$

where α and β are given functions and where $W = \{ W(t), t \geq 0 \}$ is a standard Wiener process

*Address after September 1996: Université Paris VI CHU Saint Antoine, INSERM 444 27 rue Chaligny, 75571 Paris Cedex 12, France.

with mean zero and $\text{Var}(W(t))=t$. To analyze simple diffusion processes such as the Wiener process with drift it has proven expeditious to proceed with analytical methods, based on the forward and backward Kolmogorov equations. However, in many cases the analytical method is not useful and must be replaced by numerical methods. The latter include numerical methods for solving the associated ordinary or partial differential equations and the subject of this paper, simulation. Both of these kinds of numerical approaches are useful, but simulation enables one to work directly with the process of interest and thus gives rise to insights into its physical nature and evolution. In particular, in the study of linear stochastic partial differential equations, where solutions are composed of infinite sums of one-dimensional diffusions [12,13], simulation seems to provide an efficient method of analysis of properties which would be extremely difficult to obtain otherwise.

SIMULATION METHODS

A diffusion process which satisfies the (Ito) stochastic differential equation (1) can be simulated by a variety of methods. Several such methods are outlined and analyzed in [14,15]. In the theory of such methods a primary distinction is made between strong and weak schemes of "numerical solution" and we will briefly describe this.

Let $Y_{\Delta t}$ be a numerically generated approximation to the process X when the time step for the approximation procedure is constant at Δt . Suppose that the process is simulated from $t=0$ to $t=T$ and let $n\Delta t = T$. Let $Y_{\Delta t,k}$ be the approximation to $X(k\Delta t)$, $k=0,1,\dots,n$. Then the approximating sequence $Y_{\Delta t}$ is said to converge in the strong sense if

$$E|X(T) - Y_{\Delta t,n}| \rightarrow 0, \quad (2)$$

as $\Delta t \rightarrow 0$. This definition implies that when Δt is small enough, the expectation of the absolute value of the difference between X and $Y_{\Delta t}$ at the end-time of the simulation is small so that the paths of X and Y are not very disparate. Hence $Y_{\Delta t}$ is said to provide a good pathwise approximation to X . In a *weakly convergent* scheme, the criterion (2) is replaced by one involving closeness of expectations,

$$|E(g(X(T)) - E_g(Y_{\Delta t,n}))| \rightarrow 0, \quad (3)$$

where g is from a class of suitable real-valued test functions which might include, for example, polynomials [14]. As a straightforward example, if $g(x) = x^m$, where m is a positive integer, then (3) states that the m th moment of X is well-approximated at $t=T$ by that of Y . For simplicity we will restrict ourselves to the *Euler* schemes of approximation for Ito differential equations of the form of (1). Let

$$t_k = k\Delta t. \quad (4)$$

A commonly employed procedure for generating an approximation $Y_{\Delta t}$ is (dropping the subscript Δt for convenience) via the recursive equation

$$Y_k = Y_{k-1} + \alpha(Y_{k-1}, t_{k-1})\Delta t + \beta(Y_{k-1}, t_{k-1})N_k\sqrt{\Delta t}, \quad (5)$$

where $\{N_k, k=1,2,\dots\}$ is a sequence of (simulated) independent, identically-distributed standard normal random variables with $E(N_k)=0$ and $\text{Var}(N_k) = 1$. Here the $\{N_k\}$ are actually pseudorandom variables which can be obtained from a library program. The latter may use a linear congruential method to generate uniformly distributed variables which are transformed to normal variates by using the Box-Mueller formula. However, as pointed out by the Kloeden and Platen [14], it may be more computationally efficient to use the Polar-Marsaglia method. The simulation scheme (5) has been shown to be a *strongly convergent* scheme—it is called a *strong Euler scheme*. This method of simulation has been employed frequently to analyse biological diffusion processes. It was recently used by Lánský and Lánská [16] in their study of first passage times.

We may replace the sequence of normal random variables in (5) with a sequence of Bernoulli random variables with the same means and variances. Let $\{B_k, k=1,2,\dots\}$ be such a sequence of independent and identically-distributed Bernoulli random variables. Then we put

$$Y_k = Y_{k-1} + \alpha(Y_{k-1}, t_{k-1})\Delta t + \beta(Y_{k-1}, t_{k-1})B_k\sqrt{\Delta t}. \quad (6)$$

The $\{B_k\}$ are obtained from a (library) generator of uniform (0,1) random variables, $\{U_k\}$. One puts

$$B_k = \begin{cases} 1, & \text{if } 0 < U_k < \frac{1}{2}; \\ -1, & \frac{1}{2} < U_k < 1, \end{cases} \quad (7)$$

and clearly

$$E(B_k) = 0, \quad (8)$$

$$\text{Var}(B_k) = 1, \quad (9)$$

so $B_k\sqrt{\Delta t}$ has the correct variance Δt . The simulation scheme (6) using Bernoulli rather than normal random variables is called a *weak Euler scheme*—it converges in the sense of the definition (3) above. We will compare results for strong and weak schemes in the following section.

RESULTS AND DISCUSSION

We decided to compare the above two Euler simulation schemes for two random processes of interest in biology. The first concerns random population growth and the second fluctuations of gene frequency under random mating and mutation.

A stochastic growth process

We consider the random Malthusian growth process described by the Ito stochastic differential equation

$$dX = \mu X dt + \sigma X dW, \quad (10)$$

where μ and $\sigma > 0$ are real numbers. The solution

$$X(t) = X(0) + \mu \int_0^t X(s) ds + \sigma \int_0^t X(s) dW(s), \quad (11)$$

is a process whose transition probability density is known exactly [17]:

$$p(x, t | x_0) = \frac{1}{x\sqrt{2\pi\sigma^2 t}} \exp \left[- \frac{\left(\ln \frac{x}{x_0} - \left(\mu - \frac{\sigma^2}{2} \right) t \right)^2}{2\sigma^2 t} \right], \quad x, t > 0. \quad (12)$$

For this process are also known exactly the mean at time t ,

$$M(t) = E(X(t) | X(0) = x_0) = x_0 e^{\mu t}, \quad (13)$$

and the variance at time t ,

$$V(t) = \text{Var}(X(t) | X(0) = x_0) = x_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1), \quad (14)$$

and the corresponding standard deviation $S(t) = \sqrt{V(t)}$. The process defined by (10) also provides a model for a fluctuating stock price [18]. We simulated the process X , setting both the mean and variance parameters to unity; that is $\mu = 1$ and $\sigma = 1$. The time interval was chosen to be $[0, 1]$ and the initial value was set at $x_0 = 1$.

We examined several statistical properties of the random variable $X(1)$, and in the work reported here, set the time step $\Delta t = 0.001$, with all programs written in double precision. The computations were performed on a SUN SPARC SLC workstation. The exact values of the mean and standard deviation of $X(1)$ are $M(1) = e = 2.718\dots$ and $S(1) = e\sqrt{e-1} = 3.563\dots$. The values of the number, N , of trials, used were 20 to 1000 in steps of 20. At the end of each set of N trials we calculated the mean, variance and 95% confidence limits for $E(X(1))$ and determined the machine time employed for the simulation. Here the machine time is the number of seconds of CPU time devoted to the user's process (denoted by u in response to the *TIME* command).

Table 1. Results for random Malthusian growth

| | Strong—scheme (5) | Weak—scheme (6) |
|----------------|-------------------|-----------------|
| <i>N</i> =200 | | |
| $\bar{M}(1)$ | 2.845 | 2.855 |
| $\hat{S}(1)$ | 4.436 | 4.329 |
| <i>C.I.</i> | (2.230, 3.459) | (2.255, 3.455) |
| <i>L(CI)</i> | 1.230 | 1.200 |
| τ | 18.46 | 3.90 |
| <i>N</i> =1000 | | |
| $\bar{M}(1)$ | 2.891 | 2.850 |
| $\hat{S}(1)$ | 3.939 | 3.990 |
| <i>C.I.</i> | (2.647, 3.136) | (2.603, 3.097) |
| <i>L(CI)</i> | 0.488 | 0.495 |
| τ | 92.11 | 19.27 |

In Table 1 we give the results for $N=200$ and $N=1000$ for the strong and weak schemes. In addition to the sample mean and standard deviation, the table gives *CI*, the confidence interval for the mean and its length $L(CI)$.

It is interesting to compare the results for the scheme (5) using pseudonormal random variates with those for scheme (6) which uses pseudo-Bernoulli random variates. For $N=200$ both schemes return a value for the mean of $X(1)$ which has an error of about 4%. The estimates obtained for the standard deviation of $X(1)$ are very similar, being 4.436 for scheme (5) and 4.3294 for scheme (6). Furthermore, the 95% confidence intervals for the mean are of about the same length, being 1.230 for the strong scheme and 1.200 for the weak scheme. However, if one considers the CPU time devoted to the user's process in this set of runs, one sees that the strong scheme consumes 18.46 s, whereas only 3.90 s were used when the weak scheme is employed.

For $N=1000$ the estimated means also differ only slightly for both schemes and in each case the estimated standard deviations are much closer to the true value. The 95% confidence intervals for the mean are almost of the same length and are both much smaller than for $N=200$ as expected. Again the CPU time consumed when using the strong scheme was more than that used by the weak scheme, the respective values being 92.11 s and 19.27 s.

Similar results were obtained when we performed simulations with various other numbers of trials. Figure 1(a) shows the mean and relevant 95% confidence intervals obtained for numbers of trials between 20 and 1000 when the strong scheme (5) was employed to simulate the stochastic differential equation (6). The corresponding set of results for the weak scheme is shown in Fig. 1(b). As expected the estimate of the mean becomes consistently closer to the true value as N increases and the confidence intervals become narrower. However, there is no discernible qualitative or quantitative difference between the two sets of results.

Wright-Fisher genetical model with mutation

As a second example we consider a diffusion model for the gene frequency $X \in [0,1]$ which evolves according to the Ito stochastic differential equation

$$dX = [-\gamma_1 X + \gamma_2(1-X)]dt + \sqrt{X(1-X)}dW. \quad (15)$$

Here γ_1 and γ_2 are the mutation rates for the two alleles and the variables have been scaled [2].

We simulated this process with an initial value $X(0)=0.5$ and with mutation rate parameters $\gamma_1=0.75$ and $\gamma_2=1.25$. At first a time step of $\Delta t=0.001$ was chosen for the cases of 200 and 1000 trials as for the first example. Another advantage of the use of the weak scheme became apparent, because with this choice of time step the use of normal random numbers made the run abort due to a negative gene frequency. This occurred because a randomly chosen normal variate was of an extremely large absolute value and took the gene frequency X outside its range of $[0,1]$. With the same choice of parameters, this problem did

not occur with the scheme involving Bernoulli random numbers.

With a value of $\Delta t=0.0001$ the problem of extreme values of the normal variates was avoided and the results, being the mean, standard deviation and confidence intervals for the gene frequency at $t=0.1$ are given in Table 2 for the strong and weak schemes.

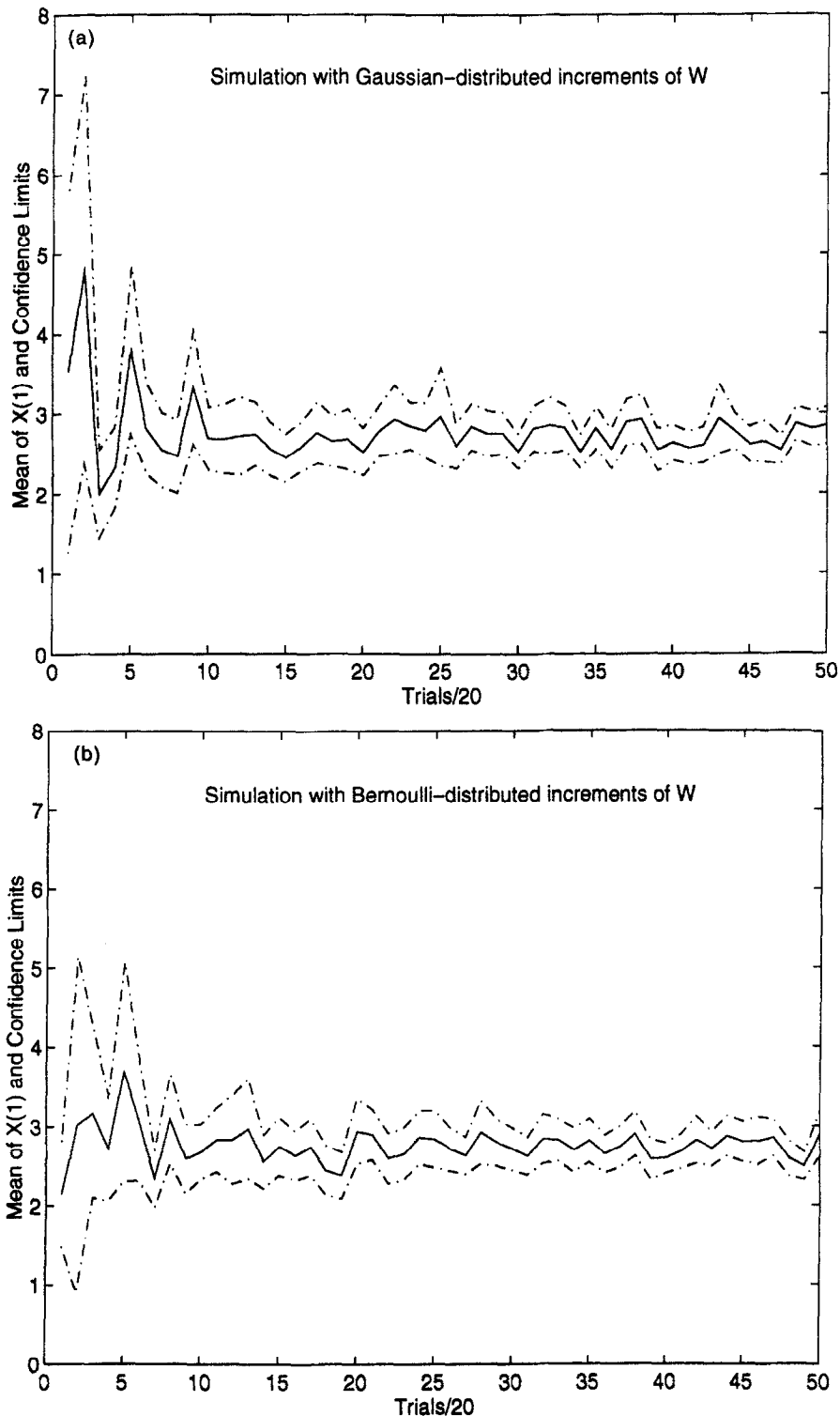


Fig. 1(a). Estimated mean of the value of $X(1)$ for the process defined by (10) with $\mu = 1$ and $\sigma = 1$ together with 95% confidence intervals when the strong Euler scheme (5) is employed, vs the number of trials. (b) As in Fig. 1(a) except that the results are shown for the weak Euler scheme (6).

Table 2. Results for Wright-Fisher diffusion with mutation

| | Strong—scheme (5) | Weak—scheme (6) |
|---------------|-------------------|-----------------|
| $N=200$ | | |
| $\hat{M}(.1)$ | 0.518 | 0.522 |
| $\hat{S}(.1)$ | 0.145 | 0.138 |
| $C.I.$ | (0.498, 0.538) | (0.504, 0.542) |
| $L(CI)$ | 0.043 | 0.038 |
| τ | 387.34 | 12.7 |
| $N=1000$ | | |
| $\hat{M}(.1)$ | 0.524 | 0.528 |
| $\hat{S}(.1)$ | 0.145 | 0.142 |
| $C.I.$ | (0.515, 0.533) | (0.519, 0.537) |
| $L(CI)$ | 0.018 | 0.018 |
| τ | 512.9 | 63.9 |

It can be seen that there are only no significant differences in the results for the statistical properties of this genetical diffusion process when the strong and weak schemes are employed. The results for the computing time consumed again lend support to the idea that the weak schemes are more efficient, although a true comparison is perhaps impeded by virtue of varying conditions of local queues for central processing times. It is apparent that the weak scheme, employing Bernoulli random variables as in (6) is preferable for this diffusion process as one may use a larger time step than that necessary for the strong scheme in order to avoid the problem of large jumps which occur when tail values of normal variates arise.

SUMMARY

Many phenomena of interest in biology can be modeled by the use of diffusion processes satisfying a stochastic differential equation. In most cases exact solutions for such models are not available and it is advantageous to proceed via computer simulations. We have considered the two main classes of simulation procedures—the so-called strong schemes involving normal random numbers and weak schemes involving Bernoulli random numbers. We have found that the statistical properties of the simulated processes are not significantly different when these two kinds of scheme are employed. Problems may arise with the use of normal random numbers when extreme values occur. This phenomenon rarely arises with Bernoulli variates and thus using a weak scheme may often enable one to use a larger time step and hence save on computer operations. There may also be an advantage in the weak schemes because the generation of Bernoulli variates requires less machine operations than the generation of normal random numbers.

Acknowledgements—The authors are grateful to the members of Laboratoire de Biométrie, Institut National de la Recherche Agronomique, Versailles, and in particular to Professor Jean-Pierre Rospars for their kind hospitality during the course of this work.

REFERENCES

1. M. Kimura, Diffusion models in population genetics, *J. appl. Prob.*, **1**, 177–226 (1964).
2. J.F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*. Harper and Row, New York (1970).
3. J. Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: an Introduction*. MacMillan, New York (1979).
4. R.C. Lewontin and D. Cohen, On population growth in randomly varying environment, *Proc. Nat. Acad. Sci. USA*, **62**, 1056–1060 (1969).
5. R.M. May, *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, New Jersey (1973).
6. H.C. Tuckwell, A study of some diffusion models of population growth, *Theor. Pop. Biol.*, **5**, 345–357 (1974).
7. G.L. Gerstein and B. Mandelbrot, Random walk models for the spike activity of a single neuron, *Biophys. J.*, **4**, 41–68 (1964).
8. B. Gluss, A model for neuron firing with exponential decay, *Bull. Math. Biophys.*, **29**, 233–243 (1967).
9. R.M. Capocelli and L.M. Ricciardi, Diffusion approximation and first passage time problem for a model neuron, *Kybernetik*, **8**, 214–223 (1971).

10. F.B. Hanson and H.C. Tuckwell, Diffusion approximations for neuronal activity including synaptic reversal potentials, *J. Theoret. Neurobiol.*, **2**, 127–153 (1983).
11. P. Lánský and V. Lánská, Diffusion approximation to the neuronal model with synaptic reversal potentials, *Biol. Cybern.*, **56**, 19–26 (1987).
12. H.C. Tuckwell and J.B. Walsh, Random currents through nerve membranes, *Biol. Cybern.*, **49**, 99–110 (1983).
13. H.C. Tuckwell, F.Y.M. Wan and Y-S. Wong, The interspike interval of a cable model neuron with white noise inputs, *Biol. Cybern.*, **49**, 155–167 (1984).
14. P.E. Kloeden and E. Platen, *The Numerical Solution of Stochastic Differential Equations*. Springer, Berlin (1992).
15. G. N. Milstein, *Numerical Integration of Stochastic Differential Equations*. Kluwer, Dordrecht (1994).
16. P. Lánský and V. Lánská, First passage time problem for simulated stochastic diffusion processes, *Comput. Biol. Med.*, **24**, 91–101 (1994).
17. H.C. Tuckwell, Diffusion processes in biology. SM Thesis, University of Chicago (1973).
18. P. Samuelson, Rational theory of warrant pricing, *Indust. Manag. Rev.*, **6**, 13–31 (1965).

About the Author—HENRY C. TUCKWELL received his PhD in Biophysics and Theoretical Biology at the University of Chicago, U.S.A., in 1974. He has worked at the University of British Columbia, University of California, Los Angeles, Monash University and has been a visiting professor at University of Washington, Australian National University and Centre de Physique Théorique, CNRS, Marseille, France. His current research is in theoretical neurobiology and applications of mathematics to medicine. He is the author of “Introduction to Theoretical Neurobiology”, Volumes 1 and 2, “Stochastic Processes in the Neurosciences” and “Elementary Applications of Probability Theory: With an Introduction to Stochastic Differential Equations” as well as several research papers.

About the Author—PETR LÁNSKÝ received his Diploma in Mathematics in 1972 and his PhD in Probability Theory and Mathematical Statistics in 1982 from Charles University, Prague, Czech Republic. He has been at the Institute of Physiology of the Academy of Sciences of the Czech Republic since 1975 and has been a visiting professor at North Carolina State University, Raleigh, U.S.A. and Laboratoire de Biométrie, Institut National de la Recherche Agronomique, Versailles, France. His current research interests are biological stochastic processes and the analysis of neurobiological data.