Population Dynamics: Poisson Approximation and Its Relation to the Langevin Process

Juan P. Aparicio* and Hernán G. Solari†

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón I, Ciudad Universitaria, 1428 Buenos Aires, Argentina

(Received 17 April 2000)

We discuss how to simulate a stochastic evolution process in terms of difference equations with Poisson distributions of independent events when the problem is naturally described by discrete variables. For large populations the Poisson approximation becomes a discrete integration of the Langevin approximation [T. G. Kurtz, J. Appl. Prob. **7**, 49 (1970); **8**, 344 (1971)]. We analyze when the latter gives a reasonable representation of the original evolution for finite size systems. A simple example of an epidemic process is used to organize the discussion and to perform statistical tests that underline the goodness of the proposed method.

DOI: 10.1103/PhysRevLett.86.4183 PACS numbers: 87.19.Xx, 02.50.Ey, 05.40.Ca, 87.23.Kg

The description and analysis of interacting populations with a discrete number of individuals is a common subject in several branches of physics and other sciences [1,2]. The members of the population can be molecules (as in chemical reactions) [3–7], photons (as in lasers) [8–10], or individual organisms (as in population biology) [11].

We are particularly interested in stochastic systems which have a certain kind of deterministic limit. Namely, for some size parameter Ω large enough, the solutions of individual realizations follow closely the solutions of a system of ordinary differential equations [12–15] (see [16] for a recent rediscovery of the inverse problem).

Early results in the physics of Brownian motion led physicists to simulate population dynamics in terms of a Langevin process. The approximation of the evolution equations for the probabilities of such systems can performed following Van Kampen's Ω expansion [1]. The Fokker-Planck equations [1,17,18] resulting from these approximations suggest that individual realizations of the stochastic dynamics follow a Langevin equation with Gaussian noise (see [14,15,19] for rigorous results). One of the conceptual difficulties with this presentation is that it cannot be used to simulate individual realizations of the stochastic process since the probability distributions for sufficiently small times cannot be considered smooth, a requirement for the applicability of the approximation.

In contrast with early attempts, recent developments have come to emphasize the "importance of being discrete" [20,21], i.e., that the replacement of populations that take discrete values by continuous variables can lead to large errors.

In what follows we introduce and discuss a finite difference Poisson approximation of the stochastic process which respects the discreteness of the populations, has the exact process as its infinitesimal (in time) limit, and can be, in turn, approximated by a Gaussian-Langevin process under conditions that are discussed. We also emphasize the relation with previous work by Kurtz.

To avoid making the present discussion unnecessarily abstract, we focus our attention on an epidemiological model for measles consisting of six stochastic events: birth and death of a susceptible individual, infection, recovery of an infected individual, death of an infected individual, and migration into the system (that can be thought of as a city) of an infected individual.

The transition probabilities and effects on the populations of susceptible (S) and infected (I) individuals exerted by each of the stochastic events in the model is presented in Table I. The interval of time between events is a randon variable exponentially distributed with mean $\tau = [w_b + w_f + w_c(S,I) + w_{ds}(S) + w_{di}(I) + w_r(I)]^{-1}$.

The model for the disease implemented corresponds to an open community (city) [11]. The model is related to those used for closed communities (no migration of infective individuals) that has been a traditional and important subject of study in mathematical epidemiology. For more

TABLE I. Event type, effects on the populations, and transition rates for the epidemic model.

Event	Effect	Transition rate
Birth	$(S,I) \to (S+1,I)$	$w_b \equiv \mu\Omega$
Inflow of infected	$(S,I) \to (S,I+1)$	$w_f \equiv \epsilon\Omega$
Contagion	$(S,I) \to (S-1,I+1)$	$w_c(S,I) \equiv \beta SI/\Omega$
Death susceptible	$(S,I) \to (S-1,I)$	$w_{ds}(S) \equiv \delta S$
Death infected	$(S,I) \to (S,I-1)$	$w_{di}(I) \equiv \delta I$
Recovery of infected	$(S,I) \to (S,I-1)$	$w_r(I) \equiv \gamma I$

information, readers are referred to a recent work by Nåsell [22]. We emphasize that the open community model does not have an absorbing state while the closed community does.

The master equations describing the epidemic process read

$$\frac{dP(S,I)}{dt} = w_b P(S-1,I) + w_f P(S,I-1) + w_c(S+1,I-1)P(S+1,I-1) + w_{ds}(S+1)P(S+1,I) + w_{di}(I+1)P(S,I+1) + w_r(I+1)P(S,I+1) - [w_b + w_f + w_c(S,I) + w_{ds}(S) + w_{di}(I) + w_f(I)]P(S,I) \quad (1)$$

and the associated deterministic differential equation reads

$$\frac{ds}{dt} = \mu - \delta s - \beta si\,, \qquad \frac{di}{dt} = \epsilon + \beta si - (\gamma + \delta)i\,, \quad (2)$$

where $s = S/\Omega$ and $i = I/\Omega$ are approximated by real numbers in the $\Omega \to \infty$ limit.

The total population consist of susceptible, S, infected, I, and recovered, R, individuals. The recovered individuals are thought to have acquired full immunity and do not affect the dynamics of (S, I). The recovered population obeys the equation

$$\frac{dR}{dt} = \gamma I - \delta R\,. \quad (3)$$

The scale factor Ω can be thought of as the total population number at the fixed point of the equation for the total population, i.e., $R + I + S = \Omega(\mu + \epsilon)/\delta$ for long enough times.

A fixed time-step (coarse) realization of the stochastic process can be realized throwing a "dice" with probability $P(S, I, \Delta t; S_0, I_0)$. This probability is a solution of (2) with initial condition $P(S, I, 0; S_0, I_0) = 1$ if $S = S_0$ and $I = I_0$ and zero otherwise. Hence, our interest is to propose a reasonable approximation to this transition probability.

The zeroth order approximation can be given in intuitive terms. If as a result of a relatively small number of steps in the random walk [of the population in the (S, I) space], the population numbers are unlikely to change significantly, and, consequently, the transition rates are likely to suffer only small alterations, we can approximate the transition rates by their initial values at $t = 0$. In this approximation each of the events respond to a Poisson distribution with $\lambda_i = \Delta t w_i$, where $i \in \{b, f, c, ds, di, r\}$.

Note that the basis for the heuristic is not directly the smallness of the time interval but rather the smallness of a likely change in the relative value of the populations.

The changes in the infected populations during the time interval Δt can be written as

$$\Delta S = n_b - n_c - n_s\,, \qquad \Delta I = n_f + n_c - n_i - n_r\,, \quad (4)$$

where n_b is the number of newborn, n_c is the number of infected individuals, n_s is the number of susceptible people death, n_f is the inflow of infected individuals, and n_r is the number of recovered individuals, all occurring in the time interval Δt. Each of the variables presents an independent Poisson statistic as explained. We further notice that in the limit $\Delta t \to 0$ this approximation corresponds with the original process.

Alternatively, the process (I) can also be represented by [13–15]

$$Z(t) = Z(0) + \sum_l \delta_l Y_l\left(\Omega \int_0^t w_l[Z(s)/\Omega]\,ds\right), \quad (5)$$

where $Z = (S, I)$, the index l runs over all the events, $l \in \{b, f, c, ds, di, r\}$, δ_l is the vector representing the change in the populations for the event l, and $Y(\lambda)$ represents a Poisson process with mean λ. The approximation introduced heuristically is equivalent to neglecting the time dependence of the stochastic variables on the right side of Eq. (5), an approximation equivalent to a Euler scheme for the Ito differential process [23].

A consistent estimation of the conditions of validity invoked in the heuristic arguments gives

$$\langle \Delta S^2 \rangle = (\Omega \Delta t)^2(\mu - \delta s - \beta si)^2 + \Delta t(w_b + w_{ds} + w_c) \ll S^2, \qquad \langle \Delta I^2 \rangle = (\Omega \Delta t)^2[\epsilon + \beta si - (\gamma + \delta)i]^2 + \Delta t(w_f + w_c + w_{di} + w_r) \ll I^2. \quad (6)$$

From (6) we see that there are two contributions, the first one corresponds to the deterministic drift and requires a small enough Δt for Eq. (2) to be realistically integrated, while the second contribution is of stochastic nature and dominates near the fixed point of the deterministic approximation.

The increment per unit time of the mean values scale with Ω while the stochastic contribution of (6) presents a dispersion that scales with $\Omega^{1/2}$ and becomes negligible in the Ω (deterministic) limit except perhaps for $\lim_{\Omega\to\infty} I/\Omega = i = 0$ or $\lim_{\Omega\to\infty} S/\Omega = s = 0$.

The marginal distribution of ΔI (ΔS) can be written in terms of $p_i = \Delta t(w_f + w_c)$, $q_i = \Delta t(w_{di} + w_r)$ [$p_s = \Delta t w_b$ $q_s = \Delta t(w_c + w_{ds})$] as the distribution of the difference between two independent Poisson-distributed random variables.

When $p_i, q_i \gg 1$ the Poisson distributions can be approximated by Gaussian functions and the same can be

said for the marginal distribution. Hence the marginal distributions are approximately normal when

$$p_i, q_i \gg 1, \qquad p_s, q_s \gg 1\,. \tag{7}$$

These conditions are necessary for the Langevin approximation to be a reasonable difference integration scheme of the stochastic process.

In the Langevin approximation the increments of the populations in the interval of time Δt take the form

$$\begin{aligned} \Delta S &= \Delta t \Omega(\mu - \delta s - \beta s i) + \psi_1 - \psi_c\,, \\ \Delta I &= \Delta t \Omega[\epsilon + \beta s i - (\gamma + \delta) i] + \psi_c + \psi_2\,, \end{aligned} \tag{8}$$

where ψ_i, $i = 1, 2$, c are Gaussian random variables with variance $\sigma_1^2 = \Delta t(w_b + w_{ds})$, $\sigma_2^2 = \Delta t(w_f + w_{di} + w_r)$, and $\sigma_c^2 = \Delta t w_c$, respectively.

Equation (8) is a numerical implementation of the Langevin process [23]. However, to take the limit $\Delta t \to 0$ on this process is definitely wrong since the conditions for its introduction would be violated.

We further add that while we have been able to find sufficient conditions for (8) to be an approximation of the process (2), none of them is optimal, and in our experience the necessary conditions (7) are a better guide to applicability.

In what follows we compare the stochastic process [Eq. (2) and Table I] and its approximations (4),(8) for three different sets of parameter values for the model (I) differing only in the values of Ω. The remaining parameter values are $\mu = 1/45$, $\epsilon = (1/45)\,(1/2000)$, $\beta = 360$, $\delta = 1/45$, $\gamma = 180/7$, and the units of all the parameters are year^{-1}. A time step of $\Delta t = 10^{-3}$ years has been used for the integration of the Poisson and Langevin approximations. Representative runs (realizations) of 100 years are obtained for the original stochastic process (two independent runs), the Poisson approximation (4), and the Langevin approximation (8).

The time series are sampled at the same rate (every 10^{-2} years) so that the finer structure of the original process does not affect the comparison.

Since the approximations can eventually produce meaningless negative populations we have set the involved population to zero every time it happens (less than once every 10^4 steps in the worst cases with population value of $I = -1$). We notice that the Poisson approximation will propagate the absence of infected individuals ($I = 0$) until a new infected individual arrives without need of new corrections. A different possibility to avoid this problem is to reduce the time step of the integration for small populations in the case of the Poisson approximation.

We use as discriminating statistics the maximum number of ill people during an epidemic outbreak. We consider that an epidemic outbreak begins (ends) when the number of infected members of the population satisfy $I >$ th ($I <$ th) th being a *detection threshold* consistently taken as the average number of ill people as predicted by the deterministic dynamics of Eq. (2), i.e., th $\sim I_0$.

The maximum infected population detected during epidemic outbreaks are put in three bins based in a single reference run of the original process. If $I_{\max}$ is the maximum number of infected individuals during the reference run, the three bins are $[0, (1/3)I_{\max}]$, $[(1/3)I_{\max}, (2/3)I_{\max}]$, and $[(2/3)I_{\max}, \infty]$. The results are presented in Table II. The number associated to each bin corresponds to the average in 100 runs of the number of peaks with the maximum number of infected individuals belonging to the bin.

We set the significance level to 0.10 meaning that one in every ten realizations of the original process will be mistakenly discriminated by the test; i.e., the significance level is set to a high value. The value for the χ^2 for the averaged distribution is $\chi_0^2 = 6.25/100 = 0.0625$. Values of χ^2 larger than χ_0^2 imply the rejection of the hypothesis that the tested distribution corresponds to the original process.

We notice that the χ^2 test is sensible to the average number of epidemic outbreaks present in each case. We further test the probability distribution associated to our discriminating statistics (obtained with the numerical simulations) comparing them with a Kolmogorov-Smirnov (KS) test [24].

In the examples shown it is not possible to reject the Poisson approximation nor a different realization of the original processes in any case at the significance level fixed.

The Langevin approximation, however, can be rejected in the cases $\Omega = 10^4$ and $\Omega = 10^5$. For $\Omega = 10^6$ the χ^2 test does not discriminate the Langevin approximation, but the KS test does by a small margin.

TABLE II. Comparison of the statistics for the stochastic process and its approximations for different values of the size parameter Ω. We show the average frequency for the three bins of the maximum epidemics values, the χ^2 value resulting from the comparison with the original process and the acceptance probability of the KS test. The data labeled original and original-2 refer to the original stochastic process run with a different seed for the random number generator, i.e., a different realization of the same stochastic process.

	$\Omega = 10^4$, th = 10					$\Omega = 10^5$, th = 100					$\Omega = 10^6$, th = 1000				
Process	Bin 1	Bin 2	Bin 3	χ^2	KS	Bin 1	Bin 2	Bin 3	χ^2	KS	Bin 1	Bin 2	Bin 3	χ^2	KS
Original	5.01	1.62	0.88			31.88	3.46	0.87			67.73	12.78	8.99		
Poisson	5.59	1.68	0.77	0.04	0.25	30.98	3.58	0.90	0.015	0.42	67.70	12.75	8.75	0.02	0.89
Langevin	104.16	0.00	0.00	92.54	$<10^{-6}$	70.66	0.00	0.00	18.99	$<10^{-6}$	65.63	13.59	9.21	0.03	0.02
Original-2	5.18	1.67	0.83	0.005	0.87	31.32	3.22	0.96	0.02	0.53	67.10	13.45	8.5i0	0.03	0.26

The performance of the Poisson approximation was excellent in all cases considered, even though, when $\Omega = 10^4$, 10^5 there are large periods of time when the infected population is zero and the conditions invoked in the heuristic arguments (6) are not satisfied. The reason for the success of the Poisson approximation for small populations is related to the fact that for $\Delta t \to 0$ it has the same limit and derivative as the original process.

It is important to notice that the mean time between epidemics obtained with the Poisson approximation presents a dependence with the "size of the city" (Ω) qualitatively similar to the dependence reported in the classical study by Bartlett [25,26]. In contrast, the Langevin approximation cannot account for this dependence.

There are several ways in which the Poisson approximation introduced should be improved. First, the algorithm should produce higher order approximations (in Δt) to the integration of (2) in the deterministic limit. Second, the *ad hoc* mechanism used to avoid negative populations should be replaced by one built in the approximation; this step might prove to be a difficult one since the main reason for the occurrence of nonzero probabilities for negative populations is the failure to satisfy the conditions (6) at very low populations. Third, linear approximations to the transition rates can be used. The need for these improvements is not yet clear given the excellent performance obtained.

In conclusion, we have introduced a Poisson approximation to the stochastic process describing population dynamics. This approximation produces reliable numerical simulations of the dynamics and has the proper limits for infinitesimal times and for large populations (the Langevin process). It is then able to produce reliable results in a range of situations substantially larger that the Langevin approximation with equivalent simplicity. One of the main reasons for these abilities is that the approximation is respectful of the discrete character of the populations.

We acknowledge support of the University of Buenos Aires under Grant No. TW04 and a grant from Fundación Antorchas. We thank Angel Capurro, Gabriel B. Mindlin, Ingemar Nåsell, Mario A. Natiello, Maxi San Miguel, and Martín Zimmermann for useful discussions.

*Present address: Department of Biometrics, 432 Warren Hall, Cornell University, Ithaca, New York 14853-7801. Electronic address: jpa9@cornell.edu

†Electronic address: solari@df.uba.ar

[1] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[2] T. G. Kurtz, in *Stochastic Nonlinear Systems in Physics, Chemistry, and Biology,* edited by L. Arnold and R. Lefever (Springer, Berlin, 1981).
[3] P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, London, 1971).
[4] R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975).
[5] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
[6] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[7] J. Güémez and M. A. Matías, Phys. Rev. E **51**, 3059 (1995).
[8] C. H. Henry, IEEE J. Quantum Electron. **18**, 259 (1982).
[9] C. H. Henry, IEEE J. Quantum Electron. **19**, 1391 (1983).
[10] W. W. Chow, S. W. Koch, and M. Sargent, *Semiconductor-Laser Physics* (Springer-Verlag, Berlin, 1994).
[11] E. Renshaw, *Modelling Biological Populations in Space and Time* (Cambridge University Press, Cambridge, England, 1991).
[12] T. G. Kurtz, J. Appl. Prob. **7**, 49 (1970).
[13] T. G. Kurtz, *Approximation of Population Processes* (Society for Industrial and Applied Mathematics, Philadelphia, 1981).
[14] S. N. Ethier and T. G. Kurtz, *Markov Processes* (John Wiley and Sons, New York, 1986).
[15] H. Andersson and T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis,* Lecture Notes in Statistics Vol. 151 (Springer-Verlag, Berlin, 2000).
[16] O. E. Aièllo, V. J. Haas, M. A. A. daSilva, and A. Caliri, Physica (Amsterdam) **282A**, 546 (2000).
[17] H. Risken, *The Fokker-Planck Equation, Methods of Solution and Application,* Springer Series in Synergetics Vol. 18 (Springer, Heidelberg, 1989), 2nd ed.
[18] H. Risken, in *Statistical Properties of Light* (Springer, Berlin, 1996), Chap. 12.
[19] T. G. Kurtz, J. Appl. Prob. **8**, 344 (1971).
[20] R. Durrett and S. A. Levin, Theor. Popul. Biol. **46**, 363 (1994).
[21] N. M. Shnerb, Y. Louzoun, E. Bettlheim, and S. Solomon, Proc. Natl. Acad. Sci. U.S.A. **97**, 10 322 (2000).
[22] I. Nåsell, J. R. Statist. Soc. B **61**, 309 (1999).
[23] W. Rümelin, SIAM J. Numer. Anal. **19**, 604 (1982).
[24] For the Langevin method the maximum epidemics values are converted to the closest integer number before running the test; otherwise, the KS test consistently returns null probabilities for all the cases.
[25] M. S. Bartlett, J. R. Statist. Soc. A **120**, 48 (1957).
[26] M. S. Bartlett, J. R. Statist. Soc. A **123**, 37 (1960).