Emergence of patterns in random processes. II. Stochastic structure in random events

William I. Newman*
Departments of Earth, Planetary, & Space Sciences, and Physics & Astronomy, and Mathematics, University of California, Los Angeles, California 90095-1567, USA
(Received 1 August 2013; revised manuscript received 30 April 2014; published 9 June 2014)

Random events can present what appears to be a pattern in the length of peak-to-peak sequences in time series and other point processes. Previously, we showed that this was the case in both individual and independently distributed processes as well as for Brownian walks. In addition, we introduced the use of the discrete form of the Langevin equation of statistical mechanics as a device for connecting the two limiting sets of behaviors, which we then compared with a variety of observations from the physical and social sciences. Here, we establish a probabilistic framework via the Smoluchowski equation for exploring the Langevin equation and its expected peak-to-peak sequence lengths, and we introduce a concept we call "stochastic structure in random events," or SSRE. We extend the Brownian model to include antipersistent processes via autoregressive (AR) models. We relate the latter to describe the behavior of Old Faithful Geyser in Yellowstone National Park, and we devise a further test for the validity of the Langevin and AR models. Given our analytic results, we show how the Langevin equation can be adapted to describe population cycles of three to four years observed among many mammalian species in biology.

DOI: 10.1103/PhysRevE.89.062113 PACS number(s): 02.50.Ey, 89.75.Kd, 05.45.Tp, 87.23.Cc

I. INTRODUCTION

When we view time series or other point process data, we try to identify a pattern in what we see, but ultimately we must address the question of whether the pattern is real or illusory. We usually expect that "random uncorrelated events," in space and/or time, should be devoid of any underlying order.

A classic example emerges in wildlife management, where the population of various animal species in different locations is tracked on an annual basis. Cole [1] presented data gathered by many others relating to the Arctic fox and wolf populations in Canada. He looked for apparent regularity in population peaks, which he and others before him identified using the years in which the population was greater than the population in both the previous and succeeding years. Cole and others observed what appeared to be a three- to four-year cycle. This ostensible cycle was, in part, the basis for research over many decades on predator-prey cycles plus a host of environmental influences, including spatial variability, etc. However, as Cole noted, he was able to observe a similar three- to four-year cycle in plotting random numbers from published tables that were in common use at that time.

His mathematician colleague, Mark Kac, provided a simple argument establishing that three-year or three-event "cycles" would emerge for random events, i.e., an independent and identically distributed (i.i.d.) time series such as Gaussian white noise. Kac's argument, presented as a brief footnote, was essentially that the probability that a given point is a peak requires that you compare its value with its immediate predecessor and successor. For i.i.d. random variables, the likelihood that it is a peak is therefore $1/3$. Hence, if $1/3$ of all events are peaks, then there must be on average three events per peak-to-peak sequence.

Newman *et al.* [2] succeeded in finding the distribution of peak-to-peak sequences for i.i.d. random variables as a function of the number of intervening events. They went on to show that a pure "Brownian walk" would also manifest similar cyclical or clustering behavior. In a Brownian walk, the amplitude of the middle of the three events mentioned above is assumed to have the same value as the preceding event plus a randomly selected addition. Further, we assume that the final event has the same value as the middle one plus another randomly selected addition. Assuming that the distribution of the additive random variable has a vanishing median, the probability of each of those two independent steps is $1/2$, making the likelihood that the central point is a peak $1/4$. Hence, there must be on average four events per peak-to-peak sequence. Newman *et al.* [2] also succeeded in finding the distribution of peak-to-peak sequences for Brownian walk associated random variables as a function of the number of intervening intervals.

Newman *et al.* [2] went on to consider a family of models that interpolate smoothly between the i.i.d. case and the Brownian walk situation. This is elaborated in [2] and is intimately related to the discrete form of the Langevin equation encountered in statistical mechanics [2–4], namely

$$x_{n+1} = \alpha x_n + \eta_n. \tag{1}$$

The fundamental equation (1) is shown above, where the parameter $0 \leqslant \alpha \leqslant 1$. Here, we observe that $\alpha = 0$ corresponds to the i.i.d. problem while $\alpha = 1$ corresponds to the Brownian walk problem. The case $\alpha = 1$ is a special and singular case, and its formal treatment is sometimes referred to as "renewal theory" [5–7] and requires specialized methods for its treatment. The fundamental results emergent from that case for peak-to-peak cycles were derived in [2]. In the original formulation of the Langevin problem, it was generally assumed that the distribution function describing the random step η_n had a thermodynamic origin and, hence, a Maxwell-Boltzmann distribution, i.e., a Gaussian.

Equation (1) also appears in a classic problem of time-series and spectral analysis, where it is known as a (first order) "autoregressive" (AR) model [8–10]. This model is

*win@ucla.edu

commonly used in a variety of signal processing applications in circumstances where $-1 \leqslant \alpha \leqslant 1$. As we will shortly see, we will now extend the domain of interest for α to include negative values that are smaller than unity in magnitude. While $\alpha > 0$ intuitively relates to situations in which there is some memory present or "persistence" of the preceding event, it follows that $\alpha < 0$ relates to "antipersistent" issues or negative feedback. Assuming that the process that produces η_n is stationary, i.e., invariant over time, it is immediately observed that we can estimate α according to

$$\alpha = \frac{E(x_{n+1}x_n)}{E\left(x_n^2\right)} = \frac{E(x_{n+1}x_n)}{E\left(x_{n+1}^2\right)}, \tag{2}$$

where we have used $E(\cdots)$ to designate the mean or expected value with respect to the quantity enclosed in parentheses. The properties of Eq. (1) depend critically on the inequality present in our description of α since this variable appears in infinite sums of the form $1 \pm \alpha + \alpha^2 \pm \alpha^3 \ldots$ as well as other expressions where $\alpha = \pm 1$ would result in divergent results. Hence, the methods of renewal theory are very different from those that we apply here for $|\alpha| < 1$. We will return to the issue of estimating α later in this paper.

Newman *et al.* went on to compare the statistics of observed complex phenomenon, including the magnitudes and time intervals between great earthquakes, the time intervals between auroral electrojet index events as a measure of the Earth's magnetospheric response to solar activity, the time intervals between successive eruptions of Old Faithful Geyser in Yellowstone National Park, and the Standard and Poors 500 relative change in daily closing stock market prices. Agreement between these natural and social phenomena with the theory presented in [2] varied from excellent (earthquakes) to poor (Old Faithful Geyser), and it provided an empirical basis for a methodology that we presently call "stochastic structure in random events" (SSRE). We will return later to the situation presented by Old Faithful and show that it is consistent with antipersistent behavior.

Newman *et al.* performed Monte Carlo simulations assuming an underlying uncorrelated Gaussian random variable to develop an appreciation for the variability of the cycle or sequence length as a function of $0 \leqslant \alpha \leqslant 1$. What we observed in [2] validated our conjecture that random events could manifest in patterns or structures despite having an underlying stochastic character. In this paper, we approach the problem of establishing the mean peak-to-peak sequence length as a function of α using analytic methods derived from probability theory, especially the Smoluchowski equation. The outcome of the present investigation is that we have a closed-form expression for the cycle length as a function of α that smoothly interpolates between the two extremes, and it agrees within the expected statistical error with the Monte Carlo simulation results presented in Fig. 7 in [2]. Moreover, it extends the range of α to negative values, thereby providing insight into peak-to-peak cycles that will be observed in antipersistent phenomena. As an illustration, we will apply our analytic results to the Old Faithful Geyser data considered in Newman *et al.* We now claim that the results presented here and in [2] demonstrate SSRE, the concept that random processes can manifest what appears to be organized behavior.

II. SMOLUCHOWSKI AND FOKKER-PLANCK EQUATIONS

Einstein [11] pioneered the theory underlying Brownian motion. Later, Smoluchowski [12] extended the coordinate-space definition of the problem to account for the influence of an external force on the Brownian particle. Importantly, he established a probabilistic approach to the problem. More accessible treatments are available in textbooks [4,13]. The problem at hand is more complicated than that which can readily be expressed using his methodology. As we shall see, we will need to address an infinite sequence of convolutions that introduces a profound methodological barrier to solving this problem. However, using the method of characteristic functions [6,14,15] that exploit the properties of Fourier transforms, these complications can be readily overcome, as we show below.

We shall assume that the probability density function $p(\eta_n)$ that describes the distribution of steps η_n taken is known. In many instances, it can be regarded as normally distributed. Moreover, it is stationary, i.e., it is independent of the step n taken. Without loss of generality, we shall assume that it has a zero mean and unit variance. We wish to obtain the relationship between the probability density function $f_{n+1}(x_{n+1})$ given that we know $f_n(x_n)$ and $p(\eta_n)$. We shall assume that we know the original probability distribution function $f_0(x_0)$. Further, we shall assume that it is initially the same as the distribution function for steps, namely

$$f_0(x_0) = p(x_0). \tag{3}$$

We have assumed that it has a mean of zero; as is well known for the Langevin equation, the fluctuation-dissipation theorem relates the variance of x_n to that of η. Importantly, the variance will remain finite as long as $|\alpha| < 1$.

To find the functional relationship between the distribution functions f_n, we employ the method of characteristic functions. To do this, we need to define Fourier transform pairs for all relevant quantities, and we shall employ an overlying caret to designate a transformed quantity. Hence, we have that

$$\hat{p}(k) = \int_{-\infty}^{\infty} d\eta\, p(\eta)\, \exp(ik\eta), \tag{4}$$

and so on. The inverse transform, in general, will satisfy

$$p(\eta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, \hat{p}(k)\, \exp(-ik\eta). \tag{5}$$

Since we are dealing with probability densities, it will generally be the case that the transform evaluated at 0 will be unity, e.g.,

$$\hat{p}(0) = \int_{-\infty}^{\infty} d\eta\, p(\eta) = 1. \tag{6}$$

Accordingly, invoking Eq. (1), it follows that

$$\begin{aligned} \hat{f}_{n+1}(k) &\equiv \int_{-\infty}^{\infty} dx_{n+1}\, f_{n+1}(x_{n+1})\, \exp(ikx_{n+1}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_n\, d\eta_n\, f_n(x_n)\, p(\eta_n)\, \exp[ik(\alpha x_n + \eta_n)] \\ &\quad \times \hat{f}_n(k\alpha)\hat{p}(k). \end{aligned} \tag{7}$$

Before proceeding, we observe that the characteristic function is the Fourier transform of a distribution function and, accordingly, presents all moments of the distribution via the associated powers of k. For example, we observe that

$$\hat{p}(k) = E(\exp ik\eta)_\eta = \sum_{\ell=0}^{\infty} \frac{i^\ell k^\ell}{\ell!} E(\eta^\ell)_\eta. \tag{8}$$

In principle, therefore, knowledge of all moments of a distribution function allows us to reconstruct it exactly. In this way, we are obtaining the relationship between the distribution functions for all of our random variables.

Before proceeding, we wish to employ Eq. (7) to obtain a general formula for $\hat{f}_n(k)$. It follows immediately that

$$\hat{f}_1(k) = \hat{f}_0(k\alpha)\hat{p}(k) = \hat{p}(k\alpha)\hat{p}(k), \tag{9}$$

where we employed Eq. (3). Similarly, it follows that

$$\hat{f}_2(k) = \hat{f}_1(k\alpha)\hat{p}(k) = \hat{p}(k\alpha^2)\hat{p}(k\alpha)\hat{p}(k), \tag{10}$$

and, by induction, that

$$\hat{f}_n(k) = \prod_{m=0}^{n} \hat{p}(k\alpha^m). \tag{11}$$

Since $\lim_{k\to 0} \hat{p}(k) = 1$, we generally expect that the limit of this sequence will exist for $|\alpha| < 1$.

As an illustrative and commonly encountered example, suppose that $p(\eta)$ is a Gaussian with zero mean and unit variance, say $g(\eta)$ given by

$$g(\eta) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{\eta^2}{2}\right). \tag{12}$$

It is easy to show that

$$\hat{g}(k) = \exp\left(-\frac{k^2}{2}\right), \tag{13}$$

which is itself a Gaussian in k with zero mean and unit variance. We can now calculate the corresponding value of $\hat{f}_g$ in the limit $n \to \infty$ for $|\alpha| < 1$, namely

$$\begin{aligned} \hat{f}_g(k) &= \lim_{n\to\infty} \prod_{m=0}^{n} \exp\left[-\frac{k^2\alpha^{2m}}{2}\right] \\ &= \exp\left[-\frac{k^2}{2}\sum_{m=0}^{\infty}\alpha^{2m}\right] \\ &= \exp\left[-\frac{k^2}{2(1-\alpha^2)}\right]. \end{aligned} \tag{14}$$

This can be Fourier inverse-transformed immediately to give

$$f_g(x) = \sqrt{\frac{1-\alpha^2}{2\pi}} \exp\left[-\frac{(1-\alpha^2)x^2}{2}\right]. \tag{15}$$

This latter result, valid for $|\alpha| < 1$, will be employed in the remainder of this paper.

We now wish to inverse-transform Eq. (7) to obtain our desired result in the original domain, in contrast with its Fourier representation. Accordingly, let us inverse-transform the following relation:

$$\begin{aligned} f_{n+1}(y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \hat{f}_{n+1}(k)\exp(-iky) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \hat{f}_n(k\alpha)\hat{p}(k)\exp(-iky). \end{aligned} \tag{16}$$

We now introduce the relationship between the Fourier transform of f_n and p into the latter to obtain

$$\begin{aligned} f_{n+1}(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \exp[-iky]\left(\int_{-\infty}^{\infty} f_n(x)\ \exp[ik\alpha x]\right) \\ \times\left(\int_{-\infty}^{\infty} p(\eta)\exp[ik\eta]\right). \end{aligned} \tag{17}$$

We perform the k integral first and employ the usual relations for Dirac δ functions, namely

$$\begin{aligned} f_{n+1}(y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx\, d\eta\ f_n(x)\, p(\eta) \\ &\quad\times \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \exp[ik(\alpha x + \eta - y)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx\, d\eta\ f_n(x)\, p(\eta)\delta(\alpha x + \eta - y) \\ &= \frac{1}{\alpha}\int_{-\infty}^{\infty} d\eta\ f_n\left[\frac{y-\eta}{\alpha}\right] p(\eta). \end{aligned} \tag{18}$$

This is the usual formulation of the Smoluchowski relation [13], sometimes called the Fokker-Planck equation [4]. As $\alpha \to 0$, the $\frac{1}{\alpha} f_n[\frac{y-\eta}{\alpha}]$ term effectively turns into a Dirac δ function making $f_n(x)$ essentially the same as $p(x)$. This seeming singularity can be eliminated by a simple variable transformation, namely

$$x \equiv \frac{y-\eta}{\alpha}; \tag{19}$$

this can be turned around and written as

$$y = \alpha x + \eta, \tag{20}$$

which is our original Langevin equation. Thus, this behaves like the mapping that takes us from x due to the random step η to our new value y. With this transformation, it follows that

$$f_{n+1}(y) = \int_{-\infty}^{\infty} dx\ f_n(x)\, p(y - \alpha x). \tag{21}$$

Unlike the usual expression of the Smoluchowski recursion Eq. (18), this expression is strictly valid for $0 \leqslant |\alpha| \leqslant 1$, although the case $\alpha = \pm 1$ does not admit a limiting distribution as $n \to \infty$. As we have already observed, the application of Fourier transforms, thanks to the convolution theorem, dramatically simplifies the calculation of the infinite sequence of $f_n(y)$ distribution functions.

III. CONDITIONAL PROBABILITIES AND MEAN PEAK-TO-PEAK SEQUENCE LENGTH

In the previous sections, we have introduced the linear stochastic model Eq. (1) and, using the method of characteristic functions exploiting Fourier transformations, obtained the Smoluchowski relation Eq. (21). The latter expression could have been derived using Bayesian conditional probability

arguments, but that would have left us with the problem of calculating the $f_n(y)$ as n approached infinity. We have already observed the utility of our Fourier-based methodology and exploited it in obtaining both the Smoluchowski relation and the evolution in time n as well as limiting values of the probability distribution.

Now, in order to evaluate the likelihood that the nth event is a peak, we need to calculate the probability that $x_{n-1} < x_n > x_{n+1}$. Given the Langevin equation (1), it follows, given that we know x_{n-1}, that we must consider the admissible range of the random variables η_{n-1} and η_n, since we must have

$$x_n = \alpha\, x_{n-1} + \eta_{n-1} > x_{n-1} \quad \text{and} \quad x_{n+1} = \alpha\, x_n + \eta_n < x_n. \tag{22}$$

(We are ignoring the possibility of equal x_n values inasmuch as they would normally constitute a set of measure zero.) For notational convenience, we shall eliminate the role of the subscripts and employ the random variables x, y, and z in place of x_{n-1}, x_n, and x_{n+1}. In this way, we are appropriately introducing the role of conditional variables into this problem, and this can now be directly introduced into our Smoluchowski relation (21).

In particular, we will restrict our range of integration to assure that $x_{n+1} < x_n$ (i.e., $z < y$) after first evaluating the probability density for $x_n > x_{n-1}$ (i.e., $y > x$). In this instance, we replace n by $n-1$ in Eq. (21) and limit the range of integration for x to $(-\infty, y)$. We will call this new conditional probability $f_n^c(y)$, which is given by

$$f_n^c(y) = \int_{-\infty}^{y} dx\; f_{n-1}(x)\, p(y - \alpha x). \tag{23}$$

For the Gaussian case described earlier in Eq. (15), we then obtain the closed-form result

$$\begin{aligned} f_g^c(y) &= \int_{-\infty}^{y} dx\; f_g(y)\, g(y - \alpha x) \\ &= \sqrt{\frac{1-\alpha^2}{8\pi}}\, \exp\left[-(1-\alpha^2)\frac{y^2}{2}\right] \mathrm{erfc}\left[-(1-\alpha)\frac{y}{\sqrt{2}}\right]. \end{aligned} \tag{24}$$

To complete this problem, we must now require that η_n be such that $z < y$ as in Eq. (22); hence, we require that $\eta_n < (1-\alpha)x_n$ or $\eta < (1-\alpha)y$, where we have introduced y and, for convenience, omitted the subscript from η. We can now calculate the probability $\mathcal{P}$ that the intermediate event y is a peak. In particular,

$$\mathcal{P} = \int_{-\infty}^{\infty} dy\; f_n^c(y) \int_{-\infty}^{(1-\alpha)y} d\eta\; p(\eta). \tag{25}$$

While this result does not appear to have an immediate simplification, the Gaussian case remarkably does simplify and the associated probability of there being a peak $\mathcal{P}_g$ can be expressed after some algebra as a single integral,

$$\mathcal{P}_g = \frac{1}{4\sqrt{\pi}} \int_{-\infty}^{\infty} d\zeta\; \exp(-\zeta^2)\, \mathrm{erfc}^2\left(\sqrt{\frac{1-\alpha}{1+\alpha}}\,\zeta\right). \tag{26}$$

For the case $\alpha = 0$, we can employ the derivative relationship between the exponential that appears in the former expression and the (complementary) error function, and we obtain $1/3$, as expected. For the case $\alpha = 1$, the error function term becomes 1 and, following some trivial algebra, we obtain $1/4$ as expected. Similarly, for $\alpha = -1$, the error function term becomes 0 for $\zeta > 0$ and becomes 2 for $\zeta < 0$, and we obtain 1, implying a mean peak-to-peak length of 2. Intuitively, this latter result follows from Eq. (1), since we expect that $E(|x_n|) \ll 1$ for large n, which converts a peak into a valley and a valley into a peak presenting an average cycle length of 2. Finally, by taking its derivative with respect to α, we observe that the expression for $\mathcal{P}_g$ is strictly monotone decreasing in the interval $[0,1]$. Our integral Eq. (26) seemingly lacks a simple analytic expression [16].

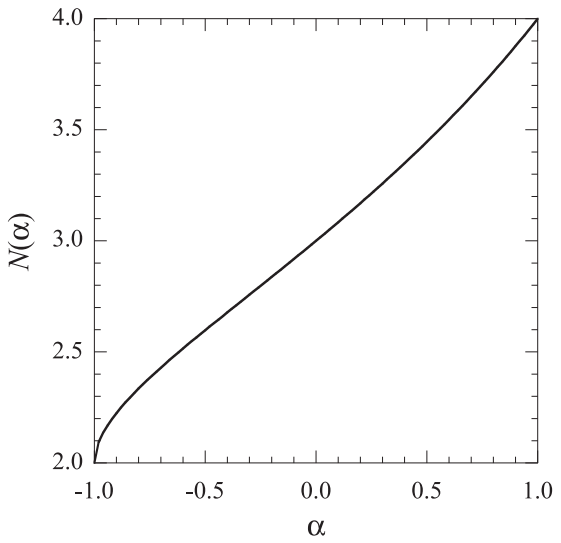

FIG. 1. Expected mean number of points per cycle $N(\alpha)$.

To illustrate this behavior, we have numerically evaluated the integral in Eq. (26) and plotted its reciprocal, which we call $N(\alpha)$ against α. In addition, we have compared our results with the Monte Carlo simulations we reported in [2], and we observed, to the requisite numerical accuracy, that they are the same.

Returning to our [2] formal treatment of Yellowstone's Old Faithful Geyser [17–20], with data extracted from [21], the observed mean peak-to-peak cycle length of 2.6685 implies using Fig. 1 that $\alpha \approx -0.42$, while from Eq. (2) for the time series $\alpha \approx -0.313$. While qualitatively similar, the two estimates differ quantitatively. By examining the empirical distribution function for the raw data, it is observed to be "long-tailed" and is far from a normal distribution; hence, the departure between the two estimates can be appreciated. The mechanism underlying the geyser phenomenon is not understood in detail, although some progress has been made [18–20]. It involves the filling and subsequent evaluation of a subterranean reservoir. If, following an eruption, the reservoir does not fill completely before its next eruption, i.e., the time to eruption is less than average, we expect a longer than average time for the reservoir to fill prior to the following eruption, and vice versa. Hence, we anticipate an alternation between shorter and longer time-to-eruption intervals, which is manifestly antipersistent.

To better visualize the outcome of the Langevin and first-order autoregressive equation (1), we present in Fig. 2 a

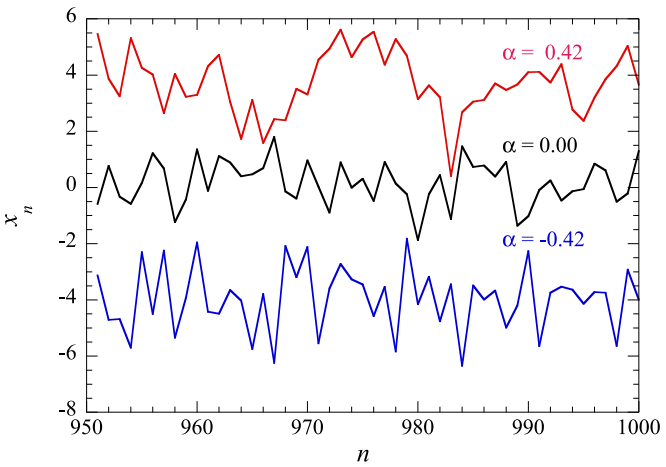

FIG. 2. (Color online) Realization of the discrete Langevin equation or first-order autoregressive process (1) for $\alpha = 0.42, 0.00$, and -0.42, from top to bottom.

realization of the process for three values of α. In each instance, we allowed the system to evolve from $x_0 = 0$ over an extended period of time, i.e., for $n = 1, \ldots, 1000$, so that initial conditions would be infinitesimal and employed a normally distributed pseudorandom generator with unit variance. We show in each case only the last 50 data points. (For clarity of presentation, the means of each data set have been shifted by four units. As a consequence of the fluctuation-dissipation theorem, the rms amplitudes of the nonzero α realizations have increased 10%.) The middle of these, shown in black, corresponds to $\alpha = 0$ and an i.i.d. process. This is the situation described by Kac mentioned earlier, and we observe that the peak-to-peak cycle length is approximately 3. The lower one of these, shown in blue, corresponds to $\alpha = -0.42$, the parameter that best fits the apparent cycle length in the Old Faithful Geyser data displayed by NTM [2] as well as in the preceding paragraph. Antipersistence is evident. The upper realization, shown in red, corresponds to $\alpha = 0.42$ and a peak-to-peak cycle of approximately 3.37. We will argue shortly that this could be significant in some biological problems. The cycle length here is more or less typical of many animal populations, although the supporting data are not especially strong. With this perspective, we now proceed to consider the implications of these results.

IV. DISCUSSION

We have sought to expand the notion of patterns emergent from randomness, which we call stochastic structure in random events (SSRE), by exploring the mean number of events in a cluster or cycle defined by the number of events separating successive peaks. We have defined here a general probabilistic theory, as well as explicitly evaluating the outcome that can be expected for an underlying Gaussian distribution. We have compared our closed-form theoretical results with the Monte Carlo simulations shown in Fig. 7 of [2], and we observe that they agree to the expected degree of accuracy presented by the simulations. While the cases wherein $\alpha = 0$ or 1 are *independent* of the underlying random step distribution function $p(\eta)$, the intervening range in α is not completely generic. We have introduced the more general first-order autoregressive model, allowing for the range of α to be extended to include negative values $-1 \leqslant \alpha \leqslant 0$ in order to describe antipersistent phenomena, such as Old Faithful Geyser eruptions. In that context, we have also introduced two methods for empirically estimating α from observational data, which can then be applied to processes with memory as well, i.e., $0 < \alpha < 1$. We will focus on this situation in the remainder of this discussion and its application to biological phenomena, such as those described at the beginning of this paper.

Our model Eq. (1) could be relevant to a wide array of problems, e.g., the seeming population cycles of many mammal species living in radically different environments, as implied by Cole and Kac [1]. Our model serves, at the very least, as a null hypothesis in a statistical sense for this phenomenon. Could there be a biologically compelling underpinning for this? Imperial College, London, and its NERC Center for Population Biology [22] has amassed a very comprehensive set of population biology related data. However, these data sets tend to be relatively short, particularly in comparison with the earthquake and magnetic substorm data presented in [2]. Biological environments and the processes prevailing there are incredibly complex. One could draw an analogy with a hydrodynamic system, which itself effectively has an infinite number of degrees of freedom, yet is simpler—due to its intrinsic homogeneity—than the biological case. Nevertheless, a single dimensionless parameter, such as a Reynolds number or a Taylor number, in hydrodynamic problems can often provide important insights. Could this also be true in the biological realm? How can Eq. (1), therefore, be interpreted.

Building on our crude analogy with complex hydrodynamic processes, let us assume that the η_n represents the number of surviving births emerging from the nth generation. (Strictly speaking, both η_n and x_n in this application will represent departures from their respective means.) While the mathematical biology literature is filled with extensive discussions of this issue, represented in many instances via solutions to time-delayed integral-differential equations, we will assume that a "random" number of births is a not unrealistic metaphor for the process. The variable x_n, naturally, describes the number of individual members of a given species included in the annual census in year n. The parameter α, therefore, can be regarded as the fraction of individuals in the nth generation who survive and are present in the census performed in year $n + 1$. (This conceptual framework excludes unexpected external events such as environmental catastrophes, etc. In Fig. 2, the upper or red realization could possibly apply to biological situations in which, given the choice of α, approximately 42% of the individuals survive from one generation to the next.) To what extent, then, can the results obtained for an underlying Gaussian process be an indicator of the behavior of our model in such circumstances? Three factors argue for the behavior being similar, albeit not precisely universal.

(i) The end points $\alpha = 0$ and 1 yield the mean number of events in a peak-to-peak sequence $N(\alpha)$ as being 3 and 4, respectively, as shown in [2] in detail as well as here in a simple manner for *all* distribution functions.

(ii) The Lindeberg-Feller theorem [23], a generalization of the central limit theorem [6], assures under relatively general conditions that f_n will tend to a Gaussian. Hence, the integral

(25) is not expected to depart substantially from Eq. (26). Monte Carlo simulations, such as those performed in [2], can be employed as a check.

(iii) The combined effect of the former items strongly suggests that the relationship for $N(\alpha)$ will remain quantitatively close to that depicted in Fig. 1.

By building upon these ideas, our application of the Langevin model (1), as well as the autoregressive model, could provide important insights into a wide arena of problems in the physical, life, and social sciences that manifest stochastic structure in random events.

ACKNOWLEDGMENTS

I wish to thank Donald Turcotte, James Sneyd, Yervant Terzian, Robert May, Simon Levin, and, especially, Joseph Rudnick and Richard Lovelace for useful discussions and suggestions.

[1] L. Cole, J. Wildlife Manage. **15**, 233 (1951).
[2] W. I. Newman, D. L. Turcotte, and B. D. Malamud, Phys. Rev. E **86**, 026103 (2012).
[3] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[4] C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958).
[5] D. R. Cox, *Renewal Theory*, Science Paperbacks, SP58 (Methuen, London, 1967).
[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. (Wiley, New York, 1971), Vol. II.
[7] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes* (Academic, New York, 1981).
[8] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. (Wiley, Hoboken, NJ, 2008).
[9] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer-Verlag, New York, 1991).
[10] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. (Chapman & Hall/CRC, Boca Raton, FL, 2004).
[11] A. Einstein, R. Fürth, and A. D. Cowper, *Theory of the Brownian Movement* (Methuen & Co., London, 1926).
[12] M. v. Smoluchowski, Annalen der Physik **48**, 1108 (1915).
[13] F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill Series in Fundamentals of Physics (McGraw-Hill, New York, 1965).
[14] R. M. Dudley, *Real Analysis and Probability* (Cambridge University Press, Cambridge, 2002).
[15] P. A. P. Moran, *An Introduction to Probability Theory* (Clarendon, Oxford, 1984).
[16] I. S. Gradshtein, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1965).
[17] A. Azzalini and A. Bowman, Appl. Statist. **39**, 357 (1990).
[18] J. Dowden, P. Kapadia, G. Brown, and H. Rymer, J. Geophys. Res. **96**, 18059 (1991).
[19] R. Hutchinson, J. Westpphal, and S. Kieffer, Geology **25**, 875 (1997).
[20] J. Rinehart, J. Geophys. Res. **74**, 566 (1969).
[21] http://www.geyserstudy.org
[22] http://www3.imperial.ac.uk/cpb
[23] J. Jacod and A. N. Shiriaev, *Limit Theorems for Stochastic Processes*, 2nd ed. (Springer, Berlin, 2003), Vol. 288.