

Diffusion Reproduction Processes

Yi-Cheng Zhang,¹ Maurizio Serva,¹ and Mikhail Polikarpov²

Received May 12, 1989; revision received September 27, 1989

Interesting physics emerges from studying a population of reproducing individuals. Each can be regarded as a random walker, but it can either duplicate or die. Novel features of the collective behavior are quite surprising: if individuals reproduce or die freely, the life expectation is proportional to the size of the population, and if it is kept constant, the center of mass moves in space as if it were a single walker (i.e., the diffusion constant is independent of population's size). Biology-inspired interactions are also considered.

KEY WORDS: Population dynamics; fluctuation; localization.

1. THE MODEL AND SURVIVAL OF A POPULATION

We consider a large population of N individuals. Each of them can undergo a Brownian motion in d -dimensional space. However, each individual can duplicate, or disappear, with equal probability. The aim in this work is to illustrate novel features of this model, in the simplest possible way. All input constants, coefficients, etc., are assumed to be unity, unless otherwise specified. The time variable t assumes integer values which we call *generations*. After each generation, each individual in the population faces two possibilities, independently and with equal probability: either it becomes two individuals (duplication) conserving whatever identity the individual at the generation before may have had, or it simply disappears (death). The duplicated new individuals are identified with their parent individual—we say it has survived.

Let us ignore their positions and motion in space for the moment.

¹ INFN, Dipartimento di Fisica, Università di Roma, 00185 Rome, Italy.

² Permanent address: Institute of Theoretical and Experimental Physics, Moscow, USSR.

First we concentrate on the question of the survival of such a randomly reproducing population. Denote by N the size of the initial population and by $N(t)$ the size of the population at time t , $N(0) = N$. We have the following evolution relation:

$$N(t+1) = N(t) + \sum_{i=1}^{N(t)} \gamma_i(t) \quad (1)$$

where the the death–reproduction rate of the i th individual $\gamma_i(t)$ assumes the values -1 and 1 with equal probability, the labeling is arbitrary. Note that $\langle N(t) \rangle = N$, where the average $\langle \cdot \rangle$ is over the reproductive noise $\gamma_i(t)$, which is defined as $\langle \gamma_i(t) \rangle = 0$, $\langle \gamma_i(t) \gamma_j(t') \rangle \sim \delta_{i,j} \delta_{t,t'}$ and the Kronecker δ -symbols are used. From the above relations we obtain the second moment, by repeated iterations,

$$\langle [N(t)]^2 \rangle = \langle [N(t-1)]^2 \rangle + N = \dots = N^2 + tN$$

from which we obtain the fluctuation of the population's size,

$$[\Delta N(t)]^2 = \langle [N(t) - \langle N(t) \rangle]^2 \rangle = Nt$$

The population's size $N(t)$ is expected typically to lie in the range $N - \Delta N(t)$ to $N + \Delta(t)$, where

$$\Delta N(t) = (Nt)^{1/2} \quad (2)$$

This shows that the amplitude of the population fluctuation $\Delta N(t)$ is an increasing function of t . There exists a *typical* lifetime $\tau \sim N$, which is defined as the time at which fluctuations $[\Delta N(\tau) \sim N]$ become as large as mean values themselves, and beyond which the population is extinct with finite probability (close to one). (This *typical* lifetime is different from the *mean* lifetime; the latter is infinite—see Appendix for more discussion.) This is very similar to the “first-passage” problem studied in the context of random walks,⁽¹⁾ where the typical situation is when a walker goes over a “cliff,” passing a point of no return.

There is typically a gain or a loss of \sqrt{N} individuals in the population. As a result, a freely reproducing population (anarchy) is doomed, no matter how large its initial size—a single large negative fluctuation can eliminate it.³ There is a “milder” way of introducing reproductive noise, though, but we shall see that it does not change the above qualitative conclusion by much. We may choose to impose a constraint on the random death–reproduction rates in Eq. (1). We require that the global death–reproduction noise $\sigma(t) = \sum_{i=1}^N \gamma_i(t)$, rather than $\gamma_i(t)$ itself, should

³ Extinction due to statistical fluctuations is known in population dynamics. See, e.g., ref. 2.

be independently distributed so that $\langle \sigma(t) \sigma(t') \rangle \sim \delta_{t,t'}$. This implies that the individual death and reproduction rates inside the population should greatly compensate each other, so that in each generation the population only increases or decreases by one individual. Following a similar analysis, we obtain $\Delta N(t) \sim \sqrt{t}$ instead of $(Nt)^{1/2}$, and the typical lifetime becomes $\tau = N^2$.

2. CONSTANT POPULATIONS AND INTERNAL FLUCTUATIONS

For many realistic biological systems, global populations are expected to vary in time at much slower rates, at least for those that have survived long enough to be of most interest to us. For instance, through food and other resource availability or ecological factors, populations may be maintained constant or slowly changing. Our concern here is, however, the relative fluctuations inside a population. In the following we shall only consider constant populations ($= N$). The arguments and the results in this work can be easily extended to nonconstant populations.

Global population conservation implies that at each generation exactly half of the individuals (randomly chosen) duplicate, the other half die. Large reproductive fluctuations inside a population, nevertheless, are still characterized by the time scale $\tau \sim N$. This is because the reproductive noise is still capable of reducing a large portion, comparable to the global population, to a few individuals, or magnifying a small portion of a few individuals to the size of the global population, in a typical time interval τ . Part of the population, eventually only one individual, proliferates at the expense of all the other individuals. After a time lapse τ , all the individuals of the population can be regarded as offspring of a lucky individual at the beginning. We conclude that in any time interval τ , the content of the population is substituted almost completely (except one survivor).

3. THE INDIVIDUALS AS RANDOM WALKERS IN SPACE

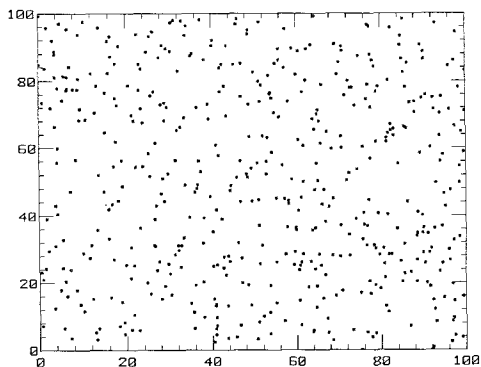
Let us now allow these individuals to move in a d -dimensional continuum space. Each of them is subject to an independent positional noise, and thus undergoes a Brownian motion. For the moment we do not consider interactions among the individuals. The two members of a duplicated pair of individua are put in the place of their parent individual, inheriting the positional information (and whatever other identities). Each member of the pair, however, will act independently of the other after birth. If we simulate this population on a computer (or in one's mind) we find that, starting from any initial distribution, after a long time ($t > N$), the popula-

tion will reach a dynamical equilibrium state. In this state the population will occupy a finite region and its physical extension will not disperse further in space, as if there were a force holding it together. A computer rendering is illustrated in Fig. 1. Some remarks are useful in understanding the figures, which are three sequential snapshots at 0th, 50th, and 400th generations, respectively. We simulate a population of 500 individuals in a two-dimensional continuum space. We take an initial configuration in which the individuals are randomly distributed (with uniform probability) in a square region, as shown in Fig. 1a. The individuals do not interact with each other and they independently duplicate or die, but are subject to the constraint of global population conservation. Each individual, in a unit time interval (generation), will make a move in a random direction, of random step length (Gaussian distributed with variance one); then it will either duplicate or disappear.

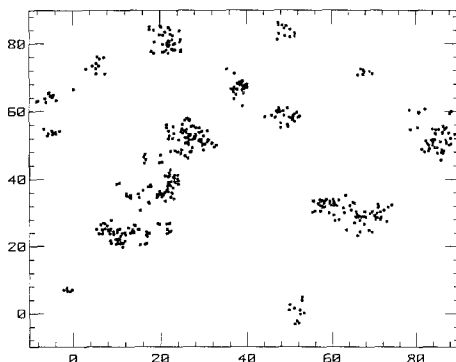
We let the population evolve after some generations say, 50, as shown in Fig. 1b. We see clusters, or islands, forming. There are large empty regions between the islands. Each island most likely consists of offspring of a single (intermediate) ancestor, i.e., it is a family with the same “surname.” In principle, inter-island travel is possible (some individuals leave an island to join another). The odds for a small split-up group to survive are discouraging, as is the perspective of not-too-large islands. Analogues can be found in gambling theories: Given that the winning-losing odds are fair, a gambler with a small budget is not likely to play long, though, in principle, he or she can become a millionaire. Such analogues can also be found in social, economical competitions. Thanks to global population conservation, someone has to survive. The largest island near the lower-left corner in Fig. 1b (it may consist of three adjacent islands) is the most hopeful candidate.

Now the population has evolved through 400 generations, as shown in Fig. 1c. As suspected, the largest island in Fig. 1b has indeed imprinted its dominance on the future of the population (we can infer this from its position). The concentration of the population is more pronounced. Note that the center of mass has drifted a considerable distance (~ 20 , on the scale of the figures), comparable to that of a single random walker. Evolving further, the population will reach a dynamical equilibrium state, in which the population is concentrated in a large, single island of linear size \sqrt{N} . There will be constant, spontaneous, short-lived breakoffs from the central island.

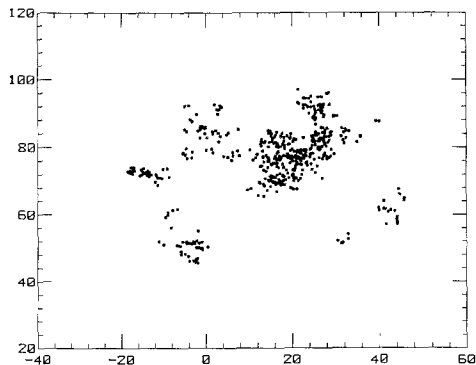
The reason a population cannot disperse indefinitely is that the existence of the individuals cannot be so preserved. The time available for any two individuals (given that both survive long enough) to diffuse apart is bounded by $\tau \sim N$. We can regard this pair as two random walkers released



(a)



(b)



(c)

Fig. 1. The population distribution at different generations. The 2D space is continuous and the scales are arbitrary. (a) The population starts with a uniform distribution; (b) the evolution after 50 generations; (c) after 400 generations. From these snapshots one may infer the future of the population; see the text for the interpretations.

from each other. After a time span τ , the distance apart is proportional to \sqrt{N} . Since this is the upper bound on the interindividual distance, the linear dimension of the population, after sufficiently long time $t > N$, is

$$R \sim \sqrt{N} \quad (3)$$

Thus, we conclude that the population is always confined in a region of linear size \sqrt{N} , independent of the dimensionality d of space.

4. THE MOBILITY OF THE CENTER OF MASS

The population as a whole or, more specifically, its center of mass can be regarded as a free random walker in space as well. Suppose that the mean displacement *per* generation (mean displacement, or MD for short) for a single individual is one; what is the corresponding value for the center of mass? If the individuals were truly independent random walkers, the mean displacement of the center of mass would be

$$x_{\text{MD}} \sim 1/\sqrt{N} \quad (4)$$

as a consequence of the central limit theorem. However, we know that there is an effective correlation—shared ancestry.

How is x_{MD} influenced by this correlation? Denote by $\mathbf{x}_c(t)$ the position of the center of mass of a population at a given time t ; we have

$$x_{\text{MD}}^2 = \langle [\mathbf{x}_c(t+1) - \mathbf{x}_c(t)]^2 \rangle \quad (5)$$

where $\langle \cdot \rangle$ denotes the average over positional noise. By definition,

$$\mathbf{x}_c(t+1) - \mathbf{x}_c(t) = \frac{1}{N} \left[\sum_{i=1}^N \mathbf{x}_i(t+1) - \sum_{j=1}^N \mathbf{x}_j(t) \right] \quad (6)$$

where $\mathbf{x}_i(t)$ denotes the i th individual's position. We recall that between two successive generations t and $t+1$, half of the N individuals of generation t disappear; each of the other hand becomes two. Take the approximation that no random walk has occurred in the interval t to $t+1$; the duplicated pairs would then stay in their parents' positions. In this approximation we have $\sum_{i=1}^N \mathbf{x}_i(t+1) = 2 \sum_{i=1}^{N/2} \mathbf{x}_i(t)$, this implies that $N/2$ terms in the above relation (6) cancel out. The RHS of Eq. (6) becomes

$$\frac{1}{N} \sum_{i=1}^{N/2} [\mathbf{x}_i(t) - \mathbf{x}_{i+N/2}(t)] \quad (7)$$

where we have used the fact that the ordering index i is arbitrary.

Let us evaluate x_{MD} only for the one-dimensional case; for $d > 1$ one can simply repeat the analysis for each of the d projections. Each difference in the sum in (7) is bounded by the length R in Eq. (3), with random \pm signs. The sum in (7) yields $\pm\sqrt{N}R$; thus, the above expression should be proportional to $\pm R/\sqrt{N}$. Using Eq. (3), we obtain the estimate

$$x_{MD} \sim R/\sqrt{N} \sim O(1) \quad (8)$$

It is easy to check that this result is also true for arbitrary dimensionality, as long as $N \gg d$. We have also performed numerical simulations for $d = 1, 2$, and 3, which confirm that x_{MD} is a constant for all N .

This result (8) implies that, say, a population of 100 and a population of one billion move equally fast; hence, the diffusion constant ($\sim\sqrt{x_{MD}}$) for the center of mass is independent of N . It is instructive to see how such coherence is achieved. From the previous discussion we learn that, after a time $\tau > N$, only one individual survives and the rest of the population can be regarded as its offspring. It is not surprising that the trace of the center of mass and the trace of the survivor can be identified in the *long* run. The above conclusion about x_{MD} is, however, a stronger statement. It asserts that even between two successive generations, the center of mass moves a distance of order one. Intuitively, this is because the linear size of a population increases with N ($R \sim\sqrt{N}$), but, since in each generation $N/2$ individuals die, there is a large rearrangement of the center of mass which compensates exactly the statistical factor $1/\sqrt{N}$.

5. ASSIGN SURNAMES TO THE INDIVIDUALS

So far we have been dealing with identical individuals, distinguished only by their different positions. We can assign distinct attributes, say surnames, to the individuals. As the population evolves, the number of the distinct surnames or the number of the islands in the figures decreases. Let us estimate the rate of this decrease. Suppose that all the N individuals are distinct at the beginning. At some given time there will be $M = N/S$ distinctive surnames left; S is the mean size of the subpopulation whose members share the same surname. From the above discussion we know that a subpopulation of size S can survive for a time span $\sim S$. This implies that during a small time interval Δt , each of the M subpopulations (of size S) has a probability proportional to $\Delta t/S$ to disappear. In other words, there will be $(M \Delta t/S)$ surnames of such size who become extinct. Using a continuum approximation, we have (the minus sign accounts for decrease and $S = N/M$)

$$dM/dt \sim -M/S \quad (9)$$

Solving this differential equation, we obtain

$$M \sim N/t \quad (10)$$

where $t \leq N$ and for $t > N$, $M = 1$.

Such a population is destined to monotony and it is rather dull. One can imagine that each individual has a small probability c ($\ll 1$) of changing its surname spontaneously (mutation). In a unit time interval there will be cN new surnames in the population. We ignore the loss of surnames due to mutations which are sub-leading-order contributions. We can include this effect in Eq. (9) by adding cN to the RHS. We expect in the large-time limit that the number of distinct surnames reaches a plateau, hence $dM/dt = 0$. Balancing the decrease due to reproductive fluctuations and the gain due to spontaneous mutations, we have $-M/S + cN = 0$. The number of distinct surnames is no longer destined to be reduced to one. Rather, in a *stationary* state, from the above relation, there are $M = \sqrt{c} N$ distinct species (surnames) of the mean size $S = 1/\sqrt{c} \gg 1$.

The above result also has implications for the previously discussed identical random walkers. Suppose that each random walker undergoes a “normal” Brownian motion, i.e., the mean free path of order one, as discussed above. However, for each walker and each time step (generation), there is a small probability c of jumping (randomly oriented) a large distance $L \gg 1$, such that it is out of the region occupied by the rest of the population. If an isolated walker is to survive, it will develop into an island, as those shown in the figures. The above result implies that there will be $\sqrt{c} N$ islands, each of them of size $1/\sqrt{c}$, on the average. Contrary to the single island dominance shown in Fig. 1c, all the islands will have more or less the same size—a democratic situation.

6. THE INDIVIDUALS WITH INTERACTIONS

6.1. Attraction

We first consider the attractive interaction: Each individual has a tendency to be close to its nearest neighboring individuals. We have a hypothetical situation in which each individual, e.g., for fear of being alone (given that small groups are more “susceptible” to adverse reproductive fluctuations) or for some other reason, seeks to be linked to the others. We assume for simplicity that the attractive force in position space between two neighboring individuals is proportional to their distance, similar to the

elastic force of a harmonic oscillator. In one dimension ($d = 1$) we have the following stochastic equation for the i th individual:

$$\frac{dx_i}{dt} = \frac{1}{2}(x_{i+1} + x_{i-1} - 2x_i) + \eta_i(t) \tag{11}$$

where $\eta_i(t)$ is the positional noise which is uncorrelated, i.e., $\langle \eta_i(t) \eta_j(t') \rangle \sim \delta_{i,j} \delta_{t,t'}$, and the boundaries ($i = 1, N$) are free. Here the ordering is only a bookkeeping notation, to give instantaneous labels to the individuals. At any time t , sequential ordering $i = 1, 2, \dots, N$ denotes the actual consecutive positions of the individuals, i.e., $x_1 < x_2 < \dots < x_N$, regardless of what their previous labels were. If in the successive generations two individuals should be exchanged in position space, so should the ordering. In a higher dimensional space, equations similar to (11) can also be given, but there are different ways in doing so. However, as long as the interactions are local, from the discussion below we will see that the conclusion should be the same.

It is easy to estimate the linear size of the population and other related physical quantities. For the population to reach dynamic equilibrium (stationary state), the deterministic attractive force that keeps the population together and the stochastic force that disperses it should balance. For a single individual the former is proportional to the mean interindividual distance and the latter is taken to be of order one. Denote by R_a the linear size of the population. The population density is N/R_a^d ; thus, the attractive force is proportional to the mean interindividual distance $R_a/N^{1/d}$. It should be equated to the stochastic force of $O(1)$, i.e., $R_a/N^{1/d} \sim 1$. We then obtain the estimate

$$R_a \sim N^{1/d} \tag{12}$$

We have treated the population as if there were no death and reproduction. They enter into play only though the characteristic time $\tau \sim N$. From the previous discussion we know that, after the characteristic time τ , the population contains only (except one) new individuals. In the absence of interactions, their inherent correlation yields $R \sim \sqrt{N}$ for all dimensions. Thus, the correct R_a is given by

$$R_a = N^\alpha, \quad \alpha = \begin{cases} 1/2, & d \leq d_c \\ 1/d, & d > d_c \end{cases} \tag{13}$$

where $d_c = 2$. For example, when $d = 1$, a completely relaxed system would reach a size proportional to N , by (12). However, due to the limited lifetime available to the majority of the population, this relaxed size can

never be reached. As is standard in physics, d is not limited to integer values; extrapolation to noninteger d is understood in (13) and in other results below.

6.2. Repulsion

We may also consider the opposite situation, i.e., when the individuals want to stay away from each other. This effect can be modeled by a local repulsive interaction. The reasons for repulsion may be that, for instance, in overcrowded regions, the reproduction rate is reduced or food shortage causes the higher death rate; or the individuals born in crowded regions tend to immigrate. The size of the population is still taken to be constant. Here also we can estimate the linear size of the population, as in the above discussion of the attraction case. Denote by R_r the linear size of the “repulsive” population; its expansion rate should be proportional to the “force” of the local repulsion. That is,

$$dR_r/dt \sim -\partial\rho/\partial R_r \quad (14)$$

where $\rho = N/R_r^d$ is the mean local population density and the repulsive potential is taken to be proportional to the density.

The above relation is very similar to that proposed by Pietronero⁽³⁾ for the “true” self-avoiding walk, which is based on the same type of kinetic local repulsion. Integrating Eq. (14) leads to $R_r^{d+2} \sim Nt$. In the stationary state, t should be replaced by $\tau \sim N$, the characteristic time. We obtain

$$R_r \sim N^{2/(d+2)} \quad (15)$$

Note that we encounter again the critical dimension $d_c = 2$. When $d > d_c$ we know that $R_r \sim N^{1/2}$, in the absence of repulsion, which implies that the local repulsive interaction now is not relevant. The correct relation is

$$R_r \sim N^\beta, \quad \beta = \begin{cases} 2/(d+2), & d \leq d_c \\ 1/2, & d > d_c \end{cases} \quad (16)$$

Recall that the mean displacement of a population is given by (8): $x_{\text{MD}} \sim R/\sqrt{N}$. Replacing R in this equation by (13) and (16), we can obtain the mean displacement x_{MD} , hence the diffusion constant, for a population with attractive or repulsive interaction, respectively.

6.3. Attraction and Repulsion Combined

We may have to face the situation in which *both* attraction and repulsion are present. Naively, one may expect that either they exactly cancel

each other or one of them is dominant such that only one manifests itself. However, the two interactions we have considered above have a different dependence on space dimensionality. Therefore, neither interaction can dominate for all d . In fact, when we consider simultaneously attraction and repulsion, which amounts to combining Eqs. (13) and (16), we obtain

$$R_{ar} \sim N^\zeta, \quad \zeta = \begin{cases} 2/(d+2), & d \leq d_c \\ 1/d, & d \geq d_c \end{cases} \quad (17)$$

where $d_c = 2$ as before. This shows that, below d_c , the repulsive interaction is dominant, whereas the attractive interaction is negligible; above d_c , the situation is reversed.

7. CONCLUSIONS

Salient features of our model that differ from traditional diffusion are due to reproductive fluctuations: some individuals may happen to reproduce more at the expense of some others. There is a characteristic time $\tau \sim N$, central to our study, during which the population can be regarded as offspring of a single lucky individual. As a result, the population cannot diffuse apart indefinitely. Rather, it is confined to a region of linear extension \sqrt{N} . Large internal reproductive noise results in the large mobility of the center of mass of the population. Simple *local* interactions are introduced to mimic a biological system. Stationary states in the presence of interactions are of particular interest; here the characteristic time always plays a crucial role.

APPENDIX

Here we derive the probability distribution function of the lifetime of a population. Consider a population whose initial size is N and apply the dynamic rules of our model (1). Denote by $P_t(n)$ the probability that the population reaches the size n at time t . Note that our rules in Eq. (1) produce *only* even integer values (including 0) of possible population size; hence in the following we consider only even population sizes (including the initial value N). We have the following master evolution equation:

$$P_{t+1}(m) = \sum_{n=0,2,\dots}^{\infty} C_{n,m/2} P_t(n)/2^n \quad (18)$$

where $C_{n,m/2} = \binom{n}{m/2}$, when $n \geq m/2$, $C_{n,m/2} = 0$ otherwise, and the initial condition $P_0(n) = \delta_{n,N}$. We are interested in the probability of extinction at time t , $P_t(0)$, which satisfies $P_t(0) \geq P_{t-1}(0) \geq \dots \geq P_0(0)$, where the

equality holds after extinction. This is because in Eq. (18) the probability of extinction $P_t(0)$ does not contribute any more to $P_{t+1}(m \neq 0)$; in other words, there is no recurrence probability. This is to be contrasted with the “first passage” problem in random walks.

Let us define the generating function at time t

$$G_t(z) \equiv \sum_{m=0,2,\dots}^{\infty} z^m P_t(m) \quad (19)$$

From the master equation we obtain the following recursive relation:

$$G_{t+1}(z) = G_t\left(\frac{1+z^2}{2}\right) \quad (20)$$

The initial condition becomes $G_0(z) = z^N$. From the above relations we can obtain the distribution function $F(t) \equiv P_t(0) = G_t(0)$, which is the probability that the population at time t becomes extinct. Using Eq. (20) and the initial condition $G_0(z) = z^N$, we can obtain $G_t(0)$ iteratively. Denote by $G_t(0) \equiv u_t^N$; we have

$$u_{t+1} = \frac{1+u_t^2}{2}, \quad u_0 = 0 \quad (21)$$

Using the continuum-time approximation, we can readily deduce the asymptotic behavior ($t \gg 1$) of u_t , $u_t \simeq (1 - 2/t)$. We finally obtain the distribution function

$$F(t) \simeq (1 - 2/t)^N \quad (22)$$

Note that for any given initial population size N ($\gg 1$), for long enough time t the population is destined to become extinct, confirming the intuitive conclusion in the text. For both asymptotically large N and t , $F(t)$ becomes appreciable only when t is proportional to N or larger, when approximately $F(t) \simeq \exp -2N/t$. This allows us to define the *typical* lifetime $\tau = N$ that with finite probability the population will become extinct. As a scaling function of large N we can say that $F(\tau)$ can be arbitrarily close to one.

However, the “tail” of $F(t)$ can be very long for large t . In fact, the distribution density is $f(t) = dF/dt \sim 1/t^2 \exp -2N/t$, which implies that the *mean* lifetime $\langle t \rangle$ is divergent (though only logarithmically, weaker than that of the “first passage” problem below).

A similar problem is the *typical* “first passage” time in the context of the classical random walk. If an unbiased random walker starts at N , the

“first passage” time t through 0 has distribution density⁽⁵⁾ $f(t) \sim 1/t^{3/2} \exp(-N^2/2t)$. In ref. 5 full discussion is given of the apparent paradox: while the random walker has probability one to cross the origin, the *mean* “first passage” time $\langle t \rangle$ is infinite. For N large we can nevertheless define a *typical* first passage time $\tau \sim N^2$, similar to the above discussion on the population dynamics.

ACKNOWLEDGMENTS

We thank the anonymous referee whose comments prompted us to furnish the above appendix. We also thank D. J. Amit for a critical reading of the work.

REFERENCES

1. M. E. Fisher, *J. Stat. Phys.* **34**:667 (1984), and references therein.
2. E. G. Leigh, Jr., *J. Theor. Biol.* **90**:213–239 (1981), and references therein.
3. L. Pietronero, *Phys. Rev. B* **38**:5647–5649 (1983).
4. C. Amitrano, L. Peliti, and M. Saber, *C. R. Acad. Sci. Paris* **307**:803 (1988).
5. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1966).