Journal of Statistical Physics, Vol. 30, No. 3, 1983

Phase Transition in a Four-Dimensional Random Walk with Application to Medical Statistics

O. E. Percus[1] and J. K. Percus[1,2]

Received October 12, 1982

A random walk in a piecewise homogeneous medium can exhibit a variety of asymptotic behaviors. In particular, it may lodge strictly in one region or divide in probability among several. This will depend upon the parameters describing (a) the walk, (b) the interregion boundary, and (c) the initial location of the walk. We analyze from this point of view a special four-dimensional walk on an integer lattice with two homogeneous regions separated by a hyperplane of codimension 1. The walk represents a continuing sequence of clinical trials of two drugs of unknown success probabilities and the two regions represent the Bayes-derived criterion as to which drug to try next. The demarcation in the parameter space of success probabilities and initial coordinates between one- and two-region asymptotics is mapped out analytically in several special cases and supporting numerical evidence given in the general case.

KEY WORDS: Phase transition; random walk; integer lattice; clinical trials; piecewise homogeneous.

1. INTRODUCTION

The subject of time-invariant random walks on a spatially homogeneous lattice is a very old one. If only finitely many prior steps must be remembered, such a walk is equivalent to a spatially homogeneous Markov chain and is routinely solvable in most cases.[1] Spatially inhomogeneous random walks have recently received much attention, primarily associated with lattices representing random media, and the possibility of qualitatively

Based upon a talk given at the Symposium on Random Walks and Their Application to the Physical and Biological Sciences, NBS, June 28–July 1, 1982.

[1] Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012.

[2] Supported in part by Department of Energy, Contract No. DE-AC102-76ERO3077.

0022-4715/83/0300-0755$03.00/0

different walks occurring under change of lattice parameters noted.[2] In this paper, we shall consider the simplest possible inhomogeneity—a piecewise homogeneous lattice of just two pieces—and show that here, too, the type of walk can change discontinuously with lattice parameters—a kind of phase transition. In fact, the model we focus on will show something much closer to the traditional phase transition of thermodynamics: a single phase, followed by two mixed phases, and then a second single phase under variation of system parameters.

The structure of the paper is as follows. We first consider a problem in the statistics of clinical trials that motivated the investigation, define an "ethical" strategy for such trials, and introduce some criteria for assessing the qualitative nature of the resulting four-dimensional random walk. We then examine in greater detail the dependence of the walk on system parameters. A highly simplified one-dimensional model is set up to show the genesis of the phase transition, and its relation to the more involved four-dimensional model indicated. At this stage, the discussion becomes more quantitative. A generating function for the general problem is introduced, and it is shown how the piecewise continuity can be accounted for by an appropriate "tag"—an additional independent variable. This technique is applied to the prototype one-dimensional model, to a treatment criterion recently analyzed by Bechhofer, and finally to special cases of the "ethical" testing procedure that supplied the original motivation.

2. CLINICAL TRIALS CONTEXT

A basic problem in applied pharmacology is that of distinguishing the relative effectiveness of drugs to be used for a given ailment. This is generally done by means of a sequence of clinical trials in which the drugs are given randomly to patients and the results compared. The longer the sampling period, the more rational a decision may be made, but the more carefully the tests must be designed to reduce the number of inappropriate treatments during the testing. For this purpose, a stopping rule may be imposed to terminate the trials when sufficient information is amassed. Alternatively, and this is the case we will emphasize, the information amassed can be used to select the course of the treatments, which then never leave the "testing" period. (One thereby pays no attention to the financial costs of accumulating, processing, and conveying data.)

The highly idealized situation we consider is as follows.[3] A sequence of equivalent patients present themselves. Each one is treated by one of two drugs A or B, and the unequivocal result, success or failure, is known before the arrival of the next patient. The drugs have fixed but unknown success probabilities p_A, p_B, failure probabilities q_A, q_B. Suppose that the information available at the arrival if a patient consists of A, the number of

times drug A has been previously tried, resulting in success, the complementary number of failures a, and the corresponding quantities B, b for drug B. We then define a treatment protocol by means of a treatment function $T(A, B, a, b)$ with the instructions

$$T(A,B,a,b)\begin{cases} >0 & \text{treat with drug } A \\ <0 & \text{treat with drug } B \\ =0 & \text{flip a coin to decide on } A \text{ or } B \end{cases} \tag{2.1}$$

Putting it another way, the drug A is to be tried with a probability

$$P(A) = \theta(T(A,B,a,b)) \tag{2.2}$$

where

$$\theta(x) = \tfrac{1}{2}(1 + \operatorname{sgn} x)$$

We now wish to assess the statistics of the sequence of treatments, and its dependence upon the unknown parameters p_A, p_B. At least two qualitative figures of merit for the function $T(A, B, a, b)$ come to mind: (i) the difference between $\max(p_A, p_B)$ and the probability of success for the Nth patient when the protocol is followed, and (ii) the corresponding difference between $\max(p_A, p_B)$ and the mean success rate for the full anticipated population of N patients (the so-called patient horizon). When (i) is the primary desideratum, we speak of the protocol as being "ethical,"[4] i.e., no patient is a guinea pig, being used solely to collect information for the benefit of his successors. A key question is how well (i) also satisfies the global criterion (ii). Of course, the way this trial procedure has been set up, the testing continues forever, presumably with an increasing use of the better drug, and hopefully with its relative frequency approaching unity, in which case T has been called "asymptotically optimal."[5] We will not in this paper pay attention to the possibility of invoking a "stopping rule,"[6] in which case a decision is made at some point that, e.g., drug A is superior, and it is used thereafter.

The process we have been discussing is a random walk on the nonnegative octant of the four-dimensional integer lattice (A, B, a, b). At each step, one and only one of the coordinates is increased by unity:

$$\begin{array}{lll} \Delta A = 1 \\ \Delta a = 1 \end{array} \quad \text{with probability} \quad \begin{matrix} p_A \\ q_A \end{matrix} \quad \text{if } A \text{ is tried}$$

$$\begin{array}{lll} \Delta B = 1 \\ \Delta b = 1 \end{array} \quad \text{with probability} \quad \begin{matrix} p_B \\ q_B \end{matrix} \quad \text{if } B \text{ is tried} \tag{2.3}$$

If the probability distribution on this lattice is $P(A, B, a, b)$, normalized so that

$$\sum_{\{A,B,a,b \mid A+B+a+b=N\}} P(A,B,a,b) = 1 \tag{2.4}$$

then the development of $P(A,B,a,b)$ clearly satisfies

$$(p_A\Delta_A + q_A\Delta_a)\big[\theta(T(A,B,a,b))P(A,B,a,b)\big] + (p_B\Delta_B + q_B\Delta_B)\big[\bar{\theta}(T(A,B,a,b))P(A,B,a,b)\big] = \delta_{(A,B,a,b),(0,0,0,0)} \tag{2.5}$$

where

$$\Delta_x f(x) \equiv f(x) - f(x-1), \qquad \bar{\theta}(T) \equiv 1 - \theta(T) \tag{2.6}$$

This is a piecewise homogeneous random walk with constant coefficients in the region $T(A,B,a,b) > 0$, and a different set of constants when $T(A,B,a,b) < 0$. The boundary behavior is rather involved, and this is the nub of the analytic difficulties.

3. PHENOMENOLOGY

From the point of view of our initial motivation, a convenient quantity with which to assess the global character of the walk is the mean number of excess failures in the first N steps. If $p_A \geqslant p_B$, this is

$$C_N = \sum_{A+B+a+b=N} \chi_N P(A,B,a,b) \tag{3.1}$$

where

$$\chi_N = a + b - Nq_A \tag{3.2}$$

It is also given by $p_A - p_B$ times the mean number of times B is tried in the first N trials:

$$C_N = (p_A - p_B) \sum_{A+B+a+b<N} \bar{\theta}(T(A,B,a,b))P(A,B,a,b) \tag{3.3}$$

For more incisive information, we may also want to consider the corresponding variance

$$\sigma_N^2 = \sum_{A+B+a+b=N} \chi_N^2 P(A,B,a,b) - C_N^2 \tag{3.4}$$

Now let us specialize to an "ethical" protocol that has been examined in fair analytic and very extensive numerical detail.[4] It is that in which a Bayes estimator is constructed for $p_A - p_B$, the difference of the unknown success probabilities, based upon the results (A,B,a,b) of the previous trials. For p_A, one has the estimator

$$\rho_A = \frac{\int\int_0^1 p_A^{A+1} p_B^B q_A^a q_B^b f_0(p_A, p_B)\, dp_A\, dp_B}{\int\int_0^1 p_A^A p_B^B q_A^a q_B^b f_0(p_A, p_B)\, dp_A\, dp_B} \tag{3.5}$$

and similarly for p_B. Matters simplify materially if the a priori joint distribution $f_0(p_A, p_B)$ is taken as a product of independent identical β-distributions controlled by parameters e and f:

$$f_0(p_A, p_B) = \left(\frac{(e+f+1)!}{e!\, f!} \right)^2 p_A^e q_A^f p_B^e q_B^f \tag{3.6}$$

in which case one has the estimator

$$\rho_A - \rho_B = \frac{A+e+1}{A+a+e+f+2} - \frac{B+e+1}{B+b+e+f+2} \tag{3.7}$$

The bias (e, f) is seen to be equivalent to an unbiased a priori, $e = f = 0$, where one starts not at the origin of the lattice but at the point (e, e, f, f). One now of course tries A, or B, or flips a coin, according to $\rho_A - \rho_B > 0$, < 0, or $= 0$, and this is equivalent to the treatment function

$$T(A, B, a, b) = (A+e+1)(b+f+1) - (B+e+1)(a+f+1) \tag{3.8}$$

which is the case we now consider.

The walk determined by (2.5) T given by (3.8) has been analyzed algebraically and by extensive computer simulation,[4] mainly from the point of view of its asymptotic large N behavior. We can distinguish two major asymptotic forms of the walk:

$$\begin{aligned} &\text{(i)} && C_\infty = \lim_{N\to\infty} C_N \text{ is finite,} \\ &\text{(ii)} && c_\infty = \lim_{N\to\infty} C_N/N > 0 \end{aligned} \tag{3.9}$$

corresponding, respectively, to saturation of mean excess failures, and to a constant failure rate no matter how long the testing, a most inappropriate situation in the context of clinical trials. An intermediate asymptotic form will also appear.

Let us start with a special case of (2.5), (3.8), simple enough to be solved analytically. It is that in which the a priori distribution is unbiased, $e = f = 0$, and the unknown probability p_A in fact has the value 1. The solution that is to be found in Section 8 then yields a finite value of C_∞ only in the interval $0 \leqslant p_B < 0.5$, while in the interval $0.5 < p_B \leqslant 1$, it is the failure rate c_∞ that is finite (although small: $c_\infty < 0.072$). Thus, (see Fig. 1)

$$\begin{aligned} C_\infty &= \frac{q_B^2}{(q_B+1)(2q_B-1)}, && 0 \leqslant p_B < \frac{1}{2} \\ c_\infty &= \frac{q_B(\frac{1}{2} - q_B)}{1 - \frac{1}{2}q_B}, && \frac{1}{2} < p_B \leqslant 1 \end{aligned} \tag{3.10}$$

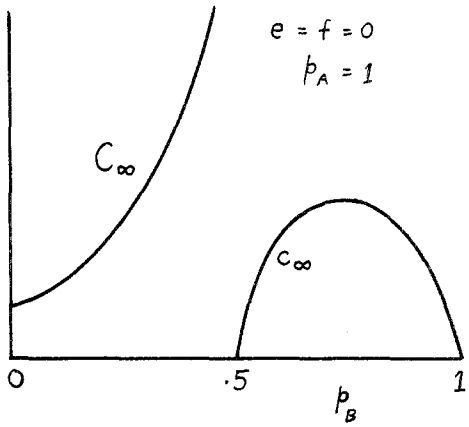

Fig. 1. Limiting number of failures, C_∞, and failure rate, c_∞, for unbiased trials with $p_A = 1$.

(it can also be shown that $\lim_{N\to\infty} C_N/N^{1/2} = \frac{1}{3}(2/\pi)^{1/2}$ *at* $p_B = 1/2$) and there is a phase transition between two distinct asymptotic forms at $p_B = 0.5$.

Further characterization of these forms is helpful. From the same analysis we find as $N \to \infty$

$$\begin{aligned} &\sigma_N = O(1), && 0 \leqslant p_B < 1/2 \\ &\sigma_N \to N q_B \left(\tfrac{1}{2} - q_B/1 - \tfrac{1}{2}q_B\right)^{1/2}, && 1/2 < p_B \leqslant 1 \end{aligned} \tag{3.11}$$

In case (i), the excess failure number

$$\chi_N = a + b - Nq_A \tag{3.2}$$

settles down to a distribution of finite mean and variance. But in case (ii),

$$\begin{aligned} E(\chi/N) &\to q_B \frac{\frac{1}{2} - q_B}{1 - \frac{1}{2}q_B} \\ \sigma(\chi/N) &\to q_B \left(\frac{1/2 - q_B}{1 - \frac{1}{2}q_B} \right)^{1/2} \end{aligned} \tag{3.12}$$

This is consistent with a bimodal distribution $p(\chi/N)$: A is selected with a mean failure rate $\chi/N = 0$, or B is selected with $\chi/N = q_B$; the former occurs with a probability $1 - \beta$, the latter with β, and the *trapping ratio* is

$$\beta = \frac{\frac{1}{2} - q_B}{1 - \frac{1}{2}q_B} \tag{3.13}$$

In fact, the description is true, and so there is literally a transition from the

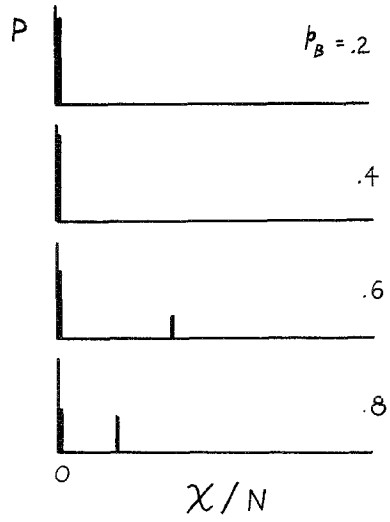

Fig. 2. Probability distributions p of mean failure rate χ/N for various values of p_B, in unbiased trials with $p_A = 1$.

distribution concentrated in A-phase to one divided between A and B-phases as one passes the $p_B = 0.5$ mark (Fig. 2).

The controlling parameters e and f, as the a priori bias towards success or failure, can now be altered. We have noted that this is equivalent to merely changing the origin of the walk. Nonetheless, this changes the nature of the system and hence the location of the transition. It will be seen that a bias towards success constrains the constant failure rate region

$$p_A = 1 \text{ bias}(e, 0) : p_{B\,\text{tr.}} = \frac{e+1}{e+2} \tag{3.14}$$

which of course is desirable in the clinical trials context.

Now proceeding to the general case of $p_A \neq 1$, there will be a transition curve, above which the mean failure rate approaches a nonzero constant—a fraction of the population—but below which the total number of failures is bounded as the process goes on to infinity. The exact analysis is very complicated, see Section 7, but extensive computer simulations[4] have been done. Typical results are shown, first in the no-bias case, with a schematic indication of the asymptotic relative frequencies of testing with A or B in the various regions. As one traverses a path in (p_A, p_B) space, one makes the transition from pure A-phase, to two-phase, to pure B-phase. The effectiveness of bias towards success is also shown, in which the $C_\infty = \infty$ region is greatly contracted, indicating a potentially practical procedure (Figs. 3 and 4).

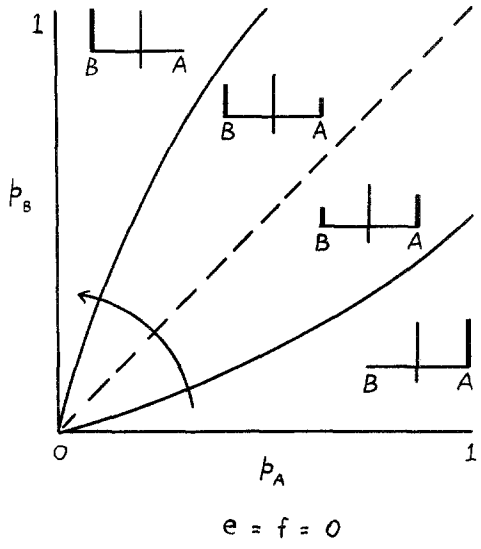

Fig. 3. Relative frequencies of testing A and B, for typical points in (p_A, p_B) plane, in unterminated unbiased trials.

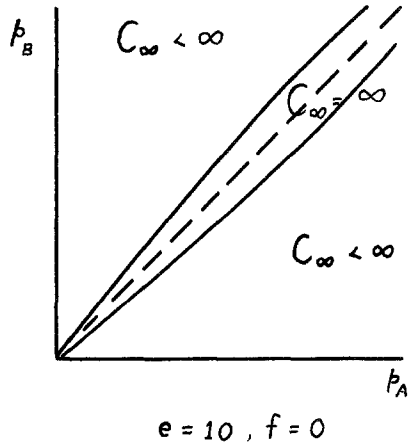

Fig. 4. Limiting number of excess failures for unterminated trials with bias $e = 10, f = 0$.

4. ONE-DIMENSIONAL EXAMPLE

We have indicated that the phase transition in our four-dimensional walk is most easily expressed in terms of the asymptotic partitioning of the walk between $\theta(T) > 0$ and $\theta(T) < 0$, with a ratio $1 - \beta : \beta$. Indeed we

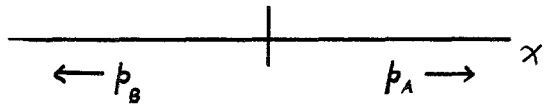

Fig. 5. Piecewise homogeneous one-dimensional walk.

have

$$c_\infty = (p_A - p_B)\,\beta \tag{4.1}$$

where the trapping probability is $\beta = 0$ in case (i) of (3.9). But we do not have to go to four-dimensional space to observe this effect. Consider the following one-dimensional walk on an integer lattice.[7,8] A move is a unit step to the left, or right, with probabilities q_A, or p_A, when $x > 0$, but p_B, or q_B, when $x < 0$, so that the medium is piecewise homogeneous unless $q_B = p_A$. Any rule can be used for a move from the origin. This walk has a much simpler phase diagram, intuitively obvious when one realizes that the mean drift per step is $p_A - q_A$ (or $p_B - q_B$), permitting a diffusive peak to escape from the origin only if $p_A - q_A > 0$ (or $p_B - q_B > 0$). The detailed form of the limiting—or asymptotic—distribution is easy to find. It depends on α, the fraction of the walk lodged on the right hand, with $\beta = 1 - \alpha$ on the left. There is then a single diffusion process when $\alpha = 0$ or $\beta = 0$, in the regions shown, a superposition of the two when $\alpha\beta > 0$, and

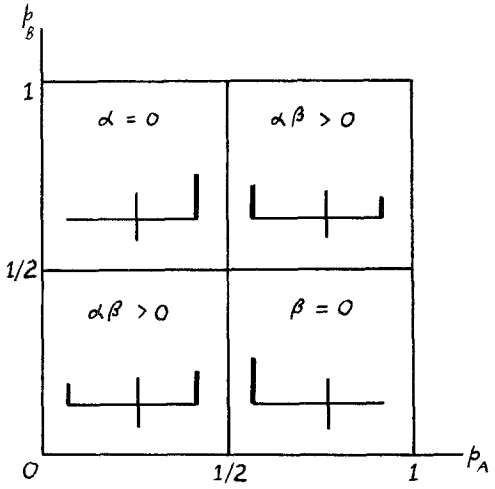

Fig. 6. (p_A, p_B) dependence of relative frequencies of remaining on left or right half line for walk of Fig. 5.

singular cases, e.g.,

$$P_\infty(x) \propto (p/q)^{|x|-1} \qquad \text{when} \quad p_A = p_B = p < 1 \tag{4.2}$$

on the boundaries of these regions.

The one-dimensional walk is in fact isomorphic to the two-dimensional walk shown, which is an obvious prototype of the four-dimensional case. Here, A denotes the number of steps to the right, B the number upward, and there are two homogeneous regions on opposite sides of the diagonal, which serves as boundary. The one-dimensional image simply chooses $x = A - B$. If, e.g., $p_A > p_B$, the walk may likewise be characterized by the relative success rate $A - B$, and it will drift strictly below the diagonal if both q_A/p_A and p_B/q_B are less than unity, the slope of the diagonal.

Our qualitative arguments, it must be observed, do not depend at all on the discrete nature of the lattice medium: any passage from top to bottom must hit a boundary point, and a continuous description is quite adequate, except for numerical details. Matters need not be this simple. Suppose that the boundary between the regions is given instead by $A = 2B$. Then the lattice points on the boundary, the solid circles, are not the only points one can pass through to cross the boundary. They are joined by the even more numerous open circles, an effect which increases with further decreasing rational slope, and becomes even more complex with irrational slope. But this is precisely what happens in our four-dimensional random

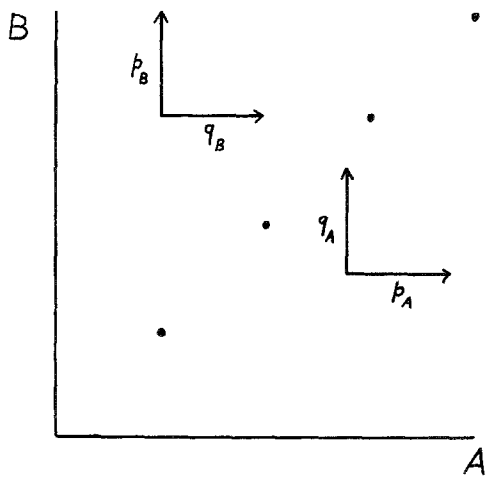

Fig. 7. Two-dimensional walk isomorphic to that of Fig. 5.

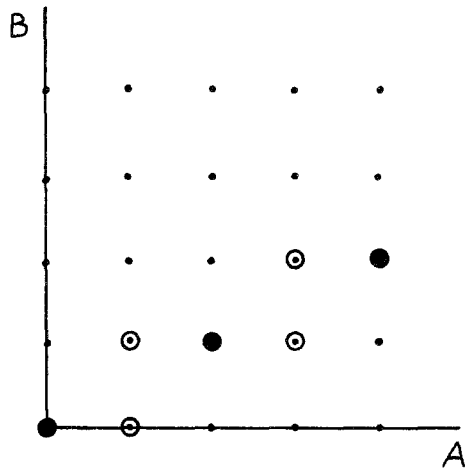

Fig. 8. The lattice points of the interregion boundary $A = 2B$ can be both on, • or adjacent to, ⊙, the boundary line.

walk: the dividing plane

$$(A + e + 1)(b + f + 1) = (B + e + 1)(a + f + 1) \tag{4.3}$$

has an unknown rational tangent insofar as the pair (A, a) is concerned. Nonetheless, it is possible in limiting cases to use continuum methods to approximate the solutions, and this is now being carried out.

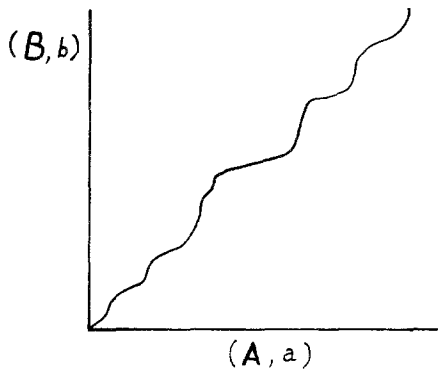

Fig. 9. Lattice points of the interregion boundary (4.3) are scattered on and near the hyperplane (4.3).

5. TECHNIQUES—TWO-DIMENSIONAL PROTOTYPE

To illustrate analytic methods available for the solution of piecewise homogeneous walks, we start with the elementary case of the two-dimensional walk of Section 4. A standard method of treating a *homogeneous* walk is to set up a generating function for the probability distribution of the Nth step. Here, we must have $A + B = N$, so that

$$\sum_{A+B=N} P_{AB} = 1 \tag{5.1}$$

but it will be convenient to define a generating function for the complete set of probabilities

$$P(z, y) = \sum_{A,B} P_{AB} z^A y^B \tag{5.2}$$

The two regions of this walk are identified by the sign of $A - B$, so we further introduce a "tag" t to make this distinction, and write

$$P(z, y, t) = \sum_{A,B} P_{AB} z^A y^B t^{A-B} \tag{5.3}$$

Let us see how (5.3) is employed. To start, we have to define the transition probability for the walk, and in conformity with the picture, Fig. 7, this will be taken as

$$\begin{aligned} P_{AB} = &\left[p_A \theta(A - 1 - B) + q_B \bar{\theta}(A - 1 - B) \right] P_{A-1,B} \\ &+ \left[q_A \theta(A + 1 - B) + p_B \bar{\theta}(A + 1 - B) \right] P_{A,B-1} + \delta_{(A,B),(0,0)} \end{aligned} \tag{5.4}$$

Now introduce the MacLaurin and Laurent parts of $P(z, y, t)$ with respect to t,

$$P(z, y, t) = P^1(z, y, t) + P^2(z, y, t)$$

where

$$\begin{aligned} P^1(z, y, t) &= \sum_{A,B} \theta(A - B) P_{AB} z^A y^B t^{A-B} \\ P^2(z, y, 1/t) &= \sum_{A,B} \theta(B - A) P_{AB} z^A y^B t^{A-B} \end{aligned} \tag{5.5}$$

These identify the two uniform regions of the lattice, and (5.4) goes over to the generating function equation

$$\begin{aligned} P^1(z, y, t) + P^2(z, y, 1/t) = &(p_A zt + q_A y/t) P^1(z, y, t) \\ &+ (q_B zt + p_B y/t) P^2(z, y, 1/t) + 1 \end{aligned} \tag{5.6}$$

or

$$(1 - p_A zt - q_A y/t)P^1(z, y, t) + (1 - q_B zt - p_B y/t)P^2(z, y, 1/t) = 1 \tag{5.7}$$

The equation (5.7) is very much typical of those that we will meet in more general cases, and can be solved by very similar tricks. We multiply by t and take the MacLaurin part. Since P^2 has only nonpositive powers of t, this yields

$$(t - p_A zt^2 - q_A y)P^1(z, y, t) = \alpha' + \beta' t + \gamma' t^2 \tag{5.8}$$

for suitable α', β', γ'. Now since $\sum_{A+B=N} P_{AB} = 1$, it is easily seen that $P^1(z, y, t)$ is analytic in t for $|t| \leqslant 1$ when $|z| < 1$, $|y| < 1$. Since $t - p_A zt^2 - q_A y$ has the two roots

$$t_{1\pm} = \frac{1}{2p_A z}\left[1 \pm (1 - 4p_A q_A yz)^{1/2}\right] \tag{5.9}$$

with $|t_{1-}| < q_A|y| < 1$ (but $|t_{1+}| > 1$), $\alpha' + \beta' t + \gamma' t^2$ must have the factor $t - t_{1-}$. Hence (5.8) takes the form

$$P^1(z, y, t) = \frac{\lambda' + \mu' t}{t - t_{1+}} \tag{5.10}$$

In exactly the same way, $(1/t) - q_B z - p_B y(1/t)^2$ has the two roots

$$1/t_{2\pm} = \frac{1}{2p_B y}\left[1 \pm (1 - 4p_B q_B yz)^{1/2}\right] \tag{5.11}$$

and we must have

$$P^2(z, y, 1/t) = \frac{\lambda'' + \mu''/t}{1/t - 1/t_{2+}} \tag{5.12}$$

The parameters $\lambda', \lambda'', \mu', \mu''$ are now determined by substituting (5.10) and (5.12) into (5.7), and using the fact, from (5.5), that

$$P^1(z, y, 0) = P^2(z, y, 0) \tag{5.13}$$

Doing so, we find after some algebra

$$\lambda' = t_{1+}/D, \qquad \mu' = q_B/p_A D$$

$$\lambda'' = 1/t_{2+} D, \qquad \mu'' = q_A/p_B D$$

where

$$D(y,z) = \left(\frac{t_{1-}}{t_{2-}} - 1\right)\left(\frac{p_B y}{t_{2+}} + p_A z t_{1+}\right)$$
$$= \frac{q_B + p_A}{2p_A}\left[q_A - p_A - (1 - 4p_A q_A yz)^{1/2}\right]$$
$$+ \frac{q_A + p_B}{2p_B}\left[q_B - p_B - (1 - 4p_B q_B yz)^{1/2}\right] \tag{5.14}$$

and hence conclude that

$$P^1(z,y,t) = \frac{(q_B/p_A)t + t_{1+}}{t - t_{1+}} \frac{1}{D(y,z)} \tag{5.15}$$

$$P^2(z,y,1/t) = \frac{(q_A/p_A)(1/t) + (1/t_{2+})}{(1/t) - (1/t_{2+})} \frac{1}{D(y,z)}$$

Consider the asymptotic distribution in $x = A - B$ for large $N = A + B$. For this purpose, we need the smallest singularity or singularities (in absolute value) in z of

$$P^1(z,z,t) = \sum P_{AB} z^{A+B} t^{A-B} \tag{5.16}$$

and similarly of $P^2(z,z,1/t)$. There are several possibilities. First, if $p_A < 1/2$, $p_B < 1/2$, then $p_A - q_A < 0$, $p_B - q_B < 0$ and the walk drifts to the origin in the one-dimensional version, or to the diagonal in the two-dimensional case, from both sides. Now the minimal singularities are simple poles at $|z| = 1$: $D(1,1) = D(-1,-1) = 0$. Thus

$$P^1(z,z,t) = -\left.\frac{D(z,z)P^1(z,z,t)}{(\partial/\partial z)D(z,z)}\right|_{z=1} \frac{1}{1-z}$$
$$-\left.\frac{D(z,z)P^1(z,z,t)}{(\partial/\partial z)D(z,z)}\right|_{z=-1} \frac{1}{1+z} + \cdots \tag{5.17}$$

leading to the required generating function

$$P_N^1(t) = \text{coef}\, z^N \qquad \text{in } P^1(z,z,t)$$
$$= \frac{(q_A - p_A)(q_B - p_B)}{(q_A + q_B)(q_A q_B - p_A p_B)}\left(\frac{1 + q_B t/q_A}{1 - p_A t/q_A} + (-1)^N \frac{1 - q_B t/q_A}{1 + p_A t/q_A}\right)$$
$$+ \cdots \tag{5.18}$$

and yielding the limiting distribution as $N \to \infty$

$$P^1(x) \propto (p_A/q_A)^x \qquad \text{for} \quad x \equiv N \pmod 2 \tag{5.19}$$

$P^2(z, z, 1/t)$ is of course analogous.

Suppose next that $p_A > 1/2$. Then $P^1(z, z, t)$ should yield a distribution drifting away from the origin (one-dimensional) or diagonal (two-dimensional) at the rate $p_A - q_A$. Now indeed the minimal singularity in $P^1(z, z, t)$ becomes a simple pole at $t_{1+} = t$ or

$$z_1 = \frac{1}{p_A t + q_A/t} \tag{5.20}$$

expanding about which

$$P^1(z, z, t) = \mathscr{A}(t)/(1 - z/z_1) + \cdots \tag{5.21}$$

for suitable $\mathscr{A}(t)$. We conclude that

$$P_N^1(t) = \mathscr{A}(t)(p_A t + q_A/t)^N + \cdots \tag{5.22}$$

But t in $\mathscr{A}(t)$ may be replaced by the maximizing value $t_0 = 1$ of $|p_A t + q_A/t|$ on the unit circle, so that

$$P_N^1(t) = C_1(p_A, p_B)(p_A t + q_A/t)^N + \cdots \tag{5.23}$$

precisely the generating function for a walk with probability p_A to the right —in the one-dimensional picture. Without going into detail, it is clear that if $p_B < q_B$, the left distribution drifts to the origin, giving just one diffusive peak, with $C_1(p_A, p_B) = 1$, but if $p_B > q_B$ there are two diffusive peaks, each drifting away from the origin, and $C_1(p_A, p_B) < 1$.

6. TECHNIQUES—FOUR-DIMENSIONAL PROTOTYPE

We proceed now to the four-dimensional walk (2.1), (2.3), and correspondingly set up the generating function [converting the arguments of $P(A, B, a, b)$ to indices]

$$P(z, y, x, w) = \sum P_{A,B,a,b} z^A y^B x^a w^b \tag{6.1}$$

Introducing the tag t, we incorporate a vehicle for the condition (2.1) by extending (6.1) to

$$P(z, y, x, w, t) = \sum P_{A,B,a,b} z^A y^B x^a w^b t^{T(A,B,a,b)} \tag{6.2}$$

and further split this into MacLaurin and Laurent parts:

$$\begin{aligned} P^1(z, y, x, w, t) &= \sum P_{A,B,a,b} \theta(T(A, B, a, b)) z^A y^B x^a w^b t^{T(A,B,a,b)} \\ P^2(z, y, x, w, 1/t) &= \sum P_{A,B,a,b} \bar{\theta}(T(A, B, a, b)) z^A y^B x^a w^b t^{T(A,B,a,b)} \end{aligned} \tag{6.3}$$

In order for the transition probability equation (2.5) to be directly expressible in terms of P^1 and P^2, it is necessary that $\Delta_A T(A,B,a,b)$—and the other differences—be linear in A,B,a,b, i.e., that $T(A,B,a,b)$ be at most bilinear, which is indeed the case for the "ethical" form (3.8). Here however we investigate a still simpler case.

Suppose that $T(A,B,a,b)$ is given by

$$T(A,B,a,b) = b - a \tag{6.4}$$

In terms of the clinical trials context, this means that the drug that has failed the fewest number of times is used next, a procedure which in fact has been suggested, with various elaborations. The most obvious is to require that if $b = a$, then $T = A - B$: the more successful drug is used next. This has been analyzed in detail.[9,10] Here, however, we stick to (6.4), in which case the transition probability equation becomes simply

$$\begin{aligned}(1 - p_A z - q_A x/t)P^1(x, y, x, w, t) \\ + (1 - p_B y - q_B wt)P^2(z, y, x, w, 1/t) = 1\end{aligned} \tag{6.5}$$

As in (5.7), we multiply by t and conclude that

$$\left[(1 - p_A z)t - q_A x\right]P^1(z, y, x, w, t) = \alpha' + \beta' t + \gamma' t^2 \tag{6.6}$$

Since $|q_A x/1 - p_A z| < 1$ in the allowed domain, the linear factor $(1 - p_A z)t - q_A x$ must divide the right-hand side, so that

$$P^1(z, y, x, w, t) = \lambda' + \mu' t \tag{6.7}$$

Similarly we have

$$P^2(z, y, x, w, 1/t) = \lambda'' + \mu''/t \tag{6.8}$$

If (6.7) and (6.8) are substituted into (6.5) and the condition $P^1(z, y, x, w, 0) = P^2(z, y, x, w, 0)$ imposed, we find at once by equating equal powers of t

$$\begin{aligned}\lambda' = \lambda'' &= (1 - p_A z)(1 - p_B y)/D \\ \mu' &= q_B w(1 - p_B y)/D \\ \mu'' &= (1 - p_A z)q_A x/D\end{aligned} \tag{6.9}$$

where $D = (2 - p_A z - p_B y)[(1 - p_A z)(1 - p_B y) - q_A q_B xw]$. Thus we conclude that

$$\begin{aligned}P^1(z, y, x, w, t) &= (1 - p_A z + q_B wt)(1 - p_B y)/D \\ P^2(z, y, x, w, 1/t) &= (1 - p_B y + q_A x/t)(1 - p_A z)/D\end{aligned} \tag{6.10}$$

That only t^{-1}, t^0, t^1 can occur is obvious: a drug that has failed less is tried

until it fails as much as the other, and there is no way that $|a-b|>1$ can occur.

A typical characterization of this walk would be the asymptotic distribution of the excess failures, $\chi \equiv a+b-Nq_A$ when $p_A > p_B$. Clearly

$$P(z,\theta) \equiv \sum P_{A,B,a,b} z^N e^{i(a+b-Nq_A)\theta}$$
$$= P(ze^{-iq_A\theta}, ze^{-iq_A\theta}, ze^{ip_A\theta}, ze^{ip_A\theta}) \tag{6.11}$$

so that the asymptotic form depends upon the smallest singularity of

$$D = \left[2-(p_A+p_B)ze^{-iq_A\theta}\right]$$
$$\times\left[(1-p_Aze^{-iq_A\theta})(1-p_Bze^{-iq_B\theta}) - q_Aq_Bz^2e^{2ip_A\theta}\right] \tag{6.12}$$

The roots occur at

$$(z_0e^{-iq_A\theta})^{-1} = \tfrac{1}{2}(p_A+p_B) + \tfrac{1}{2}u\left[(q_A+q_B)^2 + 4q_Aq_B(e^{2i\theta}-1)\right]^{1/2} \tag{6.13}$$

where $u=-1,0,1$, and so the smallest has $u=1$. The moment generating function $P_N(\theta) = \text{coef}\, z^N$ in $P(z,\theta)$ thus takes the asymptotic form

$$P_N(\theta) = \mathscr{A}(\theta) z_0^{-N} \tag{6.14}$$

$|z_0|$ is maximum on the unit circle $e^{i\theta}$ at $\theta=0$, at which $z_0=1$. Thus, $\mathscr{A}(\theta)=1$, and on expanding $\ln(z_0)$ about $\theta=0$, we find

$$P_N(\theta) = \exp\left(-q_A\frac{q_A-q_B}{q_A+q_B}iN\theta\right)\exp\left\{-2\frac{q_Aq_B}{q_A+q_B}\left[1+\frac{p_Ap_Bq_Aq_B}{(q_A+q_B)^2}\right]N\theta^2\right\}$$
$$+\cdots \tag{6.15}$$

showing that χ diffuses with a drift given by

$$\bar{\chi}_N = N\frac{q_A}{q_A+q_B}(p_A-p_B) \tag{6.16}$$

There is no "phase transition."

In the clinical trials context, the persistent relative failure rate (6.16) is hardly acceptable, and so (6.4) or its variants are used with an associated stopping rule. The testing is thus terminated at a finite stage and the nominally superior drug used thereafter. The analysis, however, is readily incorporated into (6.3) and is presented in another publication.

7. TECHNIQUES—ETHICAL STRATEGY

The "ethical" strategy characterized by

$$T(A,B,a,b) = (A+1+e)(b+1+f) - (B+1+e)(a+1+f) \tag{7.1}$$

can be treated, in principle, by the technique of the previous two sections. We have, for example,

$$\begin{aligned}
&\sum \Delta_A \left[P_{ABab} \theta\big(T(A,B,a,b)\big) \right] z^A y^B x^a w^b t^{T(A,B,a,b)} \\
&\quad = \sum P_{ABab} \theta\big(T(A,B,a,b)\big) z^A y^B x^a y^b t^{T(A,B,a,b)} \left[1 - z t^{\Delta_A T(A+1,B,a,b)} \right] \\
&\quad = \sum P_{ABab} \theta\big(T(A,B,a,b)\big) z^A y^B x^a w^b t^{T(A,B,a,b)} \left[1 - z t^{b+1+f} \right] \\
&\quad = \sum P_{ABab} \theta\big(T(A,B,a,b)\big) \Big[z^A y^B x^a w^b t^{T(A,B,a,b)} \\
&\qquad\qquad - z t^{f+1} z^A y^B x^a (wt)^b t^{T(A,B,a,b)} \Big]
\end{aligned}$$

and similarly for the remaining terms, so that the transition probability equation (2.5) then takes the form

$$\begin{aligned}
&P^1(z,y,x,w,t) - z t^{f+1} p_A P^1(z,y,x,wt,t) - x t^{-e-1} q_A P^1(z,y/t,x,w,t) \\
&\quad + P^2(z,y,x,w,1/t) - y t^{-f-1} p_A P^2(z,y,x/t,w,1/t) \\
&\quad - w t^{e+1} q_B P^2(zt,y,x,w,1/t) = 1
\end{aligned} \tag{7.2}$$

Equation (7.2) is a bit more complicated than need be. We observe, however, that

$$P_{ABab} = p_A^A p_B^B q_A^a q_B^b Q_{ABab} \tag{7.3}$$

where Q_{ABab} is a strictly combinatorial factor, the number of ways of arriving at (A,B,a,b). The corresponding generating function relation is

$$P(z,y,x,w,t) = Q(p_A z, p_B y, q_A x, q_B w, t) \tag{7.4}$$

and (7.3) transcribes at once to

$$\begin{aligned}
&Q^1(z,y,x,w,t) - z t^{f+1} Q^1(z,y,x,wt,t) - x t^{-e-1} Q^1(z,y/t,x,w,t) \\
&\quad + Q^2(z,y,x,w,1/t) - y t^{-f-1} Q^2(z,y,x/t,w,1/t) \\
&\quad - w t^{e+1} Q^2(zt,y,x,w,1/t) = 1
\end{aligned} \tag{7.5}$$

As a functional equation, not merely an algebraic one, (7.5) does not yield directly to a power series decomposition. A simple MacLaurin–Laurent separation is not possible because of the mixed powers of t and $1/t$, but this can be avoided by using only partial generating functions. For example, set

$$Q_{AB}(x,w,t) = \sum Q_{ABab} x^a w^b t^{T(A,B,a,b)} \tag{7.6}$$

with the corresponding Q_{AB}^1 and Q_{AB}^2. On taking the coefficient of $z^A y^B$ in

(7.5), we thus have

$$F_{AB}(x,w,t) + G_{AB}(x,w,1/t) = \delta_{A0}\delta_{B0}$$

where

$$F_A(x,w,t) = \left(1 - \frac{x}{t^{B+e+1}}\right)Q^1_{AB}(x,w,t) - t^{f+1}Q^1_{A-1B}(x,wt,t) \tag{7.7}$$

$$G_{AB}(x,w,1/t) = (1 - wt^{A+e+1})Q^2_{AB}(x,w,1/t) - \frac{1}{t^{f+1}}\, Q^2_{AB-1}(x/t,w,1/t)$$

to which we append the boundary condition

$$Q^1_{AB}(x,w,0) = Q^2_{AB}(x,w,0) \tag{7.8}$$

F_{AB} is seen to be of degree $\geqslant -B-e-1$ in t, G_{AB} of degree $\leqslant A+e+1$. Hence we can write

$$\begin{aligned} F_{AB}(x,w,t) &= \sum_{i=-(B+e+1)}^{A+e+1} F^{(i)}_{AB}(x,w)t^i \\ G_{AB}(x,w,1/t) &= \sum_{i=-(A+e+1)}^{B+e+1} G^{(i)}_{AB}(x,w)t^{-i} \end{aligned} \tag{7.9}$$

and (7.7) becomes

$$F^{(i)}_{AB}(x,w) + G^{(-i)}_{AB}(x,w) = \delta_{A0}\delta_{B0}\delta_{i0} \tag{7.10}$$

Further, taking the coefficient of $t^{-(B+e+1)}$ in F_{AB} and of t^{A+e+1} in G_{AB}, we have, respectively,

$$\begin{aligned} F^{(-B-e-1)}_{AB}(x,w) &= -xQ^1_{AB}(x,w,0) \\ G^{(-A-e-1)}_{AB}(x,w) &= -wQ^2_{AB}(x,w,0) \end{aligned} \tag{7.11}$$

reducing the boundary condition to

$$wF^{-(B+e+1)}_{AB}(x,w) = xG^{-(A+e+1)}_{AB}(x,w) \tag{7.12}$$

Our problem then is to solve the system (7.10), (7.12) for Q_{ABab} subject to the condition that $Q^1_{AB}(x,w,t)$ is representable by a MacLaurin series in t, $Q^2_{AB}(x,w,1/t)$ by a Laurent series. The radii of convergence of these series are crucial. From (2.4) and (7.3), we see that

$$\sum Q_{ABab}\lambda^{A+B+a+b}p_A^A p_B^B q_A^a q_B^b = 1/1-\lambda \tag{7.13}$$

for positive real arguments satisfying $0 \leqslant p_A = 1 - q_A \leqslant 1$, $0 \leqslant p_B = 1 - q_B \leqslant 1$ and $|\lambda| < 1$. It follows that if $|t| \leqslant 1$ and $|x| \leqslant \lambda < 1$, $|w| \leqslant \lambda < 1$,

then

$$\left|\sum_{ab} Q^1_{ABab} x^a w^b t^{T(A,B,a,b)}\right|$$

$$\leqslant \sum_{ab} Q^1_{ABab} |x|^a |w|^b = \sum_{ab} Q^1_{ABab} \lambda^{a+b} (|x|/\lambda)^a (|w|/\lambda)^b$$

$$= (\lambda - |x|)^{-A} (\lambda - |w|)^{-B} \sum_{ab} Q^1_{ABab}$$

$$\times \lambda^{A+a+B+b} \left(1 - \frac{|x|}{\lambda}\right)^A \left(1 - \frac{|w|}{\lambda}\right)^B$$

$$\times \left(\frac{|x|}{\lambda}\right)^a \left(\frac{|w|}{\lambda}\right)^b$$

$$\leqslant (1-\lambda)^{-1} (\lambda - |x|)^{-A} (\lambda - |w|)^{-B}$$

In other words

$$Q^1_{AB}(x,w,t) \text{ is analytic in } t \text{ for } |t| \leqslant 1 \text{ when } |x| < 1,\ |w| < 1 \quad (7.14)$$

and in similar fashion,

$$Q_{AB}(x,w,1/t) \text{ is analytic in } t \text{ for } |t| \geqslant 1 \text{ when } |x| < 1,\ |w| < 1 \quad (7.15)$$

8. FIRST SPECIAL CASE: $p_A = 1, f = 0$

The complete evaluation of Q_{ABab} has yet to be carried out, but several special cases yield rather easily. We consider first that in which $p_A = 1$ or $q_A = 0$. Then, according to (7.4), we need only the generating function $Q(z, y, 0, w)$ and hence only $Q^1_{AB}(0,w,t)$, $Q^2_{AB}(0,w,1/t)$. In fact, in this case, both indices are unnecessary, and the intermediate generating function

$$Q^1_A(y,0,w,t) = \sum Q^1_{AB}(0,w,t) y^B \quad (8.1)$$

and its analogs will be used instead. From (7.7), we have at once

$$Q^1_A(y,0,w,t) - t^{f+1} Q^1_{A-1}(y,0,wt,t) - \delta_{A,0}$$

$$= (wt^{A+e+1} - 1 + yt^{-f-1})\, Q^2_A(y,0,w,1/t) \quad (8.2)$$

where both sides are polynomials in t of degree $\leqslant A + e + 1$.

Now consider the equation

$$wt^{A+e+f+2} - t^{f+1} + y = 0 \quad (8.3)$$

corresponding to the polynomial factor of the right-hand side of (8.2). For $w \to 0$, there are clearly $f+1$ roots u_α^0 with $|u_\alpha^0| < 1$, the remaining $A + e + 1$ having $|u_\beta^1| > 1$, and application of Rouche's theorem to the unit circle extends this to a finite domain on (y, w) space. Thus writing

$$\left[Q_A^1(y,0,w,t) - t^{f+1} Q_{A-1}^1(y,0,w,t,t) - \delta_{A,0} \right] \Big/ \prod_1^{A+e+1} \left(t - u_\beta^1\right)$$

$$= w \prod_1^{f+1} (1 - u_\alpha^0/t) Q_A^2(y,0,w,1/t) \tag{8.4}$$

the left-hand side is analytic for $|t| \leqslant 1$, the right-hand side for $|1/t| \leqslant 1$ [proved as in (7.14)]. We conclude from Dirichlet's theorem[11] that

$$Q_A^1(y,0,w,t) - t^{f+1} Q_{A-1}^1(y,0,wt,t) - \delta_{A,0} = C_A(y,w) \prod_1^{A+e+1} \left(t - u_\beta^1\right) \tag{8.5}$$

$$Q_A^2(y,0,w,1/t) = C_A(y,w) \Big/ \left[w \prod_1^{f+1} (1 - y_\alpha^0/t) \right]$$

for some t-independent $C_A(y,w)$. Setting $t = 0$ in the first of (8.5), $Q_A^1(y,0,w,0) = \delta_{A,0} + \prod_1^{A+e+1}(-u_\beta^1) C_A(y,w)$, and setting $1/t = 0$ in the second of (8.5), $Q_A^2(y,0,w,0) = C_A(y,w)/w$. Since $Q_A^1(y,0,w,0) = Q_A^2(y,0,w,0)$, we thus find

$$C_A(y,w) = w\delta_{A,0} \Big/ \left[1 - w \prod_1^{A+e+1} \left(-u_\beta^1\right) \right] \tag{8.6}$$

Hence, setting $t = 1$ in (8.5), we recover the required pair of generating functions

$$Q_A^1(y,0,w) = \frac{w \prod_1^{e+1} \left(1 - u_\beta^1\right)}{1 - w \prod_1^{e+1} \left(-u_\beta^1\right)} + 1$$

$$Q_A^2(y,0,w) = \frac{\delta_{A0}}{\prod_1^{f+1} (1 - u_\alpha^0) \left[1 - w \prod_1^{e+1} \left(-u_\beta^1\right) \right]} \tag{8.7}$$

where

$$w u^{e+f+2} - u^{f+1} + y = 0, \qquad |u_\beta^0| < 1, \qquad |u_\beta^1| > 1,$$

or summing with z^A,

$$Q^1(z,y,0,w) = \frac{w}{1-z}\,\frac{\prod_1^{e+1}(1-u_\beta^1)}{1-w\prod_1^{e+1}(-u_\beta^1)} + \frac{1}{1-z}$$
$$Q^2(z,y,0,w) = \frac{1}{\prod_1^{f+1}(1-u_\alpha^0)\left[1-w\prod_1^{e+1}(-u_\beta^1)\right]} \tag{8.8}$$

Finally, adding (8.8) and using the fact that $w\prod_1^{e+1}(1-u_\alpha^0)\prod_1^{f+1}(1-u_\beta^1) = w-1+y$, we have the full generating function

$$Q(z,y,0,w) = \frac{w-z+y}{(1-z)\prod_1^{f+1}(1-u_\alpha^0)\left[1-w\prod_1^{e+1}(-u_\beta^1)\right]} + \frac{1}{1-z} \tag{8.9}$$

The special case $f=0$ has turned out empirically to be important,[4] and simplifies matters materially. Inserting $\lambda^{A+B+a+b}$ to generate the total number of steps, and observing that if $u \equiv u_0^0$, then $w(-u)\prod_1^{e+1}(-u_\beta^1) = y$, we see that then

$$P(z\lambda, y\lambda, 0, w\lambda) = Q(z\lambda, p_B y\lambda, 0, q_B w\lambda)$$
$$= \frac{(q_B w + p_B y - z)\lambda}{(1-\lambda z)(1-u)(1+p_B\lambda y/u)} + \frac{1}{1-\lambda z} \tag{8.10}$$

where

$$q_B\lambda w u^{e+2} - u + p_B\lambda y = 0, \qquad |u| < 1$$

Since $p_A = 1$, $q_A = 0$, the number of excess failures here is simply $\chi = b$. Thus the generating function for the number of steps and number of excess failures becomes

$$\sum_{N,\chi} \lambda^N w^\chi P_N(\chi) = P(\lambda,\lambda,0,w\lambda)$$
$$= \frac{q_B(w-1)}{(1-\lambda)(1-u)(1+p_B\lambda/u)} + \frac{1}{1-\lambda} \tag{8.11}$$
$$q_B\lambda w u^{e+2} - u + p_B\lambda = 0$$

To see that there are two distinct parameter regimes, let us first ask for the limit of the generating function of the χ-distribution as $N\to\infty$

$$\sum w^\chi P_\infty(\chi) = \lim_{N\to\infty} \sum w^\chi p_N(\chi) \tag{8.12}$$

We recall that in general

$$\lim_{N\to\infty} f_N = f_0 + \sum_1^\infty (f_n - f_{n-1})$$

$$= \lim_{\lambda\to 1} f_0 + \sum_1^\infty (f_n - f_{n-1})\lambda^n$$

or

$$\lim_{N\to\infty} f_N = \lim_{\lambda\to 1} (1-\lambda)\sum_0^\infty f_n \lambda^n \tag{8.13}$$

Hence multiplying (8.11) by $1-\lambda$ and letting $\lambda \to 1$,

$$\sum w^\chi P_\infty(\chi) = 1 - \frac{(1-w)q_B}{(1-u)(1+p_B/u)} \tag{8.14}$$

where

$$q_B w u^{e+2} - u + p_B = 0, \qquad |u| < 1$$

But (8.14) is the generating function of a distribution only if

$$\lim_{w\to 1} \sum w^\chi P_\infty(\chi) = \sum P_\infty(\chi) = 1$$

and hence only if the denominator of (8.14) does not vanish as $1-w$ when $w \to 1$. It is easily shown that the minimum root of $q_B w u^{e+2} - u + p_B = 0$ is at $|u| < 1$ when $w = 1$ for small enough p_B, at $u = 1$ for large p_B, with the crossover at p_B for which a double root exists at $w = 1$: $q_B u^{e+2} - u + p_B = 0$ and $(e+2)q_B u^{e+1} - 1 = 0$, implying $p_B = e + 1/e + 2$. Thus

$$P_\infty(\chi) \text{ is a limiting distribution when } p_B < \frac{e+1}{e+2} \tag{8.15}$$

and all of the moments of the mean excess failures remain finite, e.g., after a little algebra

$$C_\infty = \bar{\chi} = \frac{\partial}{\partial w} \sum w^\chi P_\infty(\chi)\Big|_{w=1}$$

$$= \frac{q_B}{(1-u_1)(1+p_B/u_1)}, \qquad u_1^{e+1} + u_1^e + \cdots + u_1 = p_B/q_B \tag{8.16}$$

As a special case,

$$\text{if} \quad e = 0, \qquad C_\infty = \frac{q_B^2}{(q_B - p_B)(1+q_B)} \qquad \text{for} \quad 0 \leqslant p_B < 1/2 \tag{8.17}$$

If $p_B > (e+1)/(e+2)$, (8.14) is no longer the generating function of a limiting distribution. Instead, the distribution becomes bimodal, and a

portion goes asymptotically to infinity, thereby not appearing in the limit (8.14). According to (8.14), the normalization of this "trapped" mode is given by

$$\beta = \lim_{w\to 1} \frac{(1-w)q_B}{(1-u)(1+p_B/u)} = \frac{1-(e+2)q_B}{2-q_B} \tag{8.18}$$

To ascertain the nature of the trapped mode, we return to (8.11) and write explicitly

$$\sum_\chi w^\chi P_N(\chi) = \frac{q}{2\pi i}\oint \left[\frac{q_B(w-1)}{(1-\lambda)(1-u)(1+p_B\lambda/u)} + \frac{1}{1-\lambda} \right]_\lambda \frac{d\lambda}{N+1} \tag{8.19}$$

The singularities in λ that dominate the large N behavior are poles at $\lambda_1 = 1$, $u(\lambda_2) = 1$ or $\lambda_2 = 1/(p_B + q_B w)$, and $u(\lambda_3) = -\lambda_3 p_B$ or $\lambda_3 = (-2/p_B^{e+1} q_B w)^{1/e+2}$, written in increasing order of magnitude. The pole $\lambda_1 = 1$ describes the limiting component (8.14). λ_2 can reach 1 for w on the unit circle, but $|\lambda_3|$ always exceeds unity. Thus the residual distribution is described asymptotically by $u = 1$, $\lambda = 1/(p_B + q_B w)$. Since

$$\frac{du}{d\lambda} = \frac{p_B + q_B w u^{e+2}}{1-(e+2)q_B\lambda w u^{e+1}} \to \frac{1}{p_B - (e+1)q_B w} \qquad \text{as} \quad u\to 1 \tag{8.20}$$

a residue evaluation of (8.19) yields at once the asymptotic form

$$p_B > \frac{e+1}{e+2} : \sum_\chi w^\chi P_N(\chi) = 1 - \frac{(1-w)q_B}{(1-u)(1+p_B/u)} + \frac{p_B - (e+1)q_B w}{2p_B + q_B w}(p_B + q_B w)^{N+3} \tag{8.21}$$

$(p_B + q_B w)^N$ is recognized as the generating function for a walk in which the walker has probability p_B of not moving, q_B of moving $+1$ to the right, precisely the probability distribution in the trapped mode, and the normalization—obtained by setting $w = 1$—which dominates at large N, is precisely β of (8.18).

9. SECOND SPECIAL CASE: $p_B = 0$, $e = f = 0$

The second special case we consider is that of $p_B = 0$ or $q_B = 1$. Since the full machinery of Section 7 must be invoked, we will avoid unnecessary complications by choosing $e = f = 0$ as well. Now only $P(z, 0, x, w)$ is required, and on dropping the constant $B = 0$ throughout (7.7)–(7.12)

reduces to

$$F_A(x,w,t) = \left(1 - \frac{x}{t}\right) Q_A^1(x,w,t) - tQ_{A-1}^1(x,wt,t) \tag{9.1}$$

$$G_A(x,w,1/t) = (1 - wt^{A+1})\, Q_A^2(x,w,1/t) \tag{9.2}$$

and

$$F_A(x,w,t) = \sum_{-1}^{A+1} F_A^{(i)}(x,w)t^i, \qquad G_A(x,w,1/t) = \sum_{-A-1}^{1} G_A^{(i)}(x,w)t^{-i} \tag{9.3}$$

$$F_A^{(i)}(x,w) + G_A^{(-i)}(x,w) = \delta_{A0}, \qquad wF_A^{(-1)}(x,w) = xG_A^{(-A-1)}(x,w)$$

Consider Eq. (9.2), written as $(1/t)^{A+1}G_A(x,w,1/t) = [(1/t)^{A+1} - w]\, Q_A^2(x,w,1/t)$. The left-hand side is a polynomial in $1/t$ of degree $A+2$. Since Q_A^2 is analytic around $|1/t| = |w|^{1/A+1}$, $(1/t)^{A+1} - w$ must be a factor of this polynomial, and so Q_A^2 must take the form

$$Q_A^2(x,w,1/t) = C_A(x,w) + D_A(x,w)/t \tag{9.4}$$

Setting $1/t = 0$, we have $C_A(x,w) = Q_A^2(x,w,0) = Q_A^1(x,w,0)$. But Eq. (9.3) tells us that $xG_A^{(-A-1)}(x,w) = -wG_A^{(1)}(x,w)$, and so, according to (9.2), $D_A = xC_A$. We see then that

$$Q_A^2(x,w,1/t) = \left(1 + \frac{x}{t}\right) Q_A^1(x,w,0) \tag{9.5}$$

and combining (9.1), (9.2), (9.3)

$$\left(1 - \frac{x}{t}\right) Q_A^1(x,w,t) - tQ_{A-1}^1(x,wt,t) + (1 - wt^{A+1})\left(1 + \frac{x}{t}\right) Q_A^1(x,w,0) = \delta_{A0} \tag{9.6}$$

An exchange of arguments and indices is now useful. If

$$\mathscr{Q}_b^1(z,x,t) = \sum Q_{A0ab}^1 z^A x^a t^{T(A,0,a,b)} \tag{9.7}$$

and similarly for $\mathscr{Q}^2$ and $\mathscr{Q}$, then on summing (9.6) with z^A and taking the coefficient of w^b, we have

$$\left(1 - \frac{x}{t} - t^{b+1}z\right)\mathscr{Q}_b^1(z,x,t) = \delta_{b0} - \left(1 + \frac{x}{t}\right)\left[\mathscr{Q}_b^1(z,x,0) - t\mathscr{Q}_{b-1}^1(zt,x,0)\right] \tag{9.8}$$

Now if $x > 0$, $z > 0$, $x + z < 1$, then $|x - t| > |zt^{b+2}|$ on the unit circle $|t| = 1$. It follows from Rouche's theorem[11] that

$$zt^{b+2} - t + x = 0 \tag{9.9}$$

has precisely one root, which we call $t_b(x,z)$, satisfying $|t_b| < 1$. Since $\mathscr{Q}_b^1(x,z,t)$ is anlaytic for $|t| < 1$, this implies that the right-hand side of

(9.8) vanishes at $t = t_b$:

$$\mathscr{Q}_b^1(z,x,0) - t_b\mathscr{Q}_{b-1}^1(zt_b,x,0) = \frac{t_b}{x+t_b}\delta_{b0} \tag{9.10}$$

and hence that

$$\begin{aligned} \mathscr{Q}_0^1(z,x,0) &= t_0/(x+t_0) \\ \mathscr{Q}_b^1(z,x,0) &= t_b\mathscr{Q}_{b-1}^1(zt_b,x,0) \qquad \text{for} \quad b>0 \end{aligned} \tag{9.11}$$

If we now choose

$$z = \frac{1-x/\tau}{\tau^{b+1}} \tag{9.12}$$

where τ is near x, then $t_b(x,z) = \tau$, and so

$$\begin{aligned} \mathscr{Q}_0^1\left(\frac{1-x/\tau}{\tau},x,0\right) &= \frac{\tau}{x+\tau} \\ \mathscr{Q}_b^1\left(\frac{1-x/\tau}{\tau^{b+1}},x,0\right) &= \tau\mathscr{Q}_{b-1}^1\left(\frac{1-x/\tau}{\tau^b},x,0\right) \end{aligned} \tag{9.13}$$

with the immediate solution

$$\mathscr{Q}_b^1\left(\frac{1-x/\tau}{\tau^{b+1}},x,0\right) = \tau^b\mathscr{Q}_0^1\left(\frac{1-x/\tau}{\tau},x,0\right) = \tau^{b+1}/x+\tau$$

Hence choosing $\tau = t_b(x,z)$ yields the desired

$$\mathscr{Q}_b^1(z,x,0) = t_b^{b+1}/(x+t_b) \tag{9.14}$$

It follows from (9.5) and (9.8) that, on letting $t\to 1$,

$$\begin{aligned} \mathscr{Q}_b^2(z,x) &= \frac{1+x}{t_b+x}t_b^{b+1} \\ \mathscr{Q}_b^1(z,x) &= \frac{1+x}{z-1-x}\left(\frac{t_b^{b+1}}{t_b+x} - \frac{t_{b-1}^b}{t_{b-1}+x}\right) - \frac{\delta_{b0}}{z-1-x} \end{aligned} \tag{9.15}$$

The single controlling parameter in this special case is p_A, but its influence is only qualitative, with no "phase transition" being evoked. Consider the mean excess failures, given according to (3.3) and (7.4) by

$$\begin{aligned} \sum C_N\lambda^N &= \frac{p_A\lambda}{1-\lambda}\,Q^2(p_A\lambda,0,q_A\lambda,\lambda) \\ &= p_A\frac{(1+q_A\lambda)}{1-\lambda}\sum_{b=0}^{\infty}\lambda^{b+1}t_b^{b+1}/(t_b+q_A\lambda) \\ p_A\lambda t_b^{b+2} - t_b + q_A\lambda &= 0, \qquad |t_b| < 1 \end{aligned} \tag{9.16}$$

Consequently,

$$C_\infty = p_A(1+q_A)\lim_{\lambda\to 1}\sum_{b=0}^{\infty}\lambda^{b+1}t_b^{b+1}/(t_b+q_A\lambda) \tag{9.17}$$

Indeed, if t_b behaves suitably as $\lambda\to 1$, the limit $\lambda\to 1$ can be taken term by term. We have seen, in the discussion of (9.9), that $t_b<1$ when $p_A\lambda+q_A\lambda=\lambda<1$. On the other hand, when $\lambda=1$, $f_b(t)=p_At^{b+2}-t+q_A=0$ does have the root $t=1$. The question is whether $t=1$ is the smaller or larger of the two positive roots when $\lambda=1$. But $f_b(0)>0$, $f_b(1)=0$, so that the condition for a root $0<t_b<1$ is that $f_b'(1)>0$ or $(b+2)p_A-1>0$. Hence

$$t_b\begin{cases}=1\\<1\end{cases}\quad\text{if}\quad b+2\begin{cases}\leqslant 1/p_A\\>1/p_A\end{cases} \tag{9.18}$$

Further, it is clear that at $\lambda=1$, $t_b\to q_A$ as $b\to\infty$. Thus we can reduce (9.17) to

$$\begin{gathered}C_\infty = p_A(1+q_A)\sum_{0}^{\infty}t_b^{b+1}/(t_b+q_A)\\ p_At_b^{b+2}-t_b+q_A=0\end{gathered} \tag{9.19}$$

which converges for all $p_A>0$. We conclude that the asymptotic mean excess failures remain finite for all $p_A>0$, implying a distribution concentrated about a repeated step of (p_A,q_A) in the (A,a) direction. There is no transition to a two-phase region.

10. SOME CONSIDERATIONS IN THE GENERAL CASE

Solution of the general walk with "ethical" treatment function, e.g., in the form (7.7)–(7.12), seems feasible, if complicated, but has yet to be accomplished. It is not difficult, however, to develop iterative approximations based upon the solved special cases. In particular, since the original motivation was to operate preferentially in the low mean excess failure region, it makes sense to expand about the point $p_A=1$, $p_B=0$ at which this quantity attains its absolute minimum. For this purpose, we may rewrite (2.5) as

$$\begin{aligned}P_{ABab}-P^1_{A-1,Bab}-P^2_{ABa,b-1}=\delta_{ABab,0000}&+q_A\left(P^1_{AB,a-1,b}-P^1_{A-1,Bab}\right)\\&+p_B\left(P^2_{A,B-1,ab}-P^2_{ABa,b-1}\right)\end{aligned} \tag{10.1}$$

and carry out an expansion with respect to q_A and p_B, not trivial due to the nonconstant coefficients implicit in P^1 and P^2. In the no-bias case, $e=f=0$, we have already analyzed two lines in the (q_A,q_B) plane: (8.17)

expands out to

$$C_\infty(0, p_B) = (1 - p_B)\left(\tfrac{1}{2} + \tfrac{3}{4} p_B + \tfrac{11}{8} p_B^2 + \tfrac{43}{16} p_B^3 + \cdots\right) \tag{10.2}$$

and (9.19) to

$$C_\infty(q_A, 0) = (1 - q_A)\left(\tfrac{1}{2} + \tfrac{5}{4} q_A + \tfrac{11}{8} q_A^2 + \cdots\right) \tag{10.3}$$

The cross-terms however must be picked up by laborious application of (10.1).

After a fair amount of routine algebra, (10.1) with boundary conditions (10.2), (10.3) can be expanded through second order, resulting in

$$C_\infty(q_A, p_B) = (p_A - p_B)\left(\tfrac{1}{2} + \tfrac{3}{4} p_B + \tfrac{5}{4} q_A + \tfrac{11}{8}\left(p_B^2 + q_A^2\right) + \tfrac{21}{4} p_B q_A + \cdots\right) \tag{10.4}$$

Even at this early stage in the expansion, significant information can be extracted. For this purpose, we use a Padé-type rational fraction representation, most easily done by introducing an expansion parameter γ: $p_B \to \gamma p_B$, $q_A \to \gamma q_A$ and then converting the γ-series to a continued fraction:

$$1 + a\gamma + b\gamma^2 + \cdots = 1 + \frac{a\gamma}{1 - (b/a)\gamma} + \cdots = \frac{\left[a + (a^2 - b)\gamma + \ldots\right]}{(a - b\gamma \cdots)}$$

In the present case, this yields

$$C_\infty(q_A, p_B) = \frac{1}{2}(p_A - p_B)\,\frac{\tfrac{3}{2} p_B + \tfrac{5}{2} q_A - \tfrac{1}{2} p_B^2 + \tfrac{7}{2} q_A^2 - 3 p_B q_A \cdots}{\tfrac{3}{2} p_B + \tfrac{5}{2} q_A - \tfrac{11}{4} p_B^2 - \tfrac{11}{4} q_A^2 - \tfrac{21}{2} p_B q_A \cdots} \tag{10.5}$$

which, taken literally, implies that the transition curve, on which C_∞ first becomes infinite, is given by

$$\tfrac{3}{2} p_B + \tfrac{5}{2} q_A - \tfrac{11}{4} p_B^2 - \tfrac{11}{4} q_A^2 - \tfrac{21}{2} p_B q_A \cdots = 0 \tag{10.6}$$

In fact, this curve, given in Fig. 3, is indistinguishable from that obtained[4] from a numerical simulation involving 100 walks of 1000 steps at each point of the (p_A, p_B) grid with grid spacing 0.1.

Numerical simulations have also been carried out[4] for a number of walks with nonzero bias and are responsible for the typical phase diagram of Fig. 4. However, even the series expansion technique above becomes rather involved for such cases. Although continued investigations of this kind have obvious importance in the context of clinical trials, they seem not at this stage to add substantially to the conceptual development of phase transitions in random walks. The prime necessity is rather that of constructing simple models which illustrate the facets of the phenomenon more

completely and which aid in the development of more powerful approximation techniques for their analysis. We intend to proceed in this direction.

REFERENCES

1. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I (John Wiley, New York, 1968).
2. Symposium on Random Walks, NBS, June 1982, to be published in *J. Stat. Phys.* **30** (1983).
3. H. E. Robbins, "Some aspects of sequential design of experiments," *Bull. Am. Math. Soc.* **58**:527–535 (1952).
4. O. E. Percus and J. K. Percus, "Ethical procedures in comparative evaluation of drugs," Courant Mathematics and Computing Laboratory Rep., New York University, New York (1980).
5. J. A. Bather, "Randomized allocation of treatment in sequential experiments," *J. Royal Stat. Soc.* (*B*) 265–292 (1981).
6. Y. S. Chow, H. E. Robbins, and D. Siegmund, *Great Expectation: The Theory of Optimal Stopping* (Houghton-Mifflin, New York, 1971).
7. M. L. Aggarwal, "Records and oscillating random walk," *Calcutta Stat. Assoc. Bull.* **25**:55–64 (1977).
8. O. E. Percus and J. K. Percus, "One dimensional random walk with phase transition," *SIAM J. Appl. Math.* **40**:485–497 (1981).
9. For numerical work, see R. E. Bechhofer and R. V. Kulkarni, "Closed adaptive sequential procedures for selecting the best of $k \geqslant 2$ Bernoulli populations," College of Engineering, Cornell Univ. Report (1981).
10. For analytical treatment, see O. E. Percus and J. K. Percus, to be published.
11. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, (Cambridge Univ. Press, Cambridge, 1962).