Reprinted from THEORETICAL POPULATION BIOLOGY
All Rights Reserved by Academic Press, New York and London

Vol. 45, No. 1, February 1994
Printed in Belgium

Barrier Transitions Driven by Fluctuations, with Applications to Ecology and Evolution

MARC MANGEL

Section of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616

Received September 15, 1992

Questions that involve "barrier crossings" arise in both ecology and evolution. When two species compete, and the population dynamics have multiple equilibria, a barrier crossing corresponds to fluctuations driving the populations across a deterministic separatrix. Similarly, fluctuation may drive quantitative traits across valleys in the surface of individual fitness, particularly when the fluctuations interact with a deterministic force such as a correlated selection reponse. In this paper, I show how to formulate and solve the barrier crossing problem. Two classes of models are considered. In models based on stochastic differential or difference equations, fluctuations are superimposed upon a system of deterministic differential or difference equations. In the case of a Markov chain, there may be no underlying deterministic system, so that the system is only characterized by Markovian transition probabilities. The validity of the diffusion approximation and the origin of macroscopic population dynamics from a underlying purely probabilistic system are investigated.

INTRODUCTION

Questions that involve "barrier crossings" arise in both ecology and evolution. Thomas Park's classic experiments (Neyman, Park, and Scott, 1956) illustrate the ecological setting. Park and his colleagues studied the processes underlying competition between flour beetles, at individual and group levels (Neyman, Park, and Scott, 1956, pp. 48 ff). Six different combinations of temperature and humidity were used to construct six different "ecologies" for the flour beetles. Single species persisted in essentially all six ecologies, with equilibrium levels of about 100 adults. However, in competition between two species, one species always persisted and the other was always eliminated, but the results of the competition were sometimes an indeterminate function of initial conditions. That is, for some sets of initial conditions, species 1 (e.g., *Triboluium confusum*) always won the competition; for other sets of initial conditions, species 2 (e.g., *Tribolium castaneum*) always won the competition; but there was a range of initial

0040-5809/94 $6.00

conditions in which species 1 won the competition only some of the time. Neyman, Park, and Scott (1956) refer to these as determinate and indeterminate regions.[1] The results are consistent with a model of stochastic competition (Fig. 1) in which population fluctuations are superimposed upon classical, deterministic equations of competition. That is, one envisions an underlying deterministic competition described by a pair of differential equations

$$\frac{dX}{dt} = b_1(X, Y) \qquad \frac{dY}{dt} = b_2(X, Y) \tag{1a}$$

or difference equations

$$\begin{aligned} X(t+1) &= X(t) + b_1(X(t), Y(t)) \\ Y(t+1) &= Y(t) + b_2(X(t), Y(t)). \end{aligned} \tag{1b}$$

Here $X(t)$ and $Y(t)$ are the population sizes at time t. The system (1) is assumed to have three equilibria. Two of them are stable (one of the two species extinct) and the third is a saddle point (Fig. 1). The population trajectory that enters the saddle point is called the separatrix. It separates the plane into two "domains of attraction": Points on one side of the separatrix are deterministically attracted to the equilibrium in which $X(t) = 0$; points on the other side to the equilibrium in which $Y(t) = 0$. When fluctuations are present, initial conditions that are deterministically attracted to $X(t) = 0$ may end up with $Y(t) = 0$ and vice-versa. This event is a "barrier crossing." If the intensity of the noise is small, then we expect

[1] In a provocative and thoughtful essay, Simberloff (1980) points out that 1859 (the year of publication of *The Origin of Species*) was the watershed year for the "revolution against determinism": not only did Darwin publish his seminal work, but Maxwell published the kinetic gas law. These are the foundations for non-deterministic thinking in biology and physics. Park's work had many goals, but one high priority was the attempt to explain the probabilistic outcome of competition between different species of flour beetles in terms of physical factors (e.g., temperature and humidity) and their relation to biological factors (e.g., fecundity, cannibalism rate). Simberloff describes the attempts to demonstrate that Park's results were illusory; but in the end the indeterminacy could not be eliminated by better controls (e.g., genetic uniformity). The stochastic nature of the competition is real and "our understanding of population phenomena will require stochastic treatment" (p. 21). Simberloff identifies three levels of indeterminacy: (i) molecular indeterminacy (e.g., random mutations in the genome of the flour beetle), (ii) chaotic behavior of non-linear systems (e.g., the flour beetle dynamics are deterministic, but so non-linear that they are chaotic), and (iii) the indeterminacy engendered by the enormous number of entities in the ecological system (e.g., each species has eggs, larvae, pupae, and adults interacting with complex behaviors). Simberloff concludes that the probabilistic revolution is far from complete and that the "variability [that] typifies ecological systems because of their complexity ··· suggests different mathematical approaches and criteria for success" (p. 27).

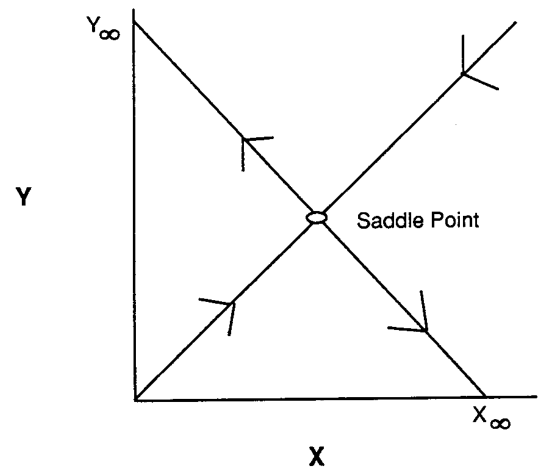

FIG. 1. An underlying deterministic phase portrait that gives rise to a region of indeterminacy in stochastic competition. We assume that the competition equations have an unstable equilibrium (saddle point). The separatrix trajectory $\mathscr{S}$ divides the phase plane into two domain of attraction. Deterministically, all points in one region tend towards $(0, Y_\infty)$ and all points in the other region tend towards $(X_\infty, 0)$. When fluctuations are present, initial conditions near the separatrix may lead to indeterminate results, since fluctuations may force trajectories across the separatrix into the opposite domain of attraction.

that points near the separatrix will be most sensitive to fluctuations. The indeterminate region described by Neyman, Park, and Scott (1956) is thus the vicinity of the separatrix. Mangel and Ludwig (1977) showed how one can compute the probability that fluctuations cause a point which is deterministically attracted to one domain of attraction to end up at the other domain of attraction. The objective of the present paper is to explain the method of Mangel and Ludwig (1977) so that it is accessible to individuals not expert in asymptotic analysis, mathematical physics, and the writing style associated with the Courant Institute of Mathematical Sciences (Courant and Hilbert, 1962).

Price *et al.* (1992) recently proposed a model for the evolution of correlated characters in a changing environment (Fig. 2). According to this model, we envision two correlated traits, X and Y, which are near one of two peaks in the fitness surface of individual fitness. After an environmental change, X is subject to directional selection. As a correlated response, Y may be driven from one fitness peak to the other. When genetic drift is taken into account, the stochastic fluctuations may produce a peak shift when it is not predicted in the deterministic model, or may prevent it when it would occur deterministically. (Crossing this valley of fitness is equivalent to crossing the deterministic separatrix in the ecological setting.)

We wish to compute the probability that fluctuations cause a shift from one outcome to another. This involves (i) specifying an underlying model,

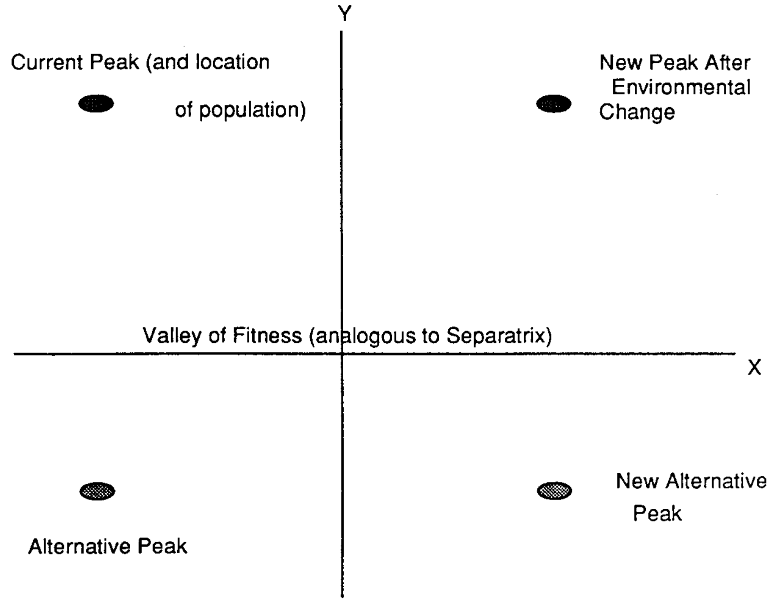

FIG. 2. The model of Price *et al.* (1991) for the evolution of correlated characters. We envision two correlated traits, X and Y, currently at a peak of fitness but for which another peak exists. After an environmental change, both peaks shift and selection on the X-trait drives the population towards the new peak. When fluctuations are present, it is possible that the population may cross the valley of fitness for the Y trait, on its way to the new peak for the X-trait. Crossing this valley of fitness is equivalent to crossing the deterministic separatrix in the ecological setting.

(ii) deriving the equations for the probability of outcome, given initial conditions, and (iii) solving those equations. This program is carried out in the remainder of the paper.

In the next section, models based on stochastic differential equations, stochastic difference equations, and a Markov chain are introduced. In addition, a "practice" problem is solved. The solution of this problem, which involves only one state variable, indicates the form of the general solution. The method of general solution is summarized and explained in more detail than presented in the extremely terse version of Mangel and Ludwig (1977). Similar ideas have also recently been used by Barton and Rouhani (1987) and Rouhani and Barton (1987a, b) to calculate the expected times associated with transitions, rather than probabilities. In the fourth section, two simple examples of the method are illustrated. In the fifth section, the method is extended to models in which there is no underlying deterministic system. A thorough study of the results illustrates why the diffusion approximation is appropriate in this case, and how macroscopic population dynamics such as (1) arise from an underlying Markov chain model.

Models

Stochastic Differential and Difference Equations

Consider two populations or two traits $X(t)$ and $Y(t)$ that satisfy the following stochastic differential equations:

$$\begin{aligned} dX &= b_1(X, Y)\,dt + \sqrt{\varepsilon a_{11}(X, Y)}\,dW_1 + \sqrt{\varepsilon a_{12}(X, Y)}\,dW_2 \\ dY &= b_2(X, Y)\,dt + \sqrt{\varepsilon a_{12}(X, Y)}\,dW_1 + \sqrt{\varepsilon a_{22}(X, Y)}\,dW_2. \end{aligned} \tag{2a}$$

In this equation, $dX = X(t+dt) - X(t)$, $dY = Y(t+dt) - Y(t)$, and the $b_i(X, Y)$ are the deterministic components of either selection or population dynamics. The intensity of fluctuations is measured by the variance–covariance matrix with components $\varepsilon a_{ij}(X, Y)$, so that $a_{12}(X, Y) = a_{21}(X, Y)$, and the dW_i are the increments in two independent Wiener processes: each dW_i has independent increments and is normally distributed with mean 0 and variance dt. The parameter ε is a measure of the intensity of fluctuations. In most settings, it is inversely proportional to a measure of population size. According to the model (2a), given that $X(t) = x$, $Y(t) = y$, dX is normally distributed with mean $b_1(x, y)\,dt + o(dt)$ and variance $[\varepsilon a_{11}(x, y) + \varepsilon a_{12}(x, y)]\,dt + o(dt)$, where $o(dt)$ denotes terms such that $\lim_{dt \to 0}(o(dt)/dt) = 0$. Similarly, dY is normally distributed with mean $b_2(x, y)\,dt + o(dt)$ and variance $[\varepsilon a_{12}(x, y) + \varepsilon a_{22}(x, y)]\,dt + o(dt)$. All higher moments are $o(dt)$.

The analog of (2a) for difference equations is

$$\begin{aligned} X(t+1) &= X(t) + b_1(X(t), Y(t)) + \xi_1(X(t), Y(t)) \\ Y(t+1) &= Y(t) + b_2(X(t), Y(t)) + \xi_2(X(t), Y(t)), \end{aligned} \tag{2b}$$

where the $\xi_i(X(t), Y(t))$ are noise terms analogous to the Brownian motion processes. The key point, which is essential for the analysis that follows, is that for either (2a) or (2b), increments in $X(t)$ and $Y(t)$ have conditional distributions that are normal.

A Markov Chain Model

It is, of course, possible that there are no underlying deterministic dynamics and that the dynamics of the system of interest are based solely on transition probabilities. In that case, we use a discrete time and state description in which changes in $X(t)$ and $Y(t)$ are summarized by a transition probability

$$\rho_{ij}(x, y) = \text{Prob}\{X(t+1) = x + \varepsilon i,\ Y(t+1) = y + \varepsilon j \mid X(t) = x,\ Y(t) = y\}. \tag{3}$$

The mean change in X is then $\Delta X = X(t+1) - X(t) = \sum_i \sum_j \rho_{ij}(X, Y)\, \varepsilon i$ and the mean change in Y is then $\Delta Y = Y(t+1) - Y(t) = \sum_i \sum_j \rho_{ij}(X, Y)\, \varepsilon j$, where the summations are taken over all possible values of i and j. Unlike the case of the stochastic differential equation model, in which third and higher moments are $o(dt)$, all moments of (3) may be of the same order.

The model (3) is substantially different from (2) in that there is no underlying deterministic system upon which fluctuations are superimposed. However, the same methods derived for the stochastic differential and difference equations can be applied here.

A "Practice" Problem and Its Exact Solution

The main ideas of the solution technique can be illustrated with a practice problem that is exceptionally simple (Mangel and Ludwig, 1977). In only one dimension, the separatrix collapses to a point, so consider the model

$$dX = X\, dt + \sqrt{2\varepsilon}\, dW \tag{4}$$

with X restricted to the range $[-1, 1]$. The underlying deterministic dynamics are then $dX/dt = X$ and if $X(0) < 0$, the system is deterministically attracted to $X = -1$, whereas if $X(0) > 0$, the system is deterministically attracted to $X = 1$.

When fluctuations are included, a trajectory starting from an initial condition that is attracted to $X = -1$ might end up at $X = 1$ (and vice versa). It is exactly the probability of this event that we wish to compute:

$$u(x) = \text{Prob}\{X(t) \text{ reaches 1 before } -1\text{, given that } X(0) = x\}. \tag{5}$$

This function is the analog of the probability of fixation in population genetics (Ewens, 1979) or the colonization probability in ecology (MacArthur and Wilson, 1967). We can derive an equation for $u(x)$ by applying the law of total probability (Mangel and Clark, 1988) to $u(x)$,

$$u(x) = E_{dX}\{u(x + dX)\}, \tag{6}$$

where E_{dX} denotes the expectation taken using the distribution of dX determined by (4). Taylor expanding $u(x + dX)$ in powers of dX gives

$$\begin{aligned} u(x) &= E_{dX}\left\{u(x) + dX \frac{du}{dx} + \frac{1}{2}(dX)^2 \frac{d^2u}{dx^2} + \cdots\right\} \\ &= u(x) + E_{dX}\{dX\} \frac{du}{dx} + \frac{1}{2} E_{dX}\{(dX)^2\} \frac{d^2u}{dx^2} + \cdots \end{aligned} \tag{7}$$

The terms ignored in this expansion are proportional to $(dX)^3$, $(dX)^4$, etc. As described above, they are $o(dt)$. According to (4) the mean of dX is $X\,dt+o(dt)$, the variance of dX is $\varepsilon\,dt+o(dt)$, and all higher moments are $o(dt)$. Thus, taking the average in (7) gives

$$u(x)=u(x)+x\,dt\,\frac{du}{dx}+\frac{1}{2}(2\varepsilon)\,dt\,\frac{d^2u}{dx^2}+o(dt), \tag{8}$$

so that dividing by dt and taking the limit as $dt\to 0$ gives the equation

$$x\frac{du}{dx}+\varepsilon\frac{d^2u}{dx^2}=0. \tag{9}$$

This equation requires two boundary conditions. If the value of $x=1$, then the probability of reaching $X=1$ before $X=-1$ is exactly 1; if $x=-1$, this probability is 0. Hence $u(1)=1$, $u(-1)=0$. The solution of (9) corresponding to these boundary conditions is

$$u(x)=\frac{\int_{-1}^{x}\exp(-y^2/2\varepsilon)\,dy}{\int_{-1}^{1}\exp(-y^2/2\varepsilon)\,dy}. \tag{10}$$

The denominator in this expression is simply a constant; introducing a change of variables $v=y/\sqrt{\varepsilon}$ allows us to rewrite (10) as

$$u(x)=\frac{\int_{-1/\sqrt{\varepsilon}}^{x/\sqrt{\varepsilon}}\exp(-v^2/2)\,dv}{\int_{-1/\sqrt{\varepsilon}}^{1/\sqrt{\varepsilon}}\exp(-v^2/2)\,dv} \tag{11}$$

When the intensity of fluctuations ε is small, the lower limit of integration in (11) $-1/\sqrt{\varepsilon}\ll -1$, the upper limit in the demoninator $1/\sqrt{\varepsilon}\gg 1$ and we can approximate (11) by

$$u(x)\sim g_0\Phi\left(\frac{x}{\sqrt{\varepsilon}}\right), \tag{12}$$

where g_0 is a constant and $\Phi(z)$ is the cumulative normal distribution function

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{u^2}{2}\right)du. \tag{13}$$

In this case, if ε is sufficiently small, g_0 is essentially 1; otherwise it is simply the normalization constant in (11).

A More General One Dimensional Problem and Its Approximate Solution

Suppose we consider a more general but still one dimensional problem,

$$dX = b(X)\,dt + \sqrt{2\varepsilon a(X)}\,dW, \tag{14}$$

with $b(0) = 0$ and $b'(0) > 0$, and $a(x) > 0$. We define $u(x)$ as above and follow the procedure that led to (9) to obtain

$$b(x)\frac{dx}{dx} + \varepsilon a(x)\frac{d^2u}{dx^2} = 0. \tag{15}$$

The boundary conditions for (15) are the same as those for (9) and it is possible to solve (15) by integration, just as (9) was solved. The result is only slightly more complicated:

$$u(x) = \int_{-1}^{x} \exp\left(-\int^{x'} \frac{b(y)}{\varepsilon a(y)}\,dy\right) dx' \Bigg/ \int_{-1}^{1} \exp\left(-\int^{x'} \frac{b(y)}{\varepsilon a(y)}\,dy\right) dx'. \tag{16}$$

An alternative procedure, however, is to obtain an approximate solution, using (9) as a guide. That is, let us seek a solution of (15) in the form

$$u(x) \sim g_0 \Phi\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right), \tag{17}$$

where $\psi(x)$ is an unknown function which must be determined and g_0 is a constant. In order to evaluate the derivatives of $u(x)$, we note from (13) that

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{18}$$

so that $\Phi''(z) = -z\Phi'(z)$. Thus, if $u(x)$ is given by (17) we have

$$\begin{aligned}
\frac{du}{dx} &= g_0 \Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right) \frac{1}{\sqrt{\varepsilon}} \frac{d\psi(x)}{dx} \\
\frac{d^2u}{dx^2} &= g_0 \Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right) \frac{1}{\sqrt{\varepsilon}} \frac{d^2\psi(x)}{dx^2} + g_0 \Phi''\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right) \left\{\frac{1}{\sqrt{\varepsilon}} \frac{d\psi(x)}{dx}\right\}^2 \\
&= g_0 \Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right) \frac{1}{\sqrt{\varepsilon}} \frac{d^2\psi(x)}{dx^2} - g_0 \Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right) \frac{\psi(x)}{\sqrt{\varepsilon}} \left\{\frac{1}{\sqrt{\varepsilon}} \frac{d\psi(x)}{dx}\right\}^2
\end{aligned} \tag{19}$$

(the last line follows from (18)). We now substitute these into (15) to obtain

$$
\begin{aligned}
b(x)&\left[g_0\Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right)\frac{1}{\sqrt{\varepsilon}}\frac{d\psi(x)}{dx}\right]\\
&+\varepsilon a(x)\left[g_0\Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right)\frac{1}{\sqrt{\varepsilon}}\frac{d^2\psi(x)}{dx^2}\right.\\
&\left.-g_0\Phi'\left(\frac{\psi(x)}{\sqrt{\varepsilon}}\right)\frac{\psi(x)}{\sqrt{\varepsilon}}\left\{\frac{1}{\sqrt{\varepsilon}}\frac{d\psi(x)}{dx}\right\}^2\right]\\
&=0
\end{aligned}
\tag{20}
$$

If we collect the terms in (20) according to powers of ε, the most important term involves $1/\sqrt{\varepsilon}$ and the coefficient of $1/\sqrt{\varepsilon}$ vanishes if we set

$$b(x)\frac{d\psi(x)}{dx}-a(x)\,\psi(x)\left\{\frac{d\psi(x)}{dx}\right\}^2=0. \tag{21}$$

At the present level of approximation we cannot set the coefficient of $\sqrt{\varepsilon}$ equal to 0. To do this, we must modify the initial guess (17) and replace it by $u(x)\sim g_0\Phi(\psi(x)/\sqrt{\varepsilon})+\sqrt{\varepsilon}\,h_0(x)\,\Phi'(\psi(x)/\sqrt{\varepsilon})$, where the function $h_0(x)$ is to be determined (Mangel and Ludwig, 1977, show how to do this). However, when $\varepsilon \ll 1$, the equation for $u(x)$ is closely approximated by (17) if (21) is satisfied.

To satisfy (21), either $d\psi(x)/dx=0$, which we reject because then $\psi(x)$ is a constant, or

$$b(x)=a(x)\,\psi(x)\frac{d\psi(x)}{dx}, \tag{22}$$

from which we conclude that

$$\frac{d}{dx}\left(\frac{1}{2}\psi(x)^2\right)=\frac{b(x)}{a(x)}, \tag{23}$$

so that

$$\psi(x)^2=2\int^x\frac{b(y)}{a(y)}\,dy \tag{24}$$

and

$$\psi(x)=\pm\sqrt{2\int^x\frac{b(y)}{a(y)}\,dy}. \tag{25}$$

Both the + and the − are needed: for $x > 0$ we choose + and for $x < 0$ we choose −, to ensure that $u(x)$ has the appropriate behavior (approaches 1 as $x \to 1$ and 0 as $x \to -1$).

We thus have constructed the solution

$$u(x) \sim g_0 \Phi\left(\pm \sqrt{2 \int^x \frac{b(y)}{a(y)}\, dy} \Big/ \sqrt{\varepsilon} \right), \tag{26}$$

where the constant g_0 is chosen for normalization. If we choose $b(x) = x$ and $a(y) = 1$, this procedure leads to the exact solution we computed in the previous section. A similar procedure is used in the two dimensional case, except that the details are more complicated.

THE MULTIDIMENSIONAL CASE: A GENERAL APPROXIMATION

Mangel and Ludwig (1977) developed methods, analogous to those used above, for coupled stochastic differential equations, such as those which occur in stochastic competition or in the evolution of correlated characters. In this section, that work is summarized and the details are made more explicit, so that readers will be able to use the method. For simplicity, only two species or two traits are considered. However, the methods generalize readily for more than two species without any conceptual difficulty and few technical difficulties, except for problem formulation. For some special

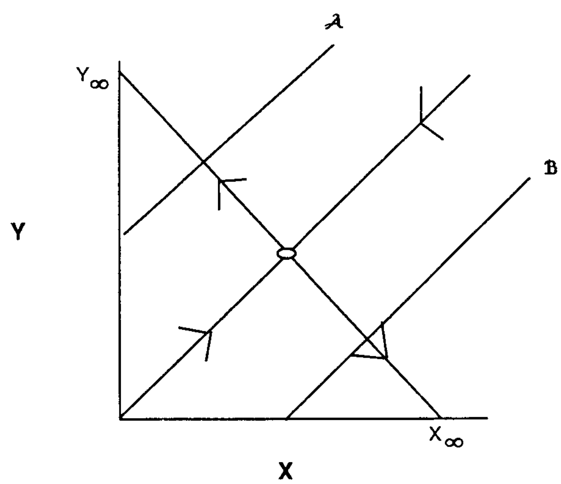

FIG. 3. Boundary conditions for the barrier transition problem. We wish to compute the probability that a system starting somewhere between $\mathscr{A}$ and $\mathscr{B}$ crosses $\mathscr{B}$ before $\mathscr{A}$. The expansion method (described in the Appendix) can be used to find $u(x, y)$ for points (x, y) that are close to the separatrix.

forms ("potentials") of the mean changes in X and Y, it is possible to obtain exact results (e.g., Barton and Rouhani (1987) and Rouhani and Barton (1987a, b)); the emphasis here is on methods that are generally applicable.

Using the language of stochastic competition, we are now interested in

$$u(x, y) = \text{Prob}\{\text{species } X \text{ wins a competition with species } Y, \text{ given } X(0)=x,\ Y(0)=y\}. \tag{27}$$

To make this more precise, surround the separatrix by a band (Fig. 3) and define

$$u(x, y) = \text{Prob}\{\text{the trajectory } (X(t), Y(t)) \text{ crosses } \mathscr{A} \text{ before } \mathscr{B}, \text{ given that } X(0)=x,\ Y(0)=y\}. \tag{28}$$

This probability is the solution of an equation analogous to (9),

$$0 = \frac{\varepsilon}{2}\left\{a_{11}(x, y)\frac{\partial^2 u}{\partial x^2} + 2a_{21}(x, y)\frac{\partial^2 u}{\partial x\,\partial y} + a_{22}(x, y)\frac{\partial^2 u}{\partial y^2}\right\} + b_1(x, y)\frac{\partial u}{\partial x} + b_2(x, y)\frac{\partial u}{\partial y}, \tag{29}$$

with associated boundary conditions $u(x, y) = 1$ for points (x, y) on the curve $\mathscr{A}$ and $u(x, y) = 0$ for points (x, y) on the curve $\mathscr{B}$.

We now seek a solution of the form

$$u(x, y) = g_0 \Phi\left(\frac{\psi(x, y)}{\sqrt{\varepsilon}}\right) + O(\sqrt{\varepsilon}), \tag{30}$$

where g_0 is a constant (determined by normalization) and $O(\sqrt{\varepsilon})$ denotes that the next term in the approximate solution is proportional to $\sqrt{\varepsilon}$. The function $\psi(x, y)$ satisfies the following equation (Mangel and Ludwig, 1977) analogous to (20):

$$b_1(x, y)\frac{\partial \psi}{\partial x} + b_2(x, y)\frac{\partial \psi}{\partial y} - \frac{1}{2}\left\{a_{11}(x, y)\,\psi(x, y)\left(\frac{\partial \psi}{\partial x}\right)^2 + 2a_{12}(x, y)\,\psi(x, y)\frac{\partial \psi}{\partial x}\frac{\partial \psi}{\partial y} + a_{22}(x, y)\,\psi(x, y)\left(\frac{\partial \psi}{\partial y}\right)^2\right\} = 0. \tag{31}$$

If we define a new function

$$\theta(x, y) = -\tfrac{1}{2}\psi(x, y)^2 \tag{32}$$

then (31) becomes

$$b_1(x, y)\frac{\partial\theta}{\partial x} + b_2(x, y)\frac{\partial\theta}{\partial y} + \frac{1}{2}\left\{a_{11}(x, y)\left(\frac{\partial\theta}{\partial x}\right)^2 + 2a_{12}(x, y)\frac{\partial\theta}{\partial x}\frac{\partial\theta}{\partial y}\right.$$

$$\left. + a_{22}(x, y)\left(\frac{\partial\theta}{\partial y}\right)^2\right\} = 0. \tag{33}$$

Equation (33) can be solved by the method of characteristics (Courant and Hilbert, 1962, p. 75), which is a classical method of applied mathematics. By doing this, we have converted the second order, linear equation for $u(x, y)$ to a first order, nonlinear equation for $\theta(x, y)$. In addition to the solution by characteristics, (33) has an associated Hamiltonian, Lagrangian, and variational principle, all of which are now explained (and are important for the biological interpretation of the equation).

The Hamilton associated with (33) is

$$H(x, y, p, q) = \tfrac{1}{2}\{a_{11}(x, y)\,p^2 + 2a_{12}(x, y)\,pq + a_{22}(x, y)\,q^2\}$$

$$+ b_1(x, y)\,p + b_2(x, y)q. \tag{34}$$

Note that (34) is the same as (33), except that the derivatives of $\theta(x, y)$ have been replaced by two new variables, p and q. The essense of the Hamilton–Jacobi variational theory is to treat these quantities as independent variables. The Hamiltonian generates trajectories according to the differential equations (Courant and Hilbert, 1962, pp. 97 ff.)

$$\begin{aligned}
\frac{dx}{ds} &= \frac{\partial H}{\partial p} \\
\frac{dy}{ds} &= \frac{\partial H}{\partial q} \\
\frac{dp}{ds} &= -\frac{\partial H}{\partial x} \\
\frac{dq}{ds} &= -\frac{\partial H}{\partial y}.
\end{aligned} \tag{35}$$

In this equation "s" parameterizes the solution curves (or characteristics). Along these solution curves, $\theta(x, y)$ satisfies the ordinary differential equation (Courant and Hilbert, 1962)

$$\frac{d\theta}{ds} = \frac{1}{2}\left[a_{11}p^2 + 2a_{12}pq + a_{22}q^2\right]. \tag{36}$$

The Lagrangian $L(x, y, dx/ds, dy/ds)$ is (Rund, 1973, pp. 67 ff.)

$$H(x, y, p, q) + L\left(x, y, \frac{dx}{ds}, \frac{dy}{ds}\right) = p\frac{dx}{ds} + q\frac{dy}{ds}, \tag{37}$$

so that in this case, the Lagrangian is

$$\begin{aligned} L\left(x, y, \frac{dx}{ds}, \frac{dy}{ds}\right) = &\frac{1}{2}\bar{a}_{11}(x, y)\left(b_1(x, y) - \frac{dx}{ds}\right)^2 \\ &+ \bar{a}_{12}(x, y)\left(b_1(x, y) - \frac{dx}{ds}\right)\left(b_2(x, y) - \frac{dy}{ds}\right) \\ &+ \frac{1}{2}\bar{a}_{22}(x, y)\left(b_2(x, y) - \frac{dy}{ds}\right)^2, \end{aligned} \tag{38}$$

where the matrix $\bar{a}(x, y)$ is the inverse of the matrix $a(x, y)$. The important reason for studying the Lagrangian is that according to Hamilton–Jacobi theory, $\theta(x, y)$ is the minimum of the integral of the Lagrangian taken over all paths joining the point (x, y) and a given initial point (x_0, y_0). If this point (x_0, y_0) is chosen to be the saddle point (Fig. 3), then the Lagrangian vanishes on the separatrix since the separatrix is the solution of the equations $dx/ds = b_1(x, y)$ and $dy/ds = b_2(x, y)$; this means that $\theta(x, y) = 0$ on the separatrix and consequently $\psi(x, y) = 0$ on the separatrix. One can now find $\theta(x, y)$ by integrating the Hamilton–Jacobi equations. In actual fact, as shown in the next section, we can often escape the numerical solution of the Hamilton–Jacobi equations by an expansion of (33) near the separatrix.

An Expansion Method Near the Separatrix

In this section, I now show how (33) can be expanded near the separatrix and then provide two examples showing how the expansion method is used. Both of these turn out to be exceedingly simple, but they illustrate the point.

A General Method for Expansion Near the Separatrix

A method for determining $\theta(x, y)$ near the separatrix by an expansion technique is tersely described in Ludwig (1975) and Mangel and Ludwig (1977); the description here makes it more easily accessible. The starting point is (33) and the assumption that the functions $b_i(x, y)$ and $a_{ij}(x, y)$ are sufficiently differentiable in the region around the separatrix. On the deterministic separatrix, we have $\psi(x, y) = 0$, and since $\theta(x, y) = -\frac{1}{2}\psi^2(x, y)$, $\theta(x, y)$ and its derivatives with respect to x and y are 0 on the separatrix.

Let us thus consider a point (x_s, y_s) on the separatrix and another point (x, y) near, but not on, the separatrix. The distance between the two points is determined by $\delta x = x - x_s$ and $\delta y = y - y_s$. We write

$$\begin{aligned}\frac{\partial\theta(x, y)}{\partial x} &= \frac{\partial^2\theta(x_s, y_s)}{\partial x^2}\,\delta x + \frac{\partial^2\theta(x_s, y_s)}{\partial x\,\partial y}\,\delta y \\ &\quad + \frac{1}{2}\left[\frac{\partial^3\theta(x_s, y_s)}{\partial x^3}(\delta x)^2 + 2\frac{\partial^3\theta(x_s, y_s)}{\partial x^2\,\partial y}\,\delta x\,\delta y + \frac{\partial^3\theta(x_s, y_s)}{\partial x\,\partial y^2}(\delta y)^2\right] \\ &\quad + \cdots \end{aligned} \tag{39}$$

and

$$\begin{aligned}\frac{\partial\theta(x, y)}{\partial y} &= \frac{\partial^2\theta(x_s, y_s)}{\partial x\,\partial y}\,\delta x + \frac{\partial^2\theta(x_s, y_s)}{\partial y^2}\,\delta y \\ &\quad + \frac{1}{2}\left[\frac{\partial^3\theta(x_s, y_s)}{\partial x^2\,\partial y}(\delta x)^2 + 2\frac{\partial^3\theta(x_s, y_s)}{\partial x\,\partial y^2}\,\delta x\,\delta y + \frac{\partial^3\theta(x_s, y_s)}{\partial y^3}(\delta y)^2\right] \\ &\quad + \cdots. \end{aligned}$$

Similarly, we can expand $b_1(x, y)$ and $b_2(x, y)$ as

$$\begin{aligned} b_1(x, y) &= b_1(x_s, y_s) + \left[\frac{\partial}{\partial x} b_1(x_s, y_s)\right]\delta x + \left[\frac{\partial}{\partial y} b_1(x_s, y_s)\right]\delta y + \cdots \\ b_2(x, y) &= b_2(x_s, y_s) + \left[\frac{\partial}{\partial x} b_2(x_s, y_s)\right]\delta x + \left[\frac{\partial}{\partial y} b_2(x_s, y_s)\right]\delta y + \cdots. \end{aligned} \tag{40}$$

As will be seen, it is only necessary to know $a(x, y)$ on the separatrix. We now substitute (39) and (40) into (33). There are terms proportional to δx, δy, $(\delta x)^2$, $\delta x\,\delta y$, and $(\delta y)^2$; since the increments in x and y are arbitrary, we require that the coefficients of these increments vanish. The terms that multiply δx in the expansion are

$$b_1(x_s, y_s)\frac{\partial^2\theta}{\partial x^2} + b_2(x_s, y_s)\frac{\partial^2\theta}{\partial x\,\partial y} \tag{41}$$

and the terms that multiply δy are

$$b_1(x_s, y_s)\frac{\partial^2\theta}{\partial x\,\partial y} + b_2(x_s, y_s)\frac{\partial^2\theta}{\partial y^2}. \tag{42}$$

Both of these terms are identically 0. To see that this is true, note that the expressions (41) and (42) can be rewritten as the product of vectors. If we let $\mathbf{t}$ denote the tangent vector along the separatrix, $\mathbf{t} = (b_1(x_s, y_s), b_2(x_s, y_s))$, and $\mathbf{Grad}(\omega)$ denote the gradient of a function ω, $\mathbf{Grad}(\omega) = (\partial\omega/\partial x, \partial\omega/\partial y)$, then the expression (41) is

$$b_1(x_s, y_s)\frac{\partial^2\theta}{\partial x^2} + b_2(x_s, y_s)\frac{\partial^2\theta}{\partial x\,\partial y} = \mathbf{t}\cdot\mathbf{Grad}\left(\frac{\partial\theta}{\partial x}\right), \tag{43}$$

where "$\cdot$" denotes the usual dot product between two vectors. Since $\partial\theta/\partial x = 0$ on the separatrix, we conclude that the expression in (43) is identically 0. Similarly, the expression in (42) can be written as $\mathbf{t}\cdot\mathbf{Grad}(\partial\theta/\partial y)$, from which we conclude that it too vanishes. Thus, the terms proportional to δx and δy are both identically zero. This means that the equation is approximately solved if we set the coefficients of the quadratic terms equal to 0. (Were this not the case, we would have to set the coefficients of the linear terms to 0.)

The next terms in the expansion are proportional to $(\delta x)^2$, $\delta x\,\delta y$, and $(\delta y)^2$. The coefficient of $(\delta x)^2$ is

$$\begin{aligned}
&\frac{1}{2}\left[a_{11}(x_s, y_s)\frac{\partial^2\theta}{\partial x^2}\frac{\partial^2\theta}{\partial x^2} + 2a_{12}(x_s, y_s)\frac{\partial^2\theta}{\partial x^2}\frac{\partial^2\theta}{\partial x\,\partial y}\right.\\
&\qquad\left. + a_{22}(x_s, y_s)\frac{\partial^2\theta}{\partial x\,\partial y}\frac{\partial^2\theta}{\partial x\,\partial y}\right]\\
&\qquad + \left[\frac{\partial b_1(x_s, y_s)}{\partial x}\frac{\partial^2\theta}{\partial x^2} + \frac{\partial b_2(x_s, y_s)}{\partial x}\frac{\partial^2\theta}{\partial x\,\partial y}\right]\\
&\qquad + \frac{1}{2}\left[b_1(x_s, y_s)\frac{\partial^3\theta}{\partial x^3} + b_2(x_s, y_s)\frac{\partial^3\theta}{\partial x^2\,\partial y}\right].
\end{aligned} \tag{44}$$

The coefficient of $\delta x\, \delta y$ is

$$\frac{1}{2} a_{11}(x_s, y_s) \frac{\partial^2\theta}{\partial x\, \partial y} \frac{\partial^2\theta}{\partial x^2}$$
$$+ a_{12}(x_s, y_s) \left[\frac{\partial^2\theta}{\partial x^2} \frac{\partial^2\theta}{\partial y^2} + \frac{\partial^2\theta}{\partial x\, \partial y} \frac{\partial^2\theta}{\partial x\, \partial y} \right]$$
$$+ \frac{1}{2} a_{22}(x_s, y_s) \frac{\partial^2\theta}{\partial x\, \partial y} \frac{\partial^2\theta}{\partial y^2}$$
$$+ \left[\frac{\partial b_1(x_s, y_s)}{\partial x} \frac{\partial^2\theta}{\partial x\, \partial y} + \frac{\partial b_2(x_s, y_s)}{\partial x} \frac{\partial^2\theta}{\partial y^2} \right]$$
$$+ \left[\frac{\partial b_1(x_s, y_s)}{\partial y} \frac{\partial^2\theta}{\partial x^2} + \frac{\partial b_2(x_s(x_s, y_s)}{\partial y} \frac{\partial^2\theta}{\partial x\, \partial y} \right]$$
$$+ \left[b_1(x_s, y_s) \frac{\partial^3\theta}{\partial x^2\, \partial y} + b_2(x_s, y_s) \frac{\partial^3\theta}{\partial x\, \partial y^2} \right]. \tag{45}$$

Finally, the coefficient of $(\delta y)^2$ is

$$\frac{1}{2} \left[a_{11}(x_s, y_s) \frac{\partial^2\theta}{\partial x\, \partial y} \frac{\partial^2\theta}{\partial x\, \partial y} \right.$$
$$\left. + 2a_{12}(x_s, y_s) \frac{\partial^2\theta}{\partial y^2} \frac{\partial^2\theta}{\partial x\, \partial y} + a_{22}(x_s, y_s) \frac{\partial^2\theta}{\partial y^2} \frac{\partial^2\theta}{\partial y^2} \right]$$
$$+ \left[\frac{\partial b_1(x_s, y_s)}{\partial y} \frac{\partial^2\theta}{\partial x\, \partial y} + \frac{\partial b_2(x_s, y_s)}{\partial y} \frac{\partial^2\theta}{\partial y^2} \right]$$
$$+ \frac{1}{2} \left[b_1(x_s, y_s) \frac{\partial^3\theta}{\partial x^2\, \partial y} + b_2(x_s, y_s) \frac{\partial^3\theta}{\partial y^3} \right]. \tag{46}$$

Setting the expressions (44)–(46) all equal to 0 provides a coupled set of three equations for the three unknowns $\partial^2\theta/\partial x^2$, $\partial^2\theta/\partial x\, \partial y$, and $\partial^2\theta/\partial y^2$. This set of equations can be written in a compact form (Mangel and Ludwig, 1977) by introducing matrices

$$P = \begin{pmatrix} \dfrac{\partial^2\theta}{\partial x^2} & \dfrac{\partial^2\theta}{\partial x\, \partial y} \\ \dfrac{\partial^2\theta}{\partial x\, \partial y} & \dfrac{\partial^2\theta}{\partial y^2} \end{pmatrix} \qquad B = \begin{pmatrix} \dfrac{\partial b_1(x, y)}{\partial x} & \dfrac{\partial b_1(x, y)}{\partial y} \\ \dfrac{\partial b_2(x, y)}{\partial x} & \dfrac{\partial b_2(x, y)}{\partial y} \end{pmatrix}$$

$$A = \begin{pmatrix} a_{11}(x, y) & a_{12}(x, y) \\ a_{12}(x, y) & a_{22}(x, y) \end{pmatrix}$$

in which case (44)–(46) can be written as

$$\frac{dP}{ds} + BP + PB^{\mathrm{T}} + PAP = 0. \tag{47}$$

In this equation, s measures "time" along the separatrix, with $s = \infty$ corresponding to the saddle point. Equation (47) is derived by noting that along the separatrix $dx_s/ds = b_1(x_s, y_s)$ and $dy_s/ds = b_2(x_s, y_s)$ so that the total derivative of any function ϑ along the separatrix is $d\vartheta/ds = b_1(x_s, y_s)\, \partial\vartheta/\partial x + b_2(x_s, y_s)\, \partial\vartheta/\partial y$, where the partial derivatives are evaluated on the separatrix. Since the saddle point corresponds to $s = \infty$, (47) is actually not directly usable for computation. In fact, at the saddle point, both $b_1(x_s, y_s)$ and $b_2(x_s, y_s)$ are zero, so that (47) becomes a nonlinear matrix equation for the second derivatives of θ:

$$BP + PB^{\mathrm{T}} + PAP = 0. \tag{48}$$

Once (48) is solved, we know the second derivatives of $\theta(x, y)$ at the saddle point. To find the values of the second derivatives along the separatrix, we first note that a saddle point is characterized by one positive eigenvalue and one negative eigenvalue. The eigenvector corresponding to the negative eigenvalue is the direction of the separatrix near the saddle point. We then consider a point close to the saddle and on the separatrix, take one more term in the expansion of (39), and evaluate $\partial^2\theta/\partial x^2$, $\partial^2\theta/\partial x\,\partial y$, and $\partial^2\theta/\partial y^2$ in the direction of the eigenvector corresponding to the negative eigenvalue by a Taylor expansion. We then run time backwards along the separatrix and use (47) to determine the second derivatives of $\theta(x_s, y_s)$ at points (x_s, y_s) on the separatrix. Mangel and Ludwig (1977) provide an example of this computation. Alternatively, Mangel and Ludwig show that (47) can be converted to a single ordinary differential equation for the second normal derivative of $\theta(x_s, y_s)$ along the separatrix. Given the power of modern computational packages, it is probably just as easy to work with (47, 48). Once these derivaties are known, since $\theta(x \cdot y) = -\frac{1}{2}\psi(x, y)^2$ and $\theta(x, y) \sim (\partial^2\theta/\partial x^2)(\delta x)^2 + 2(\partial^2\theta/\partial x\,\partial y)(\delta x\,\delta y) + (\partial^2\theta/\partial x^2)(\delta y)^2$, the value of $u(x, y)$ at a point near the separatrix is

$$u(x, y) \sim g_0 \Phi\left(\pm \sqrt{-2\frac{\partial^2\theta}{\partial x^2}(\delta x)^2 - 4\frac{\partial^2\theta}{\partial x\,\partial y}(\delta x\,\delta y) - 2\frac{\partial^2\theta}{\partial x^2}(\delta y)^2} \Big/ \sqrt{\varepsilon} \right), \tag{49}$$

where the choice of sign $(\pm)$ is determined by the location of the point relative to the sparatrix and the constant g_0 is picked so that points far from the separatrix, in the appropriate direction, have $u(x, y) \sim 1$. To implement this result, we employ the following procedure:

Step 1. Find the values of the second derivatives at the saddle point from (48).

Step 2. Find the eigenvector corresponding to the negative eigenvalue of the linearized deterministic system at the saddle point.

Step 3. Move away from the separatrix in the direction of this eigenvector.

Step 4. Integrate (47) backwards in time along the separatrix.

Step 5. Use (49) to determine $u(x, y)$ for points not on the separatrix.

However, as the following two examples show, the full machinery is often not needed to solve apparently complicated problems.

Example 1. *A Simple Model of Stochastic Competition*

The following extremely simple model of stochastic competition has been analyzed by a number of authors (e.g., May, 1973; Mangel and Ludwig, 1977):

$$\begin{aligned} dX &= X[1+c-X-cY]\,dt+\sqrt{\varepsilon}\,dW_1 \\ dY &= Y[1+c-Y-cX]\,dt+\sqrt{\varepsilon}\,dW_2. \end{aligned} \tag{50}$$

The underlying deterministic system is

$$\begin{aligned} \frac{dX}{dt} &= X[1+c-X-cY] \\ \frac{dY}{dt} &= Y[1+c-Y-cX] \end{aligned} \tag{51}$$

which has stable steady states at $(0, 1+c)$, $(1+c, 0)$, and a saddle point at $(1, 1)$ if $c > 1$. In this case the separatrix is the line $Y = X$. If we transform variables to

$$\begin{aligned} U &= \frac{\sqrt{2}}{2}[X+Y-2] \\ V &= \frac{\sqrt{2}}{2}[Y-X] \end{aligned} \tag{52}$$

then the dynamics of U and V are

$$\begin{aligned}\frac{dU}{dt}&=\frac{1}{\sqrt{2}}[-U^2-V^2-cU^2+cV^2-\sqrt{2}\,U(1+c)]\\ \frac{dV}{dt}&=\frac{1}{\sqrt{2}}[-2UV-\sqrt{2}\,V(1-c)].\end{aligned} \tag{53}$$

The steady states of this system are now $(0, 0)$, $(c-1, -1-c)\sqrt{2}/2$, and $(c-1, 1+c)\sqrt{2}/2$. The latter two are stable and the origin is a saddle point; $V=0$ is now the separatrix. On the separatrix we have

$$\frac{dU}{dt}=\frac{-1}{\sqrt{2}}(1+c)\,U[U+\sqrt{2}]. \tag{54}$$

From the analysis of (33), recall that both $\theta(U, V)$ and its tangential derivative are 0 along the separatrix. But here the tangential derivative is the same as $\partial\theta/\partial U$ and the normal derivative is $\partial\theta/\partial V$. Thus, the only non-zero second derivative of $\theta(U, V)$ is $(\partial^2\theta/\partial V^2)(U)$, which is a function of U along the separatrix. This means that (47) becomes

$$\frac{d}{ds}\left(\frac{\partial^2\theta}{\partial V^2}(U)\right)+2B(U)\frac{\partial^2\theta}{\partial V^2}(U)+A(U)\left(\frac{\partial^2\theta}{\partial V^2}(U)\right)^2=0, \tag{55}$$

where $B(U)$ and $A(U)$ are defined above (47) after transformation from (X, Y) to (U, V) coordinates.

Equation (55) is a Ricatti equation, which is solved by the substitution $W(U)=1/(\partial^2\theta/\partial V^2)(U)$. As described in the general method, at the saddle point, (55) becomes an algebraic equation for $(\partial^2\theta/\partial V^2)(U=0)=(\partial^2\theta/\partial V^2)(s=\infty)$. The solution of (55) is then chosen to satisfy this condition.

In summary, this example shows how a two dimensional problem can be converted to essentially a one dimensional problem, but because the function $B(U)$ is complicated, one still needs to integrate (55) along the separatrix. A more complicated and realistic model of competition is discussed by Mangel and Ludwig (1977), who give numerical results and compare the theory with Monte Carlo simulations.

Example 2. The Correlated Response to Selection (*Price et al.* 1992)

Price *et al.* (1992) studied the correlated response to selection for two characters X and Y with phenotypic means μ_x and μ_y, variances σ_x^2 and σ_y^2, and phenotypic correlation ρ_p. Selection was assumed to act independently upon X and Y, with X experiencing stabilizing selection of the form $w_x(x) = \exp(-x^2/2\omega_x^2)$ and Y experiencing stabilizing selection of the form $w_y(y) = \frac{1}{2}\exp(-(y-y_e)^2/2\omega_y^2) + \frac{1}{2}\exp(-(y+y_e)^2/2\omega_y^2)$, so that μ_y has stable equilibria at $\pm y_e$. In fact, for the purposes of analysis, if y_e is sufficiently large and ω_y^2 sufficiently small, one can equally consider destabilizing selection for Y of the form $w_y(y) = \exp(y^2/2b^2)$. In this simpler case there are no finite, stable equilibria for μ_y, which approaches $\pm\infty$. In this case, a peak shift occurs if Y begins near y_e (or $-y_e$) and ends up approaching ∞ $(-\infty)$ as a result of selection.

The changes in the means from one generation to the next are modeled by the standard approximation of quantitative genetics and are of the form

$$\begin{aligned} \Delta\mu_x &= g_x^2\,\beta_x + \rho_g\, g_x\, g_y\, \beta_y \\ \Delta\mu_y &= \rho_g\, g_x\, g_y\, \beta_x + g_y^2\, \beta_y, \end{aligned} \tag{56}$$

where the additive genetic variances are g_x^2 and g_y^2, the additive genetic correlation is ρ_g, and the selection gradients are linear functions of the means,

$$\begin{aligned} \beta_x &= (1/D)[-(\sigma_y^2 - b^2)\,\mu_x + \rho_p\sigma_x\sigma_y\,\mu_y] \\ \beta_y &= (1/D)[-(\sigma_x^2 + \omega_x^2)\,\mu_x + \rho_p\sigma_x\sigma_y\,\mu_x], \end{aligned} \tag{57}$$

where $D = (\sigma_y^2 - b^2)(\sigma_x^2 + \omega_x^2) - \rho_p\sigma_x\sigma_y$. Price *et al.* (1992) then show that using (57) in (56) converts (56) to a pair of linear difference equations:

$$\begin{aligned} \Delta\mu_x &= b_{11}\,\mu_x + b_{12}\,\mu_y \\ \Delta\mu_y &= b_{21}\,\mu_x + b_{22}\,\mu_y. \end{aligned} \tag{58}$$

This is the underlying deterministic system that is then perturbed by fluctuations. It can be shown that the matrix $B = (b_{ij})$ has one positive and one negative eigenvalue, corresponding to a saddle point at $(0, 0)$. Price *et al.* Show that the evolution of the vector $\mu = (\mu_x, \mu_y)^T$ can be written as $\mu(t) = (B+I)^t\,\mu(0)$, where $\mu(t)$ is the value of the vector of means at time t and $\mu(0)$ is its value at 0, and I is the identity matrix. Furthermore, since the matrix $B+I$ has distinct eigenvalues, we have $(B+I)^t = U\Lambda^t U^{-1}$, where Λ is the diagonal matrix with whose elements are the eigenvalues of $B+I$ and U is the matrix whose columns are the corresponding eigenvectors of $B+I$. We thus have a phase portrait analgous to Fig. 1, in which—because the dynamics are linear—the separatrix is once again a straight line..

Price *et al.* (1992) now add stochastic perturbations to (58) and write it as

$$\begin{aligned} \Delta\mu_x &= b_{11}\,\mu_x + b_{12}\,\mu_y + \xi_x \\ \Delta\mu_y &= b_{21}\,\mu_x + b_{22}\,\mu_y + \xi_y. \end{aligned} \tag{59}$$

Here (ξ_x, ξ_y) has covariance matrix $(1/N)\mathbf{G}$, where N is population size, and where higher moments are $o(1/N)$. Price *et al.* then approximate (59) by a two dimensional Ornstein–Uhlenbeck process with infinitesimal mean given by the right hand side of (58) and infinitesimal covariance $(1/N)\mathbf{G}$.

As in the previous example, Price *et al.* (1992) reduce the fundamental equation (47) to a single equation for a single unknown by changing coordinates from μ to $z = U^{-1}\mu$. Then the dynamics of z are approximated by a diffusion with infinitesimal mean Λz and infintesimal covariance $(1/N)\,U^{-1}G(U^{-1})^{\mathrm{T}}$. With the change of coordinates, the abscissa becomes the separatrix (as in the previous example), and we need only find the value of the second normal derivative of $\partial^2\theta/\partial z_2^2$ along the separatrix.

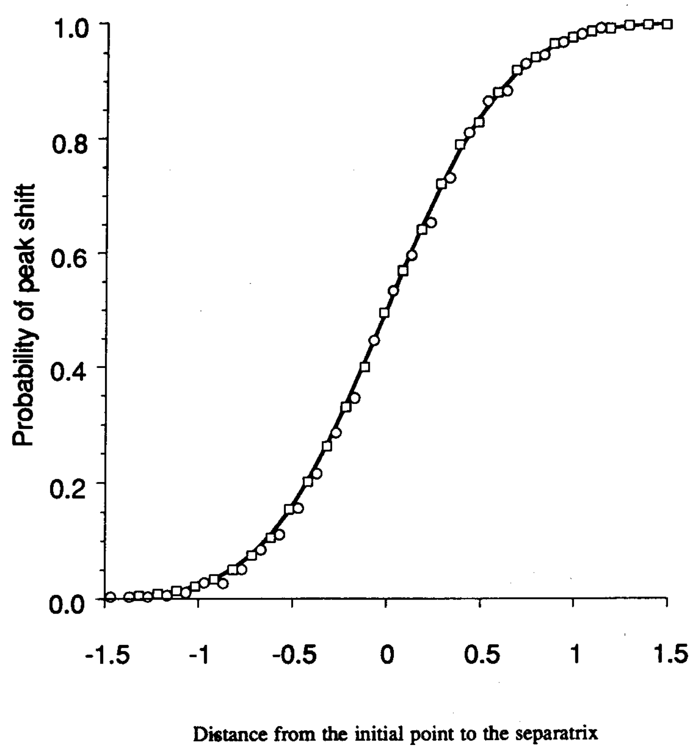

FIG. 4. The results of Price *et al.* (1992), comparing the theory (solid line) described in this paper, and used by them, with Monte Carlo simulation of the underlying stochastic difference system based on stabilizing selection with two peaks (circles) or destabilizing selection (squares). Parameters were $y_e = 5$, $\omega_y^2 = \omega_x^2 = 10$, $\sigma_y^2 = \sigma_x^2 = 1$, $h_y^2 = h_x^2 = 0.5$, and $\rho_g = \rho_e = \rho_p = 0.75$. Reproduced with permission from Price *et al.* (1992).

Because the dynamics are linear, in fact, (47) shows that $\partial^2\theta/\partial z_2^2$ is constant along the separatrix. Price *et al.* show that this constant is $\partial^2\theta/\partial z_2^2 = -2\lambda_+/(U^{-1}G(U^{-1})^T)_{22}$, where λ_+ is the positive eigenvalue. Thus, the probability of a peak shift from initial values (μ_x, μ_y) is Prob{peak shift from $(\mu_x, \mu_y)\} = \Phi(\text{dist}\sqrt{2\lambda_+/(U^{-1}G(U^{-1})^T)_{22}})$, where dist is the distance from the initial condition (μ_x, μ_y) to the separatrix (which corresponds to the stable eigenvector).

Price *et al.* compared the theoretical results with numerical computation, based directly on the difference equation rather than the approximating diffusion model. Their results (Fig. 4) show that the theory just described is extremely accurate for both the model of stabilizing selection with two peaks and the model of destabilizing selection and that the diffusion approximation for (59) is extremely accurate. This can be understood by considering the derivation of (29): the probability of peak shift is characterized by the expectation of a first exit time. Thus, the form of the time dynamics (i.e, the left hand side of (59) or its diffusion approximation—differences or differentials) is actually irrelevant to the form of the backward equation (29).

Barrier Crossings for the Markov Chain

We now consider barrier crossing for stochastic difference equations. The function $u(x, y)$ is defined as in (28). Applying the law of total expectation to this definition gives

$$u(x, y) = \sum_i \sum_j \rho_{ij}(x, y)\, u(x + \varepsilon i, y + \varepsilon j). \tag{60}$$

In this equation the summation is over all values of i and j for which $\rho_{ij}(x, y)$ is non-zero. Based on the results of the previous section, we assume that $u(x, y)$ is given by (30), with $\psi(x, y)$ satisfying

$$0 = \sum_i \sum_j \rho_{ij}(x, y)\left[1 - \exp\left(-\psi\left(i\frac{\partial\psi}{\partial x} + j\frac{\partial\psi}{\partial y}\right)\right)\right]. \tag{61}$$

The solution of this equation by the method of characteristics is virtually identical to the procedure followed previously. More importantly, Eq. (61) can be used to study (i) the validity of the diffusion approximation to the stochastic difference equation and (ii) the origin of macroscopic, deterministic population dynamics.

On the Validity of the Diffusion Approximation

The validity of a diffusion approximation is usually demonstrated (e.g., Kurtz, 1981) by proving weak convergence of a sequence of stochastic processes to the diffusion or by showing that important quantities such as moments or distribution functions coverge. This is required because the naive approach of expanding (60) in powers of ε and then assuming that terms higher than second order are insignificant provides no way of estimating the third or higher order derivatives, which could be very large. However, the procedure leading to (61) can give an idea of the region of (x, y) space in which the diffusion approximation will be valid. When ψ, ψ_x, and ψ_y are small, we Taylor expand the exponential in (61) to obtain

$$0=\sum_i \sum_j \rho_{ij}(x, y)\left\{\psi\left(i\frac{\partial\psi}{\partial x}+j\frac{\partial\psi}{\partial y}\right)-\frac{1}{2}\left[\psi\left(i\frac{\partial\psi}{\partial x}+j\frac{\partial\psi}{\partial y}\right)\right]^2\right\}. \tag{62}$$

This equation corresponds exactly to (31), when appropriate identifications of $b_i(x, y)$ and $a_{ij}(x, y)$ are made. We thus conclude that the diffusion approximation should be valid in the regions where ψ is small and its derivatives are not changing too rapidly. This will be in the region around the separatrix, which is the region of most interest from the experimental viewpoint. Thus, the diffusion approximation should be valid in the neighborhood of the separatrix. This conclusion explains, in part, why Price *et al.* (1992) obtained excellent results when comparing the diffusion and Markov chain models.

Emergence of Macroscopic Population Dynamics

Equation (61) is the starting point to show how macroscopic population dynamics emerge from the Markov chain description. As before, we set $\theta=-\frac{1}{2}\psi^2$ in (61) to obtain

$$1=\sum_i \sum_j \rho_{ij}(x, y)\exp\left(i\frac{\partial\theta}{\partial x}+j\frac{\partial\theta}{\partial y}\right). \tag{63}$$

The Hamiltonian $H(x, y, p, q)$ is now

$$H(x, y, p, q)=1-\sum_i \sum_j \rho_{ij}(x, y)\exp(ip+jq). \tag{64}$$

The trajectories generated by this Hamiltonian are the solutions of

$$\begin{aligned}
\frac{dx}{ds} &= \frac{\partial H}{\partial p} = -\sum_i \sum_j \rho_{ij}(x, y)\, i \exp(ip + jq) \\
\frac{dy}{ds} &= \frac{\partial H}{\partial q} = -\sum_i \sum_j \rho_{ij}(x, y)\, j \exp(ip + jq) \\
\frac{dp}{ds} &= -\frac{\partial H}{\partial x} = \sum_i \sum_j \frac{\partial p_{ij}(x, y)}{\partial x} \exp(ip + jq) \\
\frac{dq}{ds} &= -\frac{\partial H}{\partial y} = \sum_i \sum_j \frac{\partial \rho_{ij}(x, y)}{\partial y} \exp(ip + jq).
\end{aligned} \tag{65}$$

We now consider the sub-class of trajectories where $p(x, y, dx/ds, dy/ds) = 0$ and $q(x, y, dx/ds, dy/ds) = 0$. The first two equations in Eq. (65) become

$$\begin{aligned}
\frac{dx}{ds} &= -\sum_i \sum_j \rho_{ij}(x, y) i \\
\frac{dy}{ds} &= -\sum_i \sum_j \rho_{ij}(x, y) j.
\end{aligned} \tag{66}$$

These are equivalent to "deterministic trajectories" which we might associate with the underlying Markov chain model. That is, the "deterministic" trajectories are a sub-set of the solutions of the Hamilton–Jacobi equations for which $p(x, y, dx/ds, dy/ds)$ and $q(x, y, dx/ds, dy/ds)$ are identically 0. This interpretation of the macroscopic equations complements the work of Kurtz (1981), who provides a probabilistic interpretation for the emergence of macroscopic dynamics.

Acknowledgments

The work of Trevor Price, Michael Turelli, and Monte Slatkin motivated me to write this paper; I thank them for having applied the original method in a new and insightful manner. This work was partially completed while I held the positions of Scheinbrun Professor of Botany, Hebrew University of Jerusalem (1987–1988) and Visiting Professor Biology, Simon Fraser University (1991); I thank P. Daltrop, K. Jones, and B. D. Roitberg for hospitality. The work was partially supported by NSF Grants OCE90-16895 and BSR 86-1073 and 91-17603, and by the John Simon Guggenheim Memorial Foundation. For comments on various versions of the manuscript, I thank N. Barton, P. E. Bradshaw, J. Gillespie, D. Ludwig, M. Turelli, and M. Slatkin, and two anonymous referees.

REFERENCES

BARTON, N. H., AND ROUHANI, S. 1987. The frequency of shifts between alternative equilibria, *J. Theor. Biol.* **125**, 397–418

BORN, M., AND WOLF, E. 1980. "Principles of Optics," Pergamon, Oxford.

COSTANTINO, R. F., AND DESHARNAIS, R. A. 1991. "Population Genetics and Demography of *Tribolium*," Springer-Verlag, New York.

COURANT, R., AND HILBERT, D. 1962. "Methods of Mathematical Physics, Vol. II," Wiley–Interscience, New York.

EWENS, W. 1979. "Mathematical Population Genetics," Springer-Verlag, New York.

FELLER, W. 1971. "An Introduction to Probability Theory and Its Applications," Wiley, New York.

KURTZ, T. G. 1981. "Approximation of Population Processes," SIAM, Philadelphia.

LUDWIG, D. 1974. "Stochastic Population Theories," Springer-Verlag, New York.

LUDWIG, D. 1975. Persistence of dynamical systems under random perturbations, *SIAM Rev.* **17**, 605–640.

MACARTHUR, R. H., AND WILSON, E. O. 1967. "The Theory of Island Biogeography," Princeton Univ. Press, Princeton, NJ.

MANGEL, M., AND LUDWIG, D. 1977. Probability of extinction in a stochastic competition, *SIAM J. Appl. Math.* **33**, 256–266.

MANGEL, M., AND CLARK, C. W. 1988. "Dynamic Modeling in Behavioral Ecology," Princeton Univ. Press, Princeton, NJ.

MAY, R. M. 1973. "Stability and Complexity in Model Ecosystems," Princeton Univ. Press, Princeton, NJ.

NEYMAN, J., PARK, T., AND SCOTT, E. L. 1956. Struggle for existence. The *Tribolium* model: Bilological and statistical aspects, *in* "Third Berkeley Symposium on Mathematical Statistics and Probability" (J. Neyman, Ed.), pp. 41–78.

PARK, T. 1962. Beetles, competition and populations, *Science* **138**, 1369–1375.

PETERS, C., MANGEL, M., AND COSTANTINO, R. 1989. Stationary distribution of population size in *Tribolium*, *Bull. Math. Biol.* **51**, 625–638.

PRICE, T., TURELLI, M., AND SLATKIN, M. 1992. Peak shifts produced by correlated response to selection, *Evolution*, in press.

ROUHANI, S., AND BARTON, N. H. 1987a. The probability of peak shifts in a founder population, *J. Theor. Biol.* **126**, 51–62.

ROUHANI, S., AND BARTON, N. H. 1987b. Speciation and the "shifting balance" in a continuous population, *Theor. Pop. Biol.* **31**, 465–492

RUND, H. 1973. "The Hamilton-Jacobi Theory in the Calculus of Variations," Kreiger, Huntington, N. Y.

SCHUSS, Z. 1980. "Stochastic Differential Equations," Wiley, N. Y.

SIMBERLOFF, D. 1980. A succession of paradigms in ecology: Essentialism to materialism and probabilism, *Synthese* **43**, 3–39.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium