NUMERICAL ANALYSIS OF WEAK RANDOM DRIFT IN A CLINE

MITCHELL LUSKIN

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109

THOMAS NAGYLAKI†

*Department of Biophysics and Theoretical Biology, The University of Chicago,
920 East 58th Street, Chicago, Illinois 60637*

ABSTRACT

The equilibrium state of a diffusion model for weak random genetic drift in a cline is analyzed numerically. The monoecious organism occupies an unbounded linear habitat with constant, uniform population density. Migration is homogeneous, symmetric, and independent of genotype. A single diallelic locus with a step environment is investigated in the absence of dominance and mutation. The ratio of the variance of either gene frequency to the product of the expected gene frequencies decreases monotonically to a non-zero constant. The correlation between the gene frequencies at two points decreases monotonically to zero as the separation is increased with the average position fixed; the decrease is asymptotically exponential. The correlation decreases monotonically to a positive constant depending on the separation as the average position increasingly deviates from the center of the cline with the separation fixed. The correlation also decreases monotonically to zero if one of the points is fixed and the other is moved outward in the habitat, the ultimate decrease again being exponential. All the results are parameter free. Some asymptotic formulae are derived analytically.

IN a recent paper (NAGYLAKI 1978a), the biological importance of random genetic drift in a cline was briefly discussed, previous work was reviewed, and a diffusion model was developed. The diffusion treatment requires that the selection coefficient, s, the variance of the migrational displacement, σ^2, and the reciprocal of the population density, $1/\rho$, be all small and of the same order of magnitude (NAGYLAKI 1978b), and yields the following results (NAGYLAKI 1978a). For most selection gradients, s, σ^2, and ρ appear combined only in the single dimensionless parameter β, the ratio of the characteristic length for migration and random drift to the natural distance for migration and selection. Random drift is highly significant if $\beta << 1$; it causes only relatively small variations of order $1/\beta$ around the deterministic cline if $\beta >> 1$. With $\beta >> 1$, the correlation between the gene frequencies at any two points is very nearly independent of β, and hence is parameter free for particular forms of the selection gradient. For a very steep selection gradient, $\beta = 2\sqrt{2}s\rho\sigma$, *i.e.*, essentially the product of the

square root of the selection intensity and the neighborhood size. Thus, with weak selection a large neighborhood size is required for large β.

Since the derivation of detailed analytic results appears to be quite difficult, we must resort to numerical methods. Such treatments of random drift in linear stepping-stone clines of finite length have been given by FELSENSTEIN (1975) and SLATKIN and MARUYAMA (1975). The latter authors' study of large random drift in a cline is discussed in NAGYLAKI (1978a). Numerical exploration of this difficult problem, including an investigation of the persistence time of an allele in an environmental pocket, will commence soon. Our computations in this note relate directly to FELSENSTEIN's (1975) small random drift work.

By combining parameters into the single quantity β, the diffusion approximation enormously increases the range of parameters accessible to numerical analysis. This advantage is even greater with weak random drift, for then parameter-free results can be obtained for particular forms of the selection gradient. A single calculation with a fixed value of β covers all possible values of s, σ^2, and ρ yielding that β. Furthermore, without the diffusion approximation, numerical investigation of parameter sets with weak migration and selection and high population density is not practicable.

Our diffusion model is the limit of the following discrete scheme. A diploid monoecious population occupies an infinite chain of equally spaced colonies. We suppose migration is homogeneous, symmetric, and independent of genotype, and consider a single diallelic locus with alleles A_1 and A_2 in the absence of dominance and mutation. The generation starts with the same very large number of zygotes in each colony. Selection and migration, occurring in succession, may alter the subpopulation sizes, but the latter must remain extremely large. Any biologically reasonable variation of the subpopulation numbers under selection, in particular, no change (soft selection) and change determined by the mean fitness in each colony (hard selection), leads to the same diffusion limit (NAGYLAKI 1978b). Random genetic drift acts through population regulation, which truncates each colony to the same finite number of adults. Each of these adults produces many gametes without fertility differences. These gametes fuse at random to form the next generation of zygotes, in Hardy-Weinberg proportions in each colony.

We confine ourselves to the step environment, positing $1 + s$ sgn x, 1, $1 - s$ sgn x ($s > 0$; sgn $x = x/|x|$, $x \neq 0$) for the fitnesses of A_1A_1, A_1A_2, A_2A_2 at position x ($-\infty < x < \infty$). This choice represents precisely cases of abrupt environmental change, such as sudden shifts in soil or vegetation, and approximates fairly accurately situations in which the characteristic length of the environmental variation is less than $l = \sigma/\sqrt{2s}$ (SLATKIN 1973), the natural length for the cline corresponding to the step environment. Many of our general conclusions will concern long-distance behavior; these should hold qualitatively even for slower environmental variation (FISHER 1950; FELSENSTEIN 1975; MAY, ENDLER and MCMURTRIE 1975; NAGYLAKI 1975, 1978c; SLATKIN and MARUYAMA 1975). Notice that for simplicity we have assumed equal selection intensities in the two environments.

We investigate only the equilibrium state of the population. For the step environment, the characteristic convergence time to equilibrium is $1/s$ (NAGYLAKI 1978a). Our treatment will be restricted to the expected gene frequency in the cline and the covariance between the gene frequencies at two arbitrary points. Let us denote the mean allelic frequency at x by $\bar{p}(\xi)$, where $\xi = x/l$ is a convenient dimensionless spatial coordinate. We express the covariance between the gene frequencies at x and $y = l\eta$ as a function, $v(z,w)$, of the dimensionless average position $z = (\xi + \eta)/2$ and separation $w = (\xi - \eta)/2$.

Our approximation applies for weak random drift, defined by $\beta = 2\sqrt{2s}\rho\sigma >> 1$. Since the neighborhood size, $\rho\sigma$, almost always seems to exceed 20 or 30, and is usually at least a few hundred (WRIGHT 1978, Ch. 2), the approximation should be quite accurate unless selection is quite weak. With $\beta >> 1$, the average gene frequency is very close to the deterministic one (NAGYLAKI 1978a):

$$\bar{p}(\xi) = -(1/2) + (3/2) \tanh^2 [(\xi/2) + \tanh^{-1} \sqrt{2/3}], \quad \xi \geq 0. \tag{1}$$

By the symmetry of the problem, $\bar{p}(-\xi) = 1 - \bar{p}(\xi)$. We owe the deterministic solution to HALDANE (1948). Observe that $\bar{p}(\xi)$ increases monotonically from $1/2$ to 1 as ξ increases from 0 to ∞.

The covariance is approximately proportional to $1/\beta$: $v(z,w) = \beta^{-1}u(z,w)$ with u independent of β in the limit $\beta \to 0$. Even in the discrete model, one can see that the covariance is inversely proportional to the colony size (FELSENSTEIN 1975). Taking into account the symmetries $u(-z,w) = u(z-w) = u(z,w)$, we have the following boundary value problem for the parameter-free function u (NAGYLAKI 1978a):

$$u_{zz} + u_{ww} + 2f(z,w)u = 0, \quad z,w > 0, \tag{2a}$$

$$f(z,w) = 1 - 2\bar{p}(z + w) + \text{sgn}\,(z - w)[1 - 2\bar{p}(z - w)], \tag{2b}$$

$$u_z(0,w) = 0, \tag{2c}$$

$$u_w(z,0+) + \bar{p}(z)[1 - \bar{p}(z)] = 0, \tag{2d}$$

$$u(z,w) \to 0 \quad \text{as} \quad z^2 + w^2 \to \infty, \tag{2e}$$

in which the subscripts represent partial derivatives (e.g., $u_{zz} = \partial^2 u/\partial z^2$).

Since the gene frequency has mean $\bar{p}(z)$ and must be in $[0,1]$, the maximum possible value of the variance $v(z,0)$ is $\bar{p}(z)[1 - \bar{p}(z)]$. Hence, the ratio of these quantities, $F(z)$, is a good measure of random variation. We use the parameter-free entity

$$\beta F(z) = \frac{u(z,0)}{\bar{p}(z)[1 - \bar{p}(z)]}. \tag{3}$$

The correlation between the allelic frequencies at x and y,

$$r(z,w) = u(z,w)[u(z + w,0)\,u(z - w,0)]^{-1/2}, \tag{4}$$

is also parameter free.

If both points at which the gene frequency is measured are on the same side of the center of the cline and far away, i.e., $z - w \gg 1$, it is easy to analyze (2) asymptotically. As $z \to \infty$, (1) yields

$$p(z) \simeq 1 - be^{-z}, \quad b = 6(5 - 6\sqrt{2/3}) = 0.606. \tag{5}$$

As $z - w \to \infty$, (2) informs us that $f(z,w) \to -2$. Therefore, the boundary value problem (2) becomes asymptotically

$$u_{zz} + u_{ww} - 4u = 0, \quad w > 0, \tag{6a}$$

$$u_w(z,0+) + be^{-z} = 0, \tag{6b}$$

$$u(z,w) \to 0 \quad \text{as} \quad z^2 + w^2 \to \infty. \tag{6c}$$

Separation of variables shows that solutions of (6a) have the form

$$u(z,w) \sim ce^{-\lambda z - \mu w}, \quad \lambda^2 + \mu^2 = 4.$$

By dint of (6c), $\lambda > 0$ and $\mu > 0$. Enforcing (6b) for all large z, we find $\lambda = 1$, $\mu = \sqrt{3}$, and $c = b/\sqrt{3} = 0.350$. Thus,

$$u(z,w) \sim 0.350 \, e^{-z - \sqrt{3}w} \quad \text{as} \quad z - w \to \infty. \tag{7}$$

Substituting (5) and (7) into (3) gives

$$\beta F(z) \to \frac{1}{\sqrt{3}} = 0.577 \quad \text{as} \quad z \to \infty. \tag{8}$$

Inserting (5) and (7) into (4) leads to the correlation

$$r(z,w) \sim e^{-\sqrt{3}w} \quad \text{as} \quad z - w \to \infty. \tag{9}$$

We shall now present the standardized variance and correlation obtained from numerical solution of (2). We expect all the results to deviate at most by 1% from the (unknown) exact solution of (2).

Figure 1 shows $\beta F(z)$, which decreases monotonically from a maximum value $\beta F(0) = 0.719$ as z increases. The slope is zero at the origin because $F(-z) = F(z)$. Note the precise agreement with the analytic result (8). Figure 1 is generally consistent with FELSENSTEIN's (1975) calculations, but demonstrates that the slight rise in his variances at the extremities of his chain of colonies is an end effect.

The monotonic decrease of the correlation for fixed average position, z, and variable separation, w, is displayed in Figure 2, which shows $-\ln r(z,w)$ plotted against w. For large values of w, the following exponential approximations hold:

$$r(0,w) \approx r(1,w) \approx 1.00 \, e^{-1.05w}, \quad w \geq 3,$$

$$r(3,w) \approx 0.426 \, e^{-1.00w}, \quad w \geq 4,$$

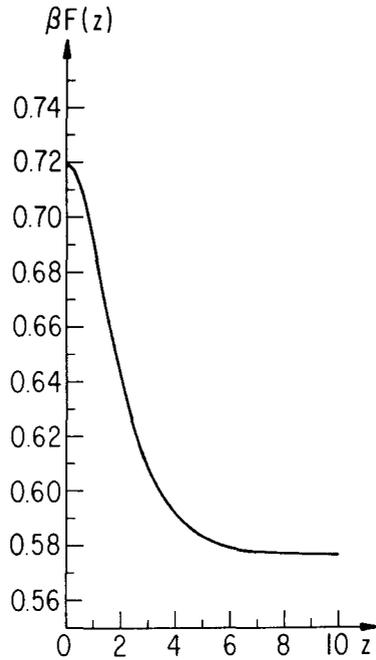$$r(5,w) \approx 0.0907 \, e^{-0.92w}, \quad w \geq 5.$$

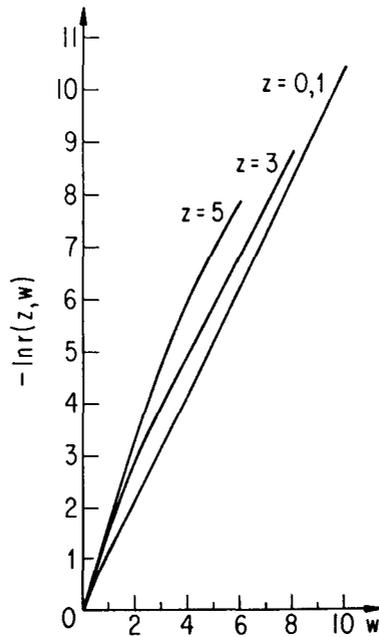FIGURE 1.—The standardized variance as a function of position.

FIGURE 2.—The correlation as a function of separation for various fixed values of the average position.

Observe that the decay rate decreases slowly with z. Plots of $u(z,w)$ for various fixed values of z show very little departure from exponential decay, even for small w, and a slowly decreasing decay constant (as a function of z again) in the apparently pure exponential decay for large w. FELSENSTEIN (1975) noticed approximate exponential decay of the correlation with separation. In the neutral case with mutation, the correlation is independent of z, owing to spatial homogeneity, and decays exponentially with w (MALÉCOT 1955; WEISS and KIMURA 1965; NAGYLAKI 1976; SAWYER 1977).

Figure 3 exhibits the monotonic decrease of the correlation as a function of the average position, z, for various values of the separation, w. The curves are horizontal at $z = 0$ because $r(-z,w) = r(z,w)$. The constant limits as $z \to \infty$ agree with (9). FELSENSTEIN (1975) obtained a slight dip in the correlation between nearest neighbors at the center of the cline. Our diffusion results do not show this nonmonotonicity, but otherwise have the same qualitative form as his computations.

Suppose now that we fix one point at η and vary the position of the other point, ξ, moving it outward from η. In Figure 4 we display $r(w + \eta, w)$ for various values of η. The monotone decrease is very close to exponential throughout; for $\eta = 3$ the results are already close to (9), which applies as $\eta \to \infty$.
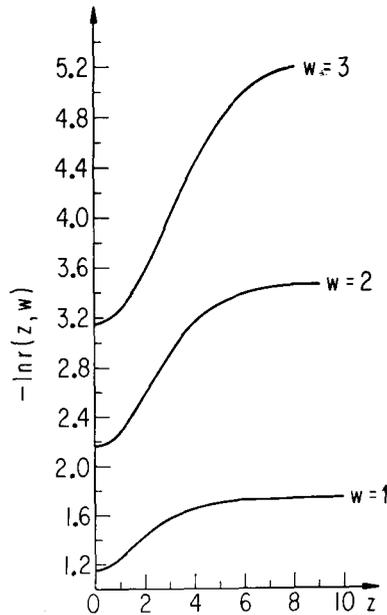
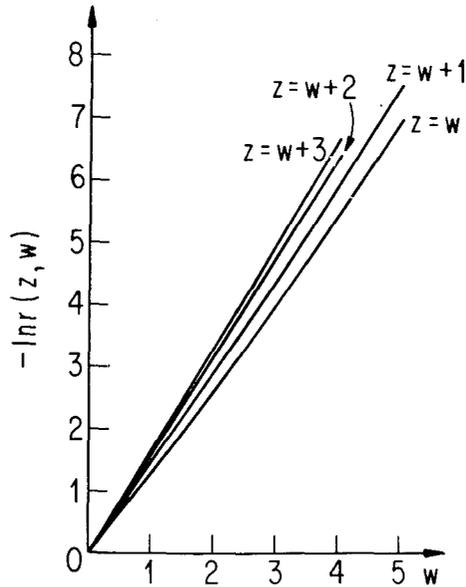FIGURE 3.—The correlation as a function of average position for various fixed values of the separation.

FIGURE 4.—The correlation as a function of separation for various fixed locations of the point closer to the environmental transition.

LITERATURE CITED

FELSENSTEIN, J., 1975   Genetic drift in clines which are maintained by migration and natural selection. Genetics **81**: 191–207.

FISHER, R. A., 1950   Gene frequencies in a cline determined by selection and diffusion. Biometrics **6**: 353–361.

HALDANE, J. B. S., 1948   The theory of a cline. J. Genet. **48**: 277–284.

MALÉCOT, G., 1955   Decrease of relationship with distance. Cold Spring Harbor Symp. Quant. Biol. **20**: 52–53.

MAY, R. M., J. A. ENDLER, and R. E. McMURTRIE, 1975   Gene frequency clines in the presence of selection opposed by gene flow. Am. Naturalist **109**: 659–676.

NAGYLAKI, T., 1975   Conditions for the existence of clines. Genetics **80**: 595–615.   ——,   1976   The decay of genetic variability in geographically structured populations. II. Theoret. Pop. Biol. **10**: 70–82.   ——,   1978a   Random genetic drift in a cline. Proc. Natl. Acad. Sci. U.S. **75**: 423–426.   ——,   1978b   A diffusion model for geographically structured populations. J. Math. Biol. **6**: 375–382.   ——,   1978c   The geographical structure of populations. In: *Studies in Mathematical Biology*. Edited by S. LEVIN. Mathematical Association of America, Washington.

SAWYER, S., 1977   Asymptotic properties of the equilibrium probability of identity in a geographically structured population. Advan. Appl. Prob. **9**: 268–282.

SLATKIN, M., 1973   Gene flow and selection in a cline. Genetics **75**: 733–756.

SLATKIN, M., and T. MARUYAMA, 1975   Genetic drift in a cline. Genetics **81**: 209–222.

WEISS, G. H., and M. KIMURA, 1965   A mathematical analysis of the stepping-stone model of genetic correlations. J. Appl. Prob. **2**: 129–149.

WRIGHT, S., 1978   *Evolution and the Genetics of Populations*. Vol. IV. The University of Chicago Press, Chicago.