Some Measure-Valued Markov Processes in Population Genetics Theory

WENDELL H. FLEMING & MICHEL VIOT

1. Introduction. This paper is concerned with stochastic population models of the following kind. Suppose that a population consists of a large, but finite, number of individuals; and that each individual in the population has a "type" described by an element x of some set S. We are concerned not with the types of particular individuals, but rather with the distributions of types among the whole population. If S is a finite set, with J elements $x_1, \cdots, x_J$, then let p_j denote the frequency of type x_j ($p_j \geq 0$, $p_1 + \cdots + p_J = 1$). The vector $p = (p_1, \cdots, p_J)$ describes the type distribution; moreover, $p \in \Sigma^J$, where Σ^J is a $(J - 1)$-dimensional simplex.

In population genetics p ordinarily refers to the frequency distribution of gametes, rather than of individuals, in a population. Thus, for diploid populations, the term "individual" in the present paper should be interpreted as "gamete". When $J = 2$ the classical Wright-Fisher model describes the changes in frequencies of two possible gametic types (or alleles) at a given gene locus, considering such factors as mutation, random genetic drift, and selective advantages among the various genotypes. By the technique of diffusion approximation, a 1-dimensional diffusion process on the interval $0 \leq p_1 \leq 1$ is obtained [4]. The corresponding diffusion approximation with J types has been discussed in [8], [12], [19] and by other authors. This diffusion process has Σ^J as state space; see §2. Its generator $\mathfrak{A}^J$ has the form (2.1), if an appropriate time scale for the diffusion is chosen. The Wright-Fisher process is a discrete-time Markov chain. Alternatively, one can start with a continuous-time Markov chain of Moran type and obtain a diffusion with the same generator $\mathfrak{A}^J$. See the Appendix.

The ladder model of Ohta and Kimura [16], [15] considers a countably infinite set of types, which may be interpreted as electrophoretically detectable states at a gene locus. In the ladder model, selective neutrality is assumed; *i.e.*, the type does not affect an individual's chances in life. We discuss the ladder model in §8.

In this paper, we allow the set S of possible types to be a compact metric space. In particular, S may be a finite interval on the line R^1, a compact subset

Indiana University Mathematics Journal ©, Vol. 28, No. 5 (1979)

of euclidean R^D for some $D > 1$, or a torus. If the type x is a real number ($S \subset R^1$), then x is often called a quantitative character. We consider a model which is an infinite-dimensional analogue of the diffusion approximation on the finite-dimensional simplex Σ^J. Any possible distribution of types in the population is described by a measure μ on S of total measure $\mu(S) = 1$. For any Borel set $A \subset S$, the frequency of individuals (or gametes) with type $x \in A$ is then $\mu(A)$.

Let $\mathcal{M}(S)$ denote the set of such measures μ, with the w^*-topology (§3). To study the evolution over time of the type distribution, we consider stochastic processes with $\mathcal{M}(S)$ as state space. These processes are required to be solutions to a certain family of martingale problems posed in §5. The associated generator $\mathcal{G}$ is described in formula (5.2). This provides an analogue of the diffusion approximation on the finite-dimensional simplex Σ^J with generator $\mathfrak{A}^J$ in (2.1), already mentioned. Our main results concern existence and uniqueness of solutions to martingale problems, as well as Markov and Feller properties of the associated measure-valued processes.

Our work was inspired, in part, by a paper by Dawson [6] on measure-valued population processes as models of rescaled branching diffusions. In [6], Dawson used semigroup methods. A more recent paper by Dawson [7] uses martingale methods.

A key role in our model is played by an operator $\mathcal{L}$, which describes a linear, deterministic mechanism for change of type. We call this mechanism "mutation". In a *purely formal* way one can introduce $\mathcal{L}$ as follows. Suppose, for instance, that $S \subset R^D$. For the moment suppose that mutation is the only mechanism affecting the change in type distribution. Moreover, suppose that the type distribution μ has a spatial density $p(x, t)$, with

$$\mu(A, t) = \int_A p(x, t)dx$$

the frequency of types $x \in A$ at time t. Moreover, suppose that the density $p(x, t)$ obeys an equation

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \qquad x \in S, t \geq 0, \tag{1.1}$$

where $\mathcal{L}^*$ is the formal adjoint to $\mathcal{L}$. Conditions on p at the boundary ∂S are imposed such that $\int_S \mathcal{L}^* p dx = 0$. This is needed to insure that $\int_S p dx = 1$ for all $t \geq 0$. If corresponding boundary conditions on $\mathcal{L}$ are imposed, then the weak form of (1.1) is

$$\frac{d}{dt}\langle \alpha, \mu \rangle = \langle \mathcal{L}\alpha, \mu \rangle \tag{1.2}$$

for every function α in the domain of $\mathcal{L}$. Here

$$(1.3) \qquad \langle \alpha, \mu \rangle = \int_S \alpha(x) d\mu(x)$$

denotes the scalar product. Unlike (1.1), the weak form (1.2) involves the measure μ, and not its density p.

For instance, $\mathscr{L} = \Delta$ = Laplace operator provides a continuous version of mutational effects in the Ohta-Kimura ladder model. The zero normal derivative condition $\partial\alpha/\partial\nu = 0$ may be imposed on ∂S. If S is a D-dimensional rectangle, one can also impose periodic boundary conditions which make S effectively a D-dimensional torus. Further examples for $\mathscr{L}$ are considered in §6 and §8.

In addition to mutation, we allow chance fluctuations in the type distribution (called "random genetic drift" by population geneticists). The type distribution will then change according to a measure-valued stochastic process, which we denote by $Y(t)$ (§4). Given a time t, $Y(t)$ is a $\mathscr{M}(S)$-valued random variable. One might ask $Y(t)$ to have a (random) spatial density $p(x, t)$, which should obey a suitable stochastic perturbation of (1.1). However, it seems that these spatial densities do not exist, except perhaps in dimension $D = 1$. For $D > 1$ a density formulation leads to unresolvable technical difficulties, as can be seen from the work of Dawson [6] on continuous branching processes. Following Dawson, we work with the measure-valued process $Y(t)$ and not with its nonexistent spatial density.

Theorem 1 (§7) gives a result about existence of a solution to the basic martingale problem, posed in §5. Our method is to introduce, as an approximation, finite subsets $S^1, S^2, \cdots$ of the space S of types, and corresponding discretizations $\mathscr{L}^1, \mathscr{L}^2, \cdots$ of the mutation operator $\mathscr{L}$. For each $r = 1, 2, \cdots$ the discretized problem deals with measures on the finite dimensional $\mathscr{M}(S^r)$, or equivalently with frequency vectors in the corresponding simplex Σ^{J^r}. Ethier's finite-dimensional results [8] give a solution to each discretized martingale problem. We get Theorem 1 by verifying a Prokhorov-type compactness condition and passing to the limit.

In §8, we suppose that changes in the type distribution $Y(t)$ are due to mutation and random genetic drift only ("selective neutrality"). Moreover, a condition (called property (U)) is imposed on the mutation operator $\mathscr{L}$. Property (U) states, in effect, that a certain system of ordinary differential equations satisfied by moments has a unique solution with given initial data. The moments in question, have the form $EF(\langle\beta_1, Y(t)\rangle, \cdots, \langle\beta_K, Y(t)\rangle)$ with F any polynomial in K variables and any $\beta_1, \cdots, \beta_K$ in the domain of $\mathscr{L}$, $K = 1, 2, \cdots$. It is shown that the solution of the martingale problem is unique (Theorem 3). Moreover, the Markov and Feller properties hold for the measure-valued process $Y(t)$. A sufficient condition for property (U) is that $\mathscr{L}$ be diagonalizable, in the sense that eigenfunctions of $\mathscr{L}$ span $C(S)$. More generally, property (U) holds if $\mathscr{L}$ is triangulizable in the sense defined in §8.

In §9, we consider the "non-neutral" case when natural selection acts. Existence of a solution to the martingale problem is contained in Theorem 1 mentioned above. The method used to obtain uniqueness in §8 is no longer valid. However, uniqueness follows from the result in §8 for the selectively neutral case together with a version of the Cameron-Martin-Girsanov formula due to Dawson [7].

The authors wish to thank T. Kurtz and G. Papanicolaou for helpful comments. Particular thanks are due to D. Dawson, whose suggestions led to significant improvements in an earlier version of §8 and §9.

2. Finite set of possible types. In this section, we review some results of Ethier [8] needed later. Suppose that the possible types belong to a finite set $\tilde{S}$ with J elements:

$$\tilde{S} = \{x_1, \cdots, x_J\}.$$

Any measure μ on $\tilde{S}$ is of the form $\mu = \sum_{i=1}^{J} p_i\delta_{x_i}$ where δ_{x_i} is a unit atomic measure at x_i. Let $\tilde{\mathcal{M}} = \mathcal{M}(\tilde{S})$ be the set of such μ which satisfy $\mu \geq 0$ and $\mu(\tilde{S}) = 1$. The vector $p = (p_1, \cdots, p_J)$ determines μ. Thus, $\tilde{\mathcal{M}}$ may be identified with the $(J - 1)$-dimensional simplex

$$\Sigma^J = \left\{p: p_i \geq 0, i = 1, \cdots, J, \sum_{i=1}^{J} p_i = 1\right\}.$$

The component p_i represents the frequency of type x_i in the population.

We suppose that the vector $p(t)$ of gene frequencies varies with time t according to a diffusion process on Σ^J with the following generator $\mathfrak{A}^J$. Let $g_1, \cdots, g_J$ be functions in $C^\infty(\Sigma^J)$ *i.e.* real-valued infinitely differentiable functions on Σ^J. We suppose that $\sum_{j=1}^{J} g_j(p) = 0$ for all $p \in \Sigma^J$, and that $g_j(p) \geq 0$ whenever $p_j = 0$. Let $\mathfrak{A}^J$ be the second order partial differential operator, defined for any $f \in C^2(\Sigma^J)$ by

$$\mathfrak{A}^J f(p) = \sum_{j=1}^{J} g_j(p) f_{p_j} + \sum_{i,j=1}^{J} (p_i\delta_{ij} - p_ip_j) f_{p_ip_j}, \tag{2.1}$$

where δ_{ij} is the Kronecker symbol. Ethier showed that $\mathfrak{A}^J$ generates a diffusion process $p(t)$ on Σ^J by solving the following martingale problem.

Fix $T > 0$ and consider the following "canonical" sample space $(\Omega^J, \mathcal{F}^J)$, where

$$\Omega^J = C([0, T]; \Sigma^J)$$

$$\mathcal{F}^J = \mathcal{B}(\Omega^J) = \{\text{Borel subsets of } \Omega^J\}.$$

We consider "canonical" processes $p(t)$ on $(\Omega^J, \mathcal{F}^J)$, namely, $p(t)(\omega) = \omega(t)$ for all $\omega \in \Omega^J$. Let $\mathcal{F}_t^J$ be the σ-algebra generated by $\{\omega(s): 0 \leq s \leq t\}$. In particular, $\mathcal{F}^J = \mathcal{F}_T^J$. Given $f \in C^2(\Sigma^J)$, let

(2.2) $$M_f(t) = f[p(t)] - f[p(0)] - \int_0^t \mathfrak{A}^J f[p(v)]dv.$$

The martingale problem is as follows: given $p^0 \in \Sigma^J$, find a probability measure P^J on Ω^J such that $P^J(p(0) = p^0) = 1$ and $M_f(t)$ is a $(\{\mathcal{F}_t^J\}, P^J)$ martingale for each $f \in C^2(\Sigma^J)$. Ethier showed existence and uniqueness for the martingale problem. (A uniqueness result is needed to establish that one has a Markov, Feller process.) Actually, in [8] Ethier used instead of Σ^J the simplex

$$\Sigma_1^J = \left\{(p_1, \cdots, p_{J-1}): p_i \geq 0, i = 1, \cdots, J-1, \sum_{i=1}^{J-1} p_i \leq 1\right\}.$$

His results are easily restated in the form quoted above, by projecting Σ^J onto Σ_1^J. Ethier obtained $p(t)$ as the solution to stochastic differential equations

(2.3) $$dp_j = g_j(p)dt + \sum_{\nu=1}^{J-1} \sigma_{j\nu}(p)dw_\nu, \qquad j = 1, \cdots, J-1$$

where in (2.3) $p_J = 1 - (p_1 + \cdots + p_{J-1})$, $w_1, \cdots, w_{J-1}$ are independent Brownian motions, and

(2.4) $$p_i\delta_{ij} - p_ip_j = \sum_{\nu=1}^{J-1} \sigma_{i\nu}(p)\sigma_{j\nu}(p), \qquad i, j \leq J-1.$$

Let

$$\sigma_{J\nu} = -\sum_{j=1}^{J-1} \sigma_{j\nu}.$$

Since $g_J = -\sum_{j=1}^{J-1} g_j$, (2.3) also holds for $j = J$. Moreover, (2.4) holds when $i = J$ or $j = J$. From the Itô stochastic differential rule,

$$M_f(t) = \sum_{\nu=1}^{J-1} \int_0^t \zeta_\nu dw_\nu, \qquad \zeta_\nu = \sum_{j=1}^{J} f_{p_j}\sigma_{j\nu}.$$

Let $\langle\langle M_f(t)\rangle\rangle$ denote the increasing process associated with the martingale $M_f(t)$. It has the property that $\langle\langle M_f(0)\rangle\rangle = 0$, and $M_f^2(t) - \langle\langle M_f(t)\rangle\rangle$ is a martingale [13]. We have

(2.5) $$\langle\langle M_f(t)\rangle\rangle = \sum_{\nu=1}^{J-1} \int_0^t \zeta_\nu^2(v)dv,$$
$$\langle\langle M_f(t)\rangle\rangle = \sum_{i,j=1}^{J} \int_0^t \{p_i(v)\delta_{ij} - p_i(v)p_j(v)\}f_{p_i}[p(v)]f_{p_j}[p(v)]dv.$$

Population genetics interpretation of (2.1). Let the coefficients g_j of the first order partial derivatives f_{p_j} in (2.1) have the form

(2.6) $$g_j(p) = L_j(p) + h_j(p), \qquad j = 1, \cdots, J,$$

where

$$L_j(p) = \sum_{i=1}^{J} p_i\theta_{ij},$$
(2.7)
$$\theta_{ij} \geq 0 \quad \text{for} \quad i \neq j,\ \theta_{ii} = -\sum_{j \neq i} \theta_{ij}.$$

For $i \neq j$, θ_{ij} represents a rate of change of type x_i to type x_j. We call θ_{ij} a "mutation rate", measured on the t time scale. The functions $h_j(p)$ will be chosen later to describe the effect of natural selection. See formula (9.1).

The terms involving the second order partial derivatives $f_{p_ip_j}$ in (2.1) come from chance replacement of individuals by new ones after random mating. In population genetics literature this effect is called "random genetic drift". In the Appendix we review how this effect arises, by obtaining $\mathfrak{A}^J f$ as the limit of $\mathfrak{A}_N^J f$ as $N \to \infty$, where $\mathfrak{A}_N^J$ is the generator of a certain continuous time Markov chain (a Moran-type model).

Sato [20] recently used a martingale method to justify the diffusion limit for a somewhat different model of Gillespie with J types. In that model natural selection enters through a device of "culling" to maintain a fixed total population size.

3. The space $\mathcal{M} = \mathcal{M}(S)$, S compact metric. Now let the space S of types be a compact metric space. Extensions of the results to certain locally compact spaces (*e.g.*, $S = R^D$) will be considered elsewhere. Let

$$\mathcal{B}(S) = \{\text{Borel subsets of } S\}$$

$$C(S) = \{\text{continuous real-valued functions on } S\}$$

$$\mathcal{M} = \mathcal{M}(S) = \{\text{measures } \mu \text{ on } \mathcal{B}(S) \colon \mu \geq 0,\ \mu(S) = 1\}.$$

We denote by $\| \ \|$ the sup norm on $C(S)$ and by $\langle \alpha, \mu \rangle$ the scalar product of $\alpha \in C(S)$ and a measure μ:

$$\langle \alpha, \mu \rangle = \int_S \alpha(x) d\mu(x).$$

We give $\mathcal{M}$ the weak* topology. Then $\mathcal{M}$ is compact, and metrizable by the following metric $d(\mu, \nu)$. We say that $\beta_1, \beta_2, \cdots$ is a *basis* for $C(S)$ if finite linear combinations of these β_k are dense in $C(S)$. Let $\beta_1, \beta_2, \cdots$ be a basis for $C(S)$ with $\|\beta_k\|$ bounded, and let

$$d(\mu, \nu) = \sum_{k=1}^{\infty} 2^{-k} |\langle \beta_k, \mu \rangle - \langle \beta_k, \nu \rangle|.$$

Let $\tilde{S} = \{x_1, \cdots, x_J\}$ be a finite subset of S, and $\eta > 0$ a number such that dist $(x, \tilde{S}) < \eta$ for all $x \in S$. There is a partition of unity $\alpha_1, \cdots, \alpha_J$ for S with the following properties: $\alpha_i \in C(S)$,

$$\alpha_i(x_j) = \delta_{ij},\ \text{diam}\{\text{spt}\, \alpha_i\} \leq \eta, \qquad i, j = 1, \cdots, J,$$

where spt α is the support of α. In fact, we may take

$$\alpha_i = \gamma_i \Big/ \sum_{j=1}^{J} \gamma_j,$$

$$\gamma_i(x) = \left\{ \prod_{j \neq i} \operatorname{dist}(x_j, x) \right\} \psi[\operatorname{dist}(x_i, x)],$$

$$\psi(r) > 0 \quad \text{for} \quad 0 \leq r < \eta, \qquad \psi(r) = 0 \quad \text{for} \quad r \geq \eta.$$

Let

$$G(\mu) = \sum_{i=1}^{J} \langle \alpha_i, \mu \rangle \delta_{x_i}.$$

Then $G: \mathcal{M}(S) \to \mathcal{M}(\hat{S})$ and $G(\mu) = \mu$ for all $\mu \in \mathcal{M}(\hat{S})$. It is easily shown that:

Lemma 1. *Given $\epsilon > 0$ there exists $\eta > 0$ such that $d[G(\mu), \mu] < \epsilon$ for all $\mu \in \mathcal{M}(S)$, if $\hat{S}$, α_i, and η are as above.*

Let

$$C(\mathcal{M}) = \{\text{continuous real-valued functions on } \mathcal{M}\}$$

with the sup norm $\| \ \|$. Let $\beta_1, \beta_2, \cdots$ be a basis for $C(S)$. Then functions depending on a finite number of scalar products $\langle \beta_k, \mu \rangle$ are dense in $C(\mathcal{M})$. In fact, let $\mathcal{D}_1$ denote the subspace of $C(\mathcal{M})$ spanned by all functions ϕ of the form

$$\phi(\mu) = F(\langle \beta_1, \mu \rangle, \cdots, \langle \beta_K, \mu \rangle)$$

with $K = 1, 2, \cdots$ and $F(z_1, \cdots, z_K) \in C(R^K)$. The space $\mathcal{D}_1$ depends, of course, on the sequence $\beta_1, \beta_2, \cdots$.

Lemma 2. *$\mathcal{D}_1$ is dense in $C(\mathcal{M})$.*

Proof. Consider $\hat{S}$, $\alpha_1, \cdots, \alpha_J$, and G as above. Given $\phi \in C(\mathcal{M})$ let

$$f(p) = \phi\left(\sum_{i=1}^{J} p_i \delta_{x_i} \right).$$

Then $f \in C(\Sigma^J)$; and by taking $p_i = \langle \alpha_i, \mu \rangle$ we get

$$(\phi \circ G)(\mu) = f(\langle \alpha_1, \mu \rangle, \cdots, \langle \alpha_J, \mu \rangle).$$

By Lemma 1, $\|\phi \circ G - \phi\| \to 0$ as $\eta \to 0$. Moreover, each α_i can be approximated in $\| \ \|$ by finite sums $\sum_{k=1}^{K} c_{ik} \beta_k$. We take

$$F(z_1, \cdots, z_K) = f\left(\sum_{k=1}^{K} c_{1k} z_k, \cdots, \sum_{k=1}^{K} c_{Jk} z_k \right).$$

This proves Lemma 2.

Note. Since $\mathcal{M}$ is compact, the set of vectors $(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)$ is a compact subset of R^K. In Lemma 2, it suffices to consider $F \in C^2(R^K)$; in fact, by the Weierstrass approximation theorem, it suffices to take F any polynomial.

4. The canonical sample space Ω. Fix $T > 0$ and consider the space of continuous, $\mathcal{M}$-valued functions

$$\Omega = C([0, T]; \mathcal{M}).$$

We give Ω the uniform metric: for $y_1, y_2 \in \Omega$.

$$d(y_1, y_2) = \max_{[0,T]} d[y_1(t), y_2(t)].$$

A sequence $y^1, y^2, \cdots$ converges to y in Ω if and only if $\|\langle\alpha, y^n\rangle - \langle\alpha, y\rangle\| \to 0$ as $n \to \infty$ for each $\alpha \in C(S)$, where $\| \quad \|$ is the sup norm in $C[0, T] = C([0, T]; R^1)$. It is easy to prove the following sufficient condition for a set $\Gamma \subset \Omega$ to be compact. Let $\beta_1, \beta_2, \cdots$ be a basis for $C(S)$ with $\|\beta_k\|$ bounded.

Lemma 3. *Let* $\Gamma = \{y \in \Omega: \langle\beta_k, y\rangle \in \Gamma_k, k = 1, 2, \cdots\}$, *where* $\Gamma_k \subset C[0, T]$ *is compact for each* k. *Then* Γ *is compact.*

Let

$$\mathcal{F} = \mathcal{B}(\Omega) = \{\text{Borel subsets of } \Omega\}.$$

We consider processes $Y(t)$ on the "canonical" sample space $(\Omega, \mathcal{F})$, namely,

$$Y(t)(y) = y(t), \qquad \text{for all } y \in \Omega.$$

Note that $y(t) \in \mathcal{M}(S)$ for each t and y. Thus, $Y(t)$ is a measure-valued process, with $\mathcal{M}(S)$ as its state space. Let $\mathcal{F}_t$ be the σ-algebra generated by $\{y(s): 0 \le s \le t\}$; in particular, $\mathcal{F}_T = \mathcal{F}$.

In the next section we shall describe a generator $\mathcal{G}$ for the Y-process, and shall pose the corresponding martingale problem.

5. Martingale problem. As explained in §1, our objective is to describe a stochastic model for the evolution of the distribution of types in S, as influenced by various mechanisms including mutation and random genetic drift.

Let $\mathcal{S}$ be a dense subset of $C(S)$ and $\mathcal{L}: \mathcal{S} \to C(S)$ a linear operator. We require that $\mathcal{S}$ contain the constant functions and that $\mathcal{L}c = 0$ for any constant function c. The operator $\mathcal{L}$ can be interpreted as describing mutational effects, as already indicated in §1. We shall require that $\mathcal{L}$ be positively discretizable in the sense to be defined in §6. This assumption is used in §7 to prove existence of a solution to the martingale problem. In §8 we make a further assumption about $\mathcal{L}$ in order to get a uniqueness result, and the Markov and Feller properties. We remark that both these assumptions hold trivially if mutation is absent ($\mathcal{L} \equiv 0$). This is the case of "pure random genetic drift", discussed at the end of §8.

Let $H \in C[\mathcal{M} \times C(S)]$ be a function with the following property: given any finite set $\tilde{S} \subset S$, namely, $\tilde{S} = \{x_1, \cdots, x_J\}$ there exist $h_1, \cdots, h_J \in C^\infty(\Sigma^J)$ such that

$$H(\mu, \beta) = \sum_{j=1}^{J} h_j(p)\beta(x_j) \tag{5.1}$$

for each $\mu = \sum_{i=1}^{J} p_i\delta_{x_i}$ in $\mathcal{M}(\tilde{S})$ and $\beta \in C(S)$. If only mutation and random genetic drift act, then $H = 0$. This will be assumed in §8. In §9 we make a specific choice for H corresponding to selective advantages among the various types. See formula (9.2). Let

$$\begin{aligned} \mathcal{D}_0 &= \{\phi\colon \phi(\mu) = F(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle), \\ &\beta_1, \cdots, \beta_K \in \mathcal{S}, F \in C^2(R^K), K = 1, 2, \cdots\}, \\ \mathcal{D} &= \text{linear subspace of } C(\mathcal{M}) \text{ spanned by } \mathcal{D}_0. \end{aligned}$$

We may take $\beta_1, \cdots, \beta_K$ elements of a basis for $C(S)$. If $\mathcal{D}_1$ is as in §3, then $\bar{\mathcal{D}}_1 \subset \bar{\mathcal{D}}$ where $^-$ denotes closure. By Lemma 2, $\mathcal{D}$ is dense in $C(\mathcal{M})$. We define $\mathcal{G}\phi$ for $\phi \in \mathcal{D}$ by requiring that $\mathcal{G}\phi$ is linear and that for $\phi \in \mathcal{D}_0$:

$$\begin{aligned} \mathcal{G}\phi(\mu) &= \sum_{k=1}^{K} (\langle\mathcal{L}\beta_k, \mu\rangle + H(\mu, \beta_k))F_{z_k}(\cdots) \\ &+ \sum_{k,\ell=1}^{K} (\langle\beta_k\beta_\ell, \mu\rangle - \langle\beta_k, \mu\rangle\langle\beta_\ell, \mu\rangle)F_{z_k z_\ell}(\cdots). \end{aligned} \tag{5.2}$$

Here $\cdots$ stands for the vector $z = (z_1, \cdots, z_K)$ with $z_k = \langle\beta_k, \mu\rangle$. In §6 we shall motivate this choice for $\mathcal{G}\phi$ and will show that $\mathcal{G}\phi$ is well defined. (see formula (6.5).) For each $\phi \in \mathcal{D}$ let

$$M_\phi(t) = \phi[Y(t)] - \phi[Y(0)] - \int_0^t \mathcal{G}\phi[Y(v)]dv, \tag{5.3}$$

where $Y(t)$ is the "canonical" process defined at the end of §4.

Martingale problem. Given $\mu^0 \in \mathcal{M}$, find a probability measure P on $\mathcal{F}$ such that:

(a) $P(Y(0) = \mu^0) = 1$;
(b) $M_\phi(t)$ is a $(\{\mathcal{F}_t\}, P)$-martingale for each $\phi \in \mathcal{D}$;
(c) The increasing process satisfies for each $\phi \in \mathcal{D}_0$ $\langle\langle M_\phi(t)\rangle\rangle = \int_0^t \psi_\phi[Y(v)]dv$, where

$$\psi_\phi(\mu) = \sum_{k,\ell=1}^{K} (\langle\beta_k\beta_\ell, \mu\rangle - \langle\beta_k, \mu\rangle\langle\beta_\ell, \mu\rangle)F_{z_k}(\cdots)F_{z_\ell}(\cdots). \tag{5.4}$$

We call μ^0 the *initial state* for the measure-valued process $Y(t)$ on the probability space $(\Omega, \mathcal{F}, P)$. We call $\mathcal{G}$ the *generator* associated with the martingale problem. Under suitable restrictions it will be shown later (§8, §9) that $\mathcal{G}$ is indeed the generator of a Markov process corresponding to the martingale problem.

6. Discretization of $\mathcal{L}$. Let $\tilde{S} \subset S$ be a finite set, namely, $\tilde{S} = \{x_1, \cdots, x_J\}$; and let

$$\tilde{\mathcal{M}} = \mathcal{M}(\tilde{S}) = \{\mu \in \mathcal{M}(S)\colon \operatorname{spt} \mu \subset \tilde{S}\}.$$

Each function $\alpha \in C(S)$ defines by restriction to $\tilde{S}$ a function on $\tilde{S}$. In place of the operator $\mathcal{L}$ in §5, let us consider a linear operator $\tilde{\mathcal{L}}$ such that

$$\tilde{\mathcal{L}}\alpha(x_i) = \sum_{j=1}^{J} \theta_{ij}\alpha(x_j),$$

$$\theta_{ij} \geq 0 \quad \text{for} \quad i \neq j,\ \theta_{ii} = -\sum_{j \neq i} \theta_{ij}.$$

By (2.7) we have for $\mu = \sum_{i=1}^{J} p_i\delta_{x_i}$ in $\tilde{\mathcal{M}}$ and $\alpha \in C(S)$

$$\langle \tilde{\mathcal{L}}\alpha, \mu\rangle = \sum_{j=1}^{J} L_j(p)\alpha(x_j). \tag{6.1}$$

The constants θ_{ij} are to be chosen so that $\tilde{\mathcal{L}}$ can be regarded as a discretization of $\mathcal{L}$; see the definition of positively discretizable operator below.

The solution to the martingale problem in §2 can be rewritten as a solution to the corresponding martingale problem for $\tilde{\mathcal{M}}$-valued processes, as follows. Consider those $\phi \in C(\mathcal{M})$ for which there is an $f \in C^2(\Sigma^J)$ such that

$$\phi\left(\sum_{i=1}^{J} p_i\delta_{x_i}\right) = f(p) \qquad \text{for all } p \in \Sigma^J.$$

Let $\tilde{\mathcal{D}}$ denote the space of all such ϕ. For $\phi \in \tilde{\mathcal{D}}$ and $\mu \in \tilde{\mathcal{M}}$ we define $\tilde{\mathcal{G}}\phi(\mu)$ by

$$\tilde{\mathcal{G}}\phi(\mu) = \mathfrak{A}^J f(p),\ \mu = \sum_{i=1}^{J} p_i\delta_{x_i},$$

with $\mathfrak{A}^J$ defined by (2.1), (2.6). If $\phi(\mu) = F(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)$ is in $\mathcal{D}_0$, then $\phi \in \tilde{\mathcal{D}}$ and the corresponding f is

$$f(p) = F\left(\sum_{i=1}^{J} \beta_1(x_i)p_i, \cdots, \sum_{i=1}^{J} \beta_K(x_i)p_i\right).$$

We compute the partial derivatives f_{p_j}, $f_{p_j p_j}$ by the chain rule, and use (2.1), (2.6), (5.1) and (6.1) to find that for $\mu \in \tilde{\mathcal{M}}$

$$\tilde{\mathcal{G}}\phi(\mu) = \sum_{k=1}^{K} (\langle\tilde{\mathcal{L}}\beta_k, \mu\rangle + H(\mu, \beta_k))F_{z_k}(\cdots)$$

(6.2)
$$+\sum_{k,\ell=1}^{K}(\langle\beta_k\beta_\ell,\mu\rangle-\langle\beta_k,\mu\rangle\langle\beta_\ell,\mu\rangle)F_{z_kz_\ell}(\cdots).$$

Note that $\tilde{\mathscr{G}}\phi$ has exactly the same form as $\mathscr{G}\phi$ in (5.2), for $\phi\in\mathscr{D}_0$, except that $\mathscr{L}$ is replaced by $\tilde{\mathscr{L}}$. Since $\mathscr{D}_0\subset\tilde{\mathscr{D}}$, the linear space $\mathscr{D}$ spanned by $\mathscr{D}_0$ is also contained in $\tilde{\mathscr{D}}$.

Let
$$\Psi(\omega)=\sum_{i=1}^{J}\omega_i\delta_{x_i},\qquad \omega=(\omega_1,\cdots,\omega_J)\in\Omega^J,$$
with $\Omega^J=C([0,T];\Sigma^J)$ as in §2. Given $\mu^0=\sum_{i=1}^{J}p_i^0\delta_{x_i}$ in $\tilde{\mathscr{M}}$, with corresponding $p^0\in\Sigma^J$, let P^J be a solution to the martingale problem in §2. Let $\tilde{P}=\Psi(P^J)$. Then $\tilde{P}$ is a probability measure on Ω with
$$\tilde{P}(Y(t)\in\tilde{\mathscr{M}},\ 0\le t\le T)=1.$$

For $\phi\in\tilde{\mathscr{D}}$ let

(6.3)
$$\tilde{M}_\phi(t)=\phi[Y(t)]-\phi[Y(0)]-\int_0^t\tilde{\mathscr{G}}\phi[Y(v)]dv.$$

Then $\tilde{M}_\phi(t)(\Psi(\omega))=M_f(t)(\omega)$ for all $\omega\in\Sigma^J$. Therefore, $\tilde{P}$ solves the following martingale problem:

(a) $\tilde{P}(Y(0)=\mu^0)=1$;
(b) $\tilde{M}_\phi(t)$ is a $(\{\mathscr{F}_t\},\tilde{P})$-martingale for each $\phi\in\tilde{\mathscr{D}}$;
(c) The increasing process satisfies, for each $\phi\in\mathscr{D}_0$,

(6.4)
$$\langle\langle\tilde{M}_\phi(t)\rangle\rangle=\int_0^t\psi_\phi[Y(v)]dv,$$

where ψ_ϕ is as in (5.4).

Equation (6.4) is obtained from (2.5) and the chain rule for the partial derivatives f_{p_i}.

Let us now consider a sequence of finite subsets of S, denoted by $S^1,S^2,\cdots$,
$$S^r=\{x_1^r,\cdots,x_{J^r}^r\},\qquad r=1,2,\cdots.$$
We suppose that dist $(x,S^r)<\eta_r$ for all $x\in S$, where $\eta_r\to0$ as $r\to\infty$. Let $\mathscr{M}^r=\mathscr{M}(S^r)$. For each $r=1,2,\cdots$ define $G^r:\mathscr{M}\to\mathscr{M}^r$ as in §3, with $G^r(\mu)=\mu$ for all $\mu\in\mathscr{M}^r$. By Lemma 1, G^r converges uniformly to the identity as $r\to\infty$.

Definition. The operator $\mathscr{L}$ is *positively discretizable* if there exist S^1, S^2, $\cdots$ as above and $\mathscr{L}^1,\mathscr{L}^2,\cdots$, such that:

(i) $\mathscr{L}^r\alpha(x_i^r)=\sum_{j=1}^{J^r}\theta_{ij}^r\alpha(x_j^r)\qquad$ for all $\alpha\in C(S)$,

$\theta_{ij}^r\ge0$ for $i\ne j$, $\theta_{ii}^r=-\sum_{j\ne i}\theta_{ij}^r$, $r=1,2,\cdots$.

(ii) $\max_{x \in S^r} |\mathscr{L}^r\alpha(x) - \mathscr{L}\alpha(x)| \to 0$ as $r \to \infty$, for all $\alpha \in \mathscr{L}$.

The operator $\mathscr{L}^r$ is called a discretization of $\mathscr{L}$. In the previous notation, $\tilde{\mathscr{L}} = \mathscr{L}^r$ when $\tilde{S} = S^r$.

Example 1. Let $S = [0, 1]$, the unit interval in R^1, and $\mathscr{L}$ the second order differential operator

$$\mathscr{L}\alpha = a(x)\alpha'' + b(x)\alpha', \qquad a(x) > 0,$$

with $a, b \in C^1(S)$. Let $\mathscr{S} = \{\alpha \in C^2(S): \alpha'(x) = 0 \text{ for } x = 0, 1\}$. Let $S^r = \{ir^{-1}: i = 0, 1, \cdots, r\}$ and let $\mathscr{L}^r$ be the usual discretization of $\mathscr{L}$, with α', α'' replaced by first and second order difference quotients. Since $\alpha'(x) = 0$ at the endpoints, when $x = 0$ we can replace in these difference quotients $\alpha(-r^{-1})$ by $\alpha(r^{-1})$ and $\alpha(1 + r^{-1})$ by $\alpha(1 - r^{-1})$ when $x = 1$. The condition $\theta^r_{ij} \geq 0$ when $i \neq j$ holds for $r \geq r_0$; if we replace r by $r + r_0$, then (i) and (ii) hold for $r = 1, 2, \cdots$.

Example 2. Let S be a rectangular region in R^D, namely, $S = I^1 \times \cdots \times I^D$, where $I^1, \cdots, I^D$ are 1-dimensional intervals. We may suppose that each $I^j = [0, 1]$. We denote points of S by $x = (x^1, \cdots, x^D)$. Let

$$\mathscr{L} = \mathscr{L}_1 + \cdots + \mathscr{L}_D,$$

$$\mathscr{L}_i = a_i(x_i) \frac{\partial^2}{\partial x_i^2} + b_i(x_i) \frac{\partial}{\partial x_i}, \qquad a_i(x_i) > 0$$

as in Example 1, and

$$\mathscr{S} = \left\{\alpha \in C^2(S): \frac{\partial \alpha}{\partial x^j} = 0 \text{ when } x^j = 0, 1, j = 1, \cdots, D\right\}.$$

For S^r we take the lattice obtained by discretizing each I^j and replacing derivatives by difference quotients, as in Example 1.

We may think of Example 2 as applying to a situation where the type x is controlled by several gene loci (D in number), and x^j is an effect contributed by the j^{th} locus. Then $\mathscr{L}_j$ represents the action of mutation at the j^{th} locus.

Example 3. Let S be a D-dimensional torus, which we may identify with the rectangular region in Example 2 with periodic boundary conditions; and take $\mathscr{L}$ as in Example 2.

Let $\mathscr{G}^r$ denote the generator of the discretized process, *i.e.*, $\mathscr{G}^r = \tilde{\mathscr{G}}$ in (6.2) when $\mathscr{L}^r = \tilde{\mathscr{L}}$. For $\mu \in \mathscr{M}^r$ and $\alpha \in \mathscr{S}$

$$|\langle \mathscr{L}^r\alpha, \mu\rangle - \langle \mathscr{L}\alpha, \mu\rangle| \leq \max_{x \in S^r} |\mathscr{L}^r\alpha(x) - \mathscr{L}\alpha(x)|.$$

The right side tends to 0 as $r \to \infty$. If $\phi(\mu) = F(\langle \beta_1, \mu\rangle, \cdots, \langle \beta_K, \mu\rangle)$ is in $\mathscr{D}_0$, then by (5.2) and (6.2)

$$|\mathcal{G}^r\phi(\mu) - \mathcal{G}\phi(\mu)| \le C \max_{\substack{x \in S^r \\ k = 1, \cdots, K}} |\mathcal{L}^r\beta_k(x) - \mathcal{L}\beta_k(x)|,$$

where C is a constant such that

$$\sum_{k=1}^{K} |F_{z_k}(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)| \le C.$$

Therefore, for each $\phi \in \mathcal{D}_0$,

$$\lim_{r\to\infty} \max_{\mu \in \mathcal{M}^r} |\mathcal{G}^r\phi(\mu) - \mathcal{G}\phi(\mu)| = 0. \tag{6.5}$$

Since $\mathcal{G}\phi \in C(\mathcal{M})$ and G^r tends to the identity uniformly on $\mathcal{M}$,

$$\lim_{r\to\infty} \max_{\mu \in \mathcal{M}} |\mathcal{G}\phi[G^r(\mu)] - \mathcal{G}\phi(\mu)| = 0,$$

for each $\phi \in \mathcal{D}_0$. Hence

$$\lim_{r\to\infty} \|(\mathcal{G}^r\phi) \circ G^r - \mathcal{G}\phi\| = 0 \tag{6.6}$$

for each $\phi \in \mathcal{D}_0$. Since $\mathcal{G}^r$ is linear, formula (6.6) then holds also for each $\phi \in \mathcal{D}$. This shows, in particular, that $\mathcal{G}$ is well defined as a linear operator from $\mathcal{D}$ into $C(\mathcal{M})$.

7. Existence of a solution to the martingale problem. Let us return to the martingale problem posed in §5. The purpose of this section is to prove the following.

Theorem 1. *Let $\mathcal{L}$ be positively discretizable. Then given any initial state $\mu^0 \in \mathcal{M}$ there exists a probability measure P which solves the martingale problem.*

The proof will proceed as follows. We introduce discretizations S^r, $\mathcal{L}^r$ as in §6. Moreover, we take $\mu^{0r} \in \mathcal{M}^r$ tending to μ^0 as $r \to \infty$. The martingale problem in which μ^0 is replaced by μ^{0r} as initial state and the generator $\mathcal{G}$ by $\mathcal{G}^r$ has a solution P^r (see §6, with $\tilde{S} = S^r$, $\tilde{\mathcal{G}} = \mathcal{G}^r$). We verify a Prokhorov-type compactness condition for the set $\{P^1, P^2, \cdots\}$ of measures. The measure P in Theorem 1 will be the limit of P^r as $r \to \infty$ through a subsequence. Convergence of measures is in the sense that $\int_\Omega \Phi(y)dP^r(y)$ tends to $\int_\Omega \Phi(y)dP(y)$ for all continuous bounded functions Φ.

We begin with the following lemma. Given $\alpha \in C(S)$ we use the notation $Y_\alpha = \langle\alpha, Y\rangle$.

Lemma 4. *Given $\alpha \in \mathcal{S}$ and $\epsilon > 0$ there exists a compact set $K_\alpha^\epsilon \subset C[0, T]$ such that $P^r(Y_\alpha \in K_\alpha^\epsilon) \ge 1 - \epsilon$, $r = 1, 2, \cdots$.*

Proof. Let

$$M_\alpha^r(t) = Y_\alpha(t) - \mu_\alpha^{0r} - \int_0^t \mathscr{G}^r\phi_\alpha[Y(v)]dv,$$

$$\mu_\alpha^{0r} = \langle \alpha, \mu^{0r} \rangle, \ \phi_\alpha(\mu) = \langle \alpha, \mu \rangle, \ M_\alpha^r = M_{\phi_\alpha}^r.$$

Then M_α^r is a P^r-martingale. By (6.4) and (5.4), with $K = 1$, $F(z) = z$,

$$\langle\langle M_\alpha^r(t) \rangle\rangle - \langle\langle M_\alpha^r(s) \rangle\rangle = \int_s^t \{\langle \alpha^2, Y(v) \rangle - Y_\alpha^2(v)\}dv, \qquad s \le t. \tag{7.1}$$

Given $\eta > 0$ let

$$\gamma_\alpha(\eta) = \sup \{|Y_\alpha(t) - Y_\alpha(s)|: s, t \in [0, T], |s - t| \le \eta\},$$

$$\delta_\alpha^r(\eta) = \sup \{|M_\alpha^r(t) - M_\alpha^r(s)|: s, t \in [0, T], |s - t| \le \eta\}.$$

Now $\mathscr{G}^r\phi_\alpha(\mu) = \langle \mathscr{L}^r\alpha, \mu \rangle$ for $\mu \in \mathscr{M}^r$. Since $\mathscr{L}$ is positively discretizable, $|\langle \mathscr{L}^r\alpha, \mu \rangle - \langle \mathscr{L}\alpha, \mu \rangle|$ is uniformly small on $\mathscr{M}^r$. Since $|\langle \mathscr{L}\alpha, \mu \rangle|$ is bounded on $\mathscr{M}$, $|\mathscr{G}^r\phi_\alpha(\mu)| \le C_\alpha$ for some C_α. Then

$$\gamma_\alpha(\eta) \le C_\alpha\eta + \delta_\alpha^r(\eta). \tag{7.2}$$

Since $0 \le \langle \alpha^2, Y(v) \rangle - Y_\alpha^2(v) \le 2\|\alpha\|^2$, by (7.1)

$$0 \le \langle\langle M_\alpha^r(t) \rangle\rangle - \langle\langle M_\alpha^r(s) \rangle\rangle \le 2\|\alpha\|^2(t - s), \qquad s \le t. \tag{7.3}$$

Thus $\langle\langle M_\alpha^r \rangle\rangle$ is Lipschitz with constant $2\|\alpha\|^2$. This implies that there exists $\Psi_\alpha(\epsilon, \eta)$ decreasing to 0 as $\eta \to 0$ (for fixed ϵ) with the following property: for each $\epsilon > 0$

$$P^r(\delta_\alpha^r(\eta) \le \Psi_\alpha(\epsilon, \eta) \text{ for all } \eta > 0) \ge 1 - \epsilon. \tag{7.4}$$

See [18, Chapter 2, Proposition 9], also [21, Lemma 2.2]. The set

$$K_\alpha^\epsilon = \{y \in \Omega: \gamma_\alpha(\eta) \le C_\alpha\eta + \Psi_\alpha(\epsilon, \eta) \text{ for all } \eta > 0\}$$

has the property required in Lemma 4.

Note. One can choose $\Psi_\alpha(\epsilon, \eta)$ to depend only on T, $\|\alpha\|$, C_α, ϵ, η. This observation is used in proving Lemma 5 below.

Proof of Theorem 1. We apply Lemma 4 as follows. Choose a basis $\beta_1, \beta_2, \cdots$, for $C(S)$ with $\beta_k \in \mathscr{S}$ for each $k = 1, 2, \cdots$ and $\|\beta_k\|$ bounded. Given $\epsilon > 0$, let $\epsilon_k = 2^{-k}\epsilon$ and

$$K^\epsilon = \{y \in \Omega: \langle \beta_k, y \rangle \in K_{\beta_k}^{\epsilon_k}, k = 1, 2, \cdots\}.$$

By Lemma 3 (§3), K^ϵ is compact. Moreover,

$$P^r(Y \notin K^\epsilon) \le \sum_{k=1}^\infty P^r(Y_{\beta_k} \notin K_{\beta_k}^{\epsilon_k}) \le \epsilon \sum_{k=1}^\infty 2^{-k} = \epsilon.$$

This is equivalent to the Prokhorov condition $P^r(Y \in K^\epsilon) \geq 1 - \epsilon$, $r = 1, 2, \cdots$. Therefore, a subsequence of $P^1, P^2, \cdots$ converges to a limit P. Let us again denote the subsequence by $P^1, P^2, \cdots$.

Consider any $\phi \in \mathscr{D}_0$, $\phi(\mu) = F(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)$ with $\beta_1, \cdots, \beta_K \in \mathscr{S}$, $F \in C^2(R^K)$. Then P^r-almost surely, $Y(t) \in \mathscr{M}^r$, $G^r[Y(t)] = Y(t)$, and

$$M_\phi^r(t) = (\phi \circ G^r)(Y(t)) - (\phi \circ G^r)(\mu^{0r}) - \int_0^t [(\mathscr{G}^r\phi) \circ G^r(Y(v))dv.$$

We use (5.3), (6.3), (6.6) and the facts that $\mu^{0r} \to \mu^0$, $\|\phi \circ G^r - \phi\| \to 0$ as $r \to \infty$ to get

$$\|M_\phi^r - M_\phi\| \leq \sigma_r \tag{7.5}$$

where $\sigma_r \to 0$ as $r \to \infty$. Since $M_\phi^r(t)$ is a $(\{\mathscr{F}_t\}, P^r)$-martingale for each r, $M_\phi(t)$ is a $(\{\mathscr{F}_t\}, P)$-martingale, for any $\phi \in \mathscr{D}_0$. By linearity of $\mathscr{G}$, this is true also for any $\phi \in \mathscr{D}$. Moreover, $P(Y(0) = \mu^0) = 1$. These are properties (a), (b) in the martingale problem formulation, §5. Finally, to prove (c) we have for $\phi \in \mathscr{D}_0$

$$\langle\langle M_\phi^r(t)\rangle\rangle = \int_0^t \psi_\phi[Y(v)]dv = (M_\phi^r(t))^2 + Z^r(t)$$

where $Z^r(t)$ is a $(\mathscr{F}_t, P^r)$ martingale with $Z^r(0) = 0$. By (7.5), $Z^r(t)$ tends uniformly to $Z(t)$ with

$$\int_0^t \psi_\phi[Y(v)]dv = (M_\phi(t))^2 + Z(t). \tag{7.6}$$

Moreover, $Z(t)$ is a $(\mathscr{F}_t, P)$-martingale with $Z(0) = 0$. Thus the left side of (7.6) is $\langle\langle M_\phi(t)\rangle\rangle$ as required in (c). This proves Theorem 1.

The same method gives the following lemma about dependence of the measure P solving the martingale problem on the initial state μ^0.

***Lemma* 5.** *Let P_r be a solution to the martingale problem for initial state $\mu_r^0 \in \mathscr{M}$, $r = 1, 2, \cdots$, ($\mathscr{G}$ fixed). Let $\mu_r^0 \to \mu^0$ as $r \to \infty$. Then P_r tends to a limit P for a subsequence of r. Moreover, P is a solution to the martingale problem for initial state μ^0.*

8. Uniqueness (selectively neutral case). The existence theorem in §7 was obtained using approximation and tightness arguments of rather wide applicability. However, to prove uniqueness, we shall use special methods which depend on the particular form of the generator $\mathscr{G}$. In this section, we take $H = 0$ in (5.2). Thus, for $\phi(\mu) = F(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)$ in $\mathscr{D}_0$

$$\begin{aligned} \mathscr{G}\phi(\mu) = {} & \sum_{k=1}^K \langle\mathscr{L}\beta_k, \mu\rangle F_{z_k}(\cdots) \\ & + \sum_{k,\ell=1}^K (\langle\beta_k\beta_\ell, \mu\rangle - \langle\beta_k, \mu\rangle\langle\beta_\ell, \mu\rangle)F_{z_k z_\ell}(\cdots). \end{aligned} \tag{8.1}$$

When $H = 0$ the model treats the interplay of mutation, which enters through the terms $\langle \mathscr{L}\beta_k, \mu \rangle$ multiplying F_{z_k} in (8.1), and random genetic drift which enters through the coefficients multiplying $F_{z_k z_\ell}$. This is the case of "selectively neutral" genes. In §9 natural selection is allowed to enter also.

If P is a solution to the martingale problem in §5, let us write E^P for expectation with respect to the probability measure P. Our method depends on uniqueness of moments of certain finite-dimensional distributions, namely, the uniqueness of $E^P(\langle \beta_1, Y(t) \rangle \cdots \langle \beta_K, Y(t) \rangle)$ given the initial state μ^0. For this purpose we first write a system of ordinary differential equations, (8.2), (8.3) below, which these moments obey. We assume that these equations, with given initial data, have a unique solution (uniqueness property (U)). We then show (Theorem 2) that the expectations $E^P\phi[Y(t)]$ define a semigroup on $C(\mathscr{M})$; and that the solution to the martingale problem is unique (Theorem 3). We then give two sufficient conditions (Theorems 4, 5) for property (U) to hold.

Let $u(\beta_1, \cdots, \beta_K, t)$ denote a continuous real-valued function on $C(S) \times \cdots \times C(S) \times [0, \infty)$, such that u is multilinear and symmetric in $\beta_1, \cdots, \beta_K$. Let $(x_1, \cdots, x_K) \in S \times \cdots \times S$. The notation $\mathscr{L}_{x_k}$ indicates that $\mathscr{L}$ acts in the variable x_k. Let $\mathscr{L}^*_{x_k}$ denote the (weak sense) adjoint of $\mathscr{L}_{x_k}$:

$$\mathscr{L}^*_{x_k} u(\beta_1, \cdots, \beta_K, t) = u(\beta_1, \cdots, \beta_{k-1}, \mathscr{L}\beta_k, \beta_{k+1}, \cdots, \beta_K, t),$$
$$\text{for all } \beta_1, \cdots, \beta_K \in \mathscr{S}.$$

For $K = 1$ the equation for $u(\beta, t)$ is

$$\frac{du}{dt} = \mathscr{L}^* u, \qquad \text{if } \beta \in \mathscr{S},\ t \geq 0. \tag{8.2}$$

For $K > 1$, let

$$w = 2 \sum_{k < \ell} u(\beta_1, \cdots, \beta_{k-1}, \beta_k\beta_\ell, \beta_{k+1}, \cdots, \beta_{\ell-1}, 1, \beta_{\ell+1}, \cdots, \beta_K, t).$$

The equation for $u(\beta_1, \cdots, \beta_K, t)$ is

$$\frac{du}{dt} = \sum_{k=1}^{K} \mathscr{L}^*_{x_k} u - K(K-1)u + w, \text{ if } \beta_1, \cdots, \beta_K \in \mathscr{S},\ t \geq 0. \tag{8.3}$$

Equations (8.2), (8.3) are considered with the following initial data. Given $\mu^0 \in \mathscr{M}$

$$u(\beta_1, \cdots, \beta_K, 0) = \langle \beta_1, \mu^0 \rangle \cdots \langle \beta_K, \mu^0 \rangle. \tag{8.4}$$

If P solves the martingale problem in §5, then a solution to (8.2)-(8.4) is

$$u(\beta_1, \cdots, \beta_K, t) = E^P(\langle \beta_1, Y(t) \rangle \cdots \langle \beta_K, Y(t) \rangle). \tag{8.5}$$

This corresponds to taking in (8.1) $F(z) = z_1 \cdots z_K$, a monomial in $z_1, \cdots, z_K$.

Let us now assume that the operator $\mathscr{L}$ has:

Uniqueness Property (U). Equations (8.2), (8.3) and the initial data (8.4) uniquely determine $u(\beta_1, \cdots, \beta_K, t)$ for all $K = 1, 2, \cdots, \beta_1, \cdots, \beta_K \in C(S)$, $t \geq 0$.

By property (U) and (8.5), $E^P\phi[Y(t)]$ is unique for $\phi(\mu) = F(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle)$ if $F(z_1, \cdots, z_K)$ is any polynomial. The same then holds for any $F \in C(R^K)$, by the Weierstrass approximation theorem and the fact that $\{(\langle\beta_1, \mu\rangle, \cdots, \langle\beta_K, \mu\rangle): \mu \in \mathcal{M}\}$ is a bounded subset of R^K. Thus, given μ^0, $E^P\phi[Y(t)]$ is unique for all $\phi \in \mathcal{D}_0$, hence for all $\phi \in C(\mathcal{M})$. We may then also denote $E^P\phi[Y(t)]$ by $E_{\mu^0}\phi[Y(t)]$. In §7 a solution to the martingale problem was given on a finite interval $[0, T]$. However, included in property (U) is the condition that $E_{\mu^0}\phi[Y(t)]$ does not depend on the final time $T \geq t$. Let

$$T_t\phi(\mu) = E_\mu\phi[Y(t)], \qquad t \geq 0. \tag{8.6}$$

The operator T_t is linear and $\|T_t\phi\| \leq \|\phi\|$.

Theorem 2. *Assume property (U). Then $\{T_t\}$ is a semigroup on $C(\mathcal{M})$.*

Proof. We must show: (a) $T_t\phi \in C(\mathcal{M})$ if $\phi \in C(\mathcal{M})$, and (b) the semigroup property. To prove (a), since T_t is linear, contracting, and $\mathcal{D}_0$ spans $C(\mathcal{M})$, it suffices to show that $T_t\phi \in C(\mathcal{M})$ when $\phi \in \mathcal{D}_0$. Let $T \geq t$, $\mu_r \in \mathcal{M}$, $r = 1, 2, \cdots$ and $\mu_r \to \mu$ as $r \to \infty$. By Theorem 1 there exists a solution P^r to the martingale problem for initial state μ_r. By Lemma 5, P^r tends to a limit P' for a subsequence and P' solves the martingale problem for initial state μ. By property (U) $E^{P'}\phi[Y(t)] = T_t\phi(\mu)$. Moreover, as $r \to \infty$

$$T_t\phi(\mu_r) = E^{P_r}\phi[Y(t)] \to E^{P'}\phi[Y(t)].$$

Since $T_t\phi(\mu_r) \to T_t\phi(\mu)$ as $r \to \infty$ through any such subsequence, $T_t\phi \in C(\mathcal{M})$. This proves (a).

To prove (b), consider again the discretized process in §6 and §7. For each $r = 1, 2, \cdots$ the discretized process is a Markov diffusion on the finite-dimensional space $\mathcal{M}^r$. We denote the corresponding semigroup by $\{T_t^r\}$. Thus,

$$T_t^r\phi(\mu) = E^{P^r}\phi[Y(t)], \qquad \mu \in \mathcal{M}^r,$$

where P^r solves the martingale problem for generator $\mathcal{G}^r$ and initial state μ. Fix $T > 0$ and let $\phi \in \mathcal{D}$. Let

$$\delta_r = \max\{|T_t^r\phi(\mu) - T_t\phi(\mu)|: \mu \in \mathcal{M}^r, 0 \leq t \leq T\}.$$

Let us show that

$$\lim_{r\to\infty} \delta_r = 0. \tag{*}$$

Suppose not. Then there exist sequences μ^r, t^r and $\delta_0 > 0$ with

$$|T_{t^r}^r\phi(\mu^r) - T_{t^r}\phi(\mu^r)| \geq \delta_0.$$

Since $\mathcal{M}$ and $[0, T]$ are compact, we may assume that $\mu^r \to \mu$, $t^r \to t$ as $r \to \infty$. Since M_ϕ^r is a $(\{\mathcal{F}_t\}, P^r)$-martingale,

$$T_{t^r}^r\phi(\mu^r) - T_t^r\phi(\mu^r) = E_{\mu^r}\int_t^{t^r} \mathcal{G}^r\phi[Y(v)]dv.$$

Since $\mathcal{G}^r\phi$ is bounded, the right side tends to 0 as $r \to \infty$. Similarly, $T_{t^r}\phi(\mu^r) - T_t\phi(\mu^r)$ tends to 0. By the proof of Theorem 1 and property (U), $E_{\mu^r}\phi[Y(t)] \to E_\mu\phi[Y(t)]$ as $r \to \infty$. Hence, $T_t^r\phi(\mu^r) \to T_t\phi(\mu)$. By (a), $T_t\phi(\mu^r) \to T_t\phi(\mu)$ as $r \to \infty$. This is a contradiction. Thus (*) holds. Given $\mu \in \mathcal{M}$, let $\mu^r \in \mathcal{M}^r$ tend to μ. For $\phi \in C(\mathcal{M})$ and $s, t \geq 0$,

$$(**) \qquad T_{s+t}\phi(\mu) = \lim_{r\to\infty} T_{s+t}^r\phi(\mu^r) = \lim_{r\to\infty} T_s^r(T_t^r\phi)(\mu^r).$$

Let $\Phi = T_t\phi$, $\Phi^r = T_t^r\phi$. By (a), $\Phi \in C(\mathcal{M})$. By (*), $\|\Phi^r - \Phi\| \to 0$ where $\|\ \ \|$ is the sup norm with Φ restricted to $\mathcal{M}^r$. Then

$$\lim_{r\to\infty} |T_s^r\Phi^r(\mu^r) - T_s^r\Phi(\mu^r)| \leq \lim_{r\to\infty} \|\Phi^r - \Phi\| = 0,$$

while by Lemma 5 and property (U)

$$\lim_{r\to\infty} T_s\Phi(\mu^r) = T_s\Phi(\mu).$$

Hence, the right side of (**) is $T_s(T_t\phi)(\mu)$. Since $T_{s+t}\phi = T_s(T_t\phi)$ for any $\phi \in \mathcal{D}$ and the contracting property holds, this proves Theorem 2.

In order to show uniqueness of the solution P to the martingale problem, let us first prove the following lemma. Let $\mathcal{F}_s$ denote the σ-algebra generated by $\{y(v), 0 \leq v \leq s\}$, as in §4.

Lemma 6. *Let P be a solution to the martingale problem, $P(Y(0) = \mu^0) = 1$. Then, for all $\phi \in C(\mathcal{M})$, $0 \leq s < t \leq T$,*

$$E^P\{\phi[Y(t)]|\mathcal{F}_s\} = T_{t-s}\phi[Y(s)], \textit{ P-almost surely.}$$

Proof. It suffices to consider $\phi(\mu) = \langle\beta_1, \mu\rangle \cdots \langle\beta_K, \mu\rangle$, with $\beta_1, \cdots, \beta_K \in C(S)$. A regular conditional distribution Q_s^P for $Y(\cdot)$ given $\mathcal{F}_s$ exists [11, pp. 141-3]. $Q_s^P(\omega, \cdot)$ is a probability measure; and $Q_s^P(\cdot, B)$ is a version of $P(Y(\cdot) \in B|\mathcal{F}_s)$. We take a version of the conditional expectation such that, for any bounded $\mathcal{F}_T$-measurable function ψ on Ω

$$E^P\{\psi[Y(\cdot)]|\mathcal{F}_s\} = \int_\Omega \psi[y(\cdot)]Q_s^P(\omega, dy), \text{ P-almost surely.}$$

We subtract $M_\phi(s)$ from $M_\phi(t)$ in (5.3) and take conditional expectations. Since $M_\phi(t)$ is a P-martingale

$$(8.7) \qquad 0 = E^P\{\phi[Y(t)]|\mathcal{F}_s\} - \phi[Y(s)] - \int_s^t E^P\{\mathcal{G}\phi[Y(v)]|\mathcal{F}_s\}dv,$$

P-almost surely for each t, hence by continuity in t, P-almost surely for all $t \in [s, T]$. Let

$$u(\beta_1, \cdots, \beta_K, r, \omega) = \int_\Omega \phi[y(s + r)]Q_s^P(\omega, dy),$$

$$\phi(\mu) = \langle \beta_1, \mu \rangle \cdots \langle \beta_K, \mu \rangle, \qquad 0 \le r \le T - s.$$

By (8.1), (8.7) for P-almost all $\omega \in \Omega$, u satisfies the differential equations (8.3), or (8.2) if $K = 1$. Moreover,

$$\phi[\omega(s)] = \int_\Omega \phi[y(s)]Q_s^P(\omega, dy), \text{ } P\text{-almost surely.}$$

Hence, P-almost surely, u satisfies the initial data

$$u(\beta_1, \cdots, \beta_K, 0, \omega) = \langle \beta_1, \omega(s) \rangle \cdots \langle \beta_K, \omega(s) \rangle.$$

However,

$$\tilde{u}(\beta_1, \cdots, \beta_K, r, \omega) = T_{s+r}\phi[\omega(s)],$$

with ϕ as above, is also a solution to (8.3) with the same initial data. Both u and $\tilde{u}$ are multilinear in $\beta_1, \cdots, \beta_K$, symmetric, and continuous (for fixed ω). By property (U), $\tilde{u} = u$. This implies Lemma 6.

Theorem 3. *Assume property (U). Then the solution P to the martingale problem is unique (given the initial state μ^0).*

Proof. By property (U), $E^P\phi[Y(t)]$ is unique for each $\phi \in C(\mathcal{M})$. For $m > 1$ consider any $\phi_1, \cdots, \phi_m \in C(\mathcal{M})$ and $0 \le t_1 < t_2 < \cdots < t_m \le T$. Then

$$\begin{aligned} &E^P[\phi_1(Y(t_1))\phi_2(Y(t_2)) \cdots \phi_m(Y(t_m))] \\ &= E^P[\phi_1(Y(t_1)) \cdots \phi_{m-1}(Y(t_{m-1}))E^P\{\phi_m(Y(t_m)) | \mathcal{F}_{t_{m-1}}\}] \\ &= E^P[\phi_1(Y(t_1)) \cdots \phi_{m-1}(Y(t_{m-1}))(T_{t_m - t_{m-1}}\phi_m)(Y(t_{m-1}))]. \end{aligned}$$

At the last step we have used Lemma 6. Using induction on m and the fact that $T_{t_m - t_{m-1}}\phi_m \in C(\mathcal{M})$ by Theorem 2, $E^P[\phi_1(Y(t_1)) \cdots \phi_m(Y(t_m))]$ is unique. This implies that P is unique.

Corollary. *The collection of measures $P = P(\mu^0)$, $\mu^0 \in \mathcal{M}$, defines a Markov, Feller process with $\mathcal{M}$ as its state space. The semigroup associated with this process is $\{T_t\}$.*

Sufficient conditions for property (U). Let us next give two conditions, each of which implies property (U). The first of these is that $C(S)$ be spanned by eigenfunctions of $\mathcal{L}$.

Definition. $\mathcal{L}$ is called *diagonalizable* (over the real number field R) if there exists a basis $\alpha_1, \alpha_2, \cdots$ for $C(S)$ such that $\alpha_k \in \mathcal{S}$ and $\mathcal{L}\alpha_k = \lambda_k\alpha_k$ (λ_k real), $k = 1, 2, \cdots$.

We may assume that $\alpha_1(x) \equiv 1$, $\lambda_1 = 0$, since $\mathscr{L}1 = 0$. Recall that in this paper by basis we mean that finite linear combinations of $\alpha_1, \alpha_2, \cdots$ are dense in $C(S)$. This holds, in particular, if each β in a dense subset of $C(S)$ is the sum of a uniformly convergent series, $\beta = \sum_{k=1}^{\infty} c_k\alpha_k$.

By Sturm-Liouville theory [17], the differential operator in Example 1, §6 is diagonalizable. By separation of variables, $\mathscr{L}$ is diagonalizable in Examples 2, 3, §6.

Theorem 4. *Property (U) holds if $\mathscr{L}$ is diagonalizable.*

Proof. Let us show by induction on K that equations (8.2), (8.3) with initial data (8.4) uniquely determine $u(\beta_1, \cdots, \beta_K, t)$ for all $\beta_1, \cdots, \beta_K \in C(S)$, $t \geq 0$. For $K = 1$ and β an eigenfunction, with eigenvalue λ, (8.2) becomes

$$\frac{d}{dt} u(\beta, t) = u(\mathscr{L}\beta, t) = \lambda u(\beta, t),$$

which is a linear differential equation for $u(\beta, t)$. Thus, $u(\beta, t)$ is uniquely determined by $u(\beta, 0) = \langle \beta, \mu^0 \rangle$. Since $u(\cdot, t)$ is linear and continuous on $C(S)$, $u(\beta, t)$ is unique for all $t \geq 0$, $\beta \in C(S)$. For $K > 1$, we first consider eigenfunctions $\beta_1, \cdots, \beta_K$, $\beta_k = \alpha_{j_k}$. By (8.3)

$$\frac{d}{dt} u(\beta_1, \cdots, \beta_K, t) = Cu(\beta_1, \cdots, \beta_K, t) + w, \tag{8.8}$$

where $C = \lambda_{j_1} + \cdots + \lambda_{j_K} - K(K-1)$ and w is a sum of terms of the form

$$\tilde{u}(\beta_1, \cdots, \beta_{k-1}, \beta_k\beta_\ell, \beta_{k+1}, \cdots, \beta_{\ell-1}, \beta_{\ell+1}, \cdots, \beta_K, t)$$

with $\tilde{u}(\gamma_1, \cdots, \gamma_{K-1}, t)$ multilinear, symmetric, and continuous. By induction on K, we suppose that each $\tilde{u}$ is unique. Then (8.8) uniquely determines $u(\beta_1, \cdots, \beta_K, t)$ if $\beta_1, \cdots, \beta_K$ are eigenfunctions. Since u is multilinear and continuous, $u(\beta_1, \cdots, \beta_K, t)$ is then uniquely determined for all $\beta_1, \cdots, \beta_K \in C(S)$ and $t \geq 0$. This proves Theorem 4.

Example. Let $S = [-a, a]$, $\mathscr{L} = d^2/dx^2$ with reflecting boundary conditions, $\mathscr{S} = \{\beta \in C^2[-a, a]: \beta'(\pm a) = 0\}$. As noted below this corresponds to a continuous-type version of the Ohta-Kimura ladder model [16] except that we have imposed the bound $|x| \leq a$ on the type x. $\mathscr{L}$ is diagonalizable. Equation (8.2) is now simply the heat equation written in a weak form, with reflecting boundary conditions. The solution $u(\beta, t) = E^P(\langle\beta, Y(t)\rangle)$ corresponds for $t > 0$ to a C^∞ function $m(x, t)$ such that

$$\frac{\partial m}{\partial t} = \frac{\partial^2 m}{\partial x^2}, \qquad \frac{\partial m}{\partial x}(\pm a, t) = 0,$$

$$E^P(\langle\beta, Y(t)\rangle) = \int_{-a}^{a} \beta(x)m(x, t)dx.$$

We call $m(x, t)$ the *mean density* of type x at time t. For $K = 2$, equation (8.3) becomes

$$\frac{d}{dt} u(\beta_1, \beta_2, t) = u(\mathscr{L}\beta_1, \beta_2, t) + u(\beta_1, \mathscr{L}\beta_2, t) - 2u(\beta_1, \beta_2, t) + 2u(\beta_1\beta_2, t),$$

which is satisfied by

$$u(\beta_1, \beta_2, t) = E^P(\langle\beta_1, Y(t)\rangle\langle\beta_2, Y(t)\rangle).$$

There exists a kernel $m_2(x_1, x_2, t)$ for $x_1 \neq x_2$, $t > 0$ such that

$$\frac{\partial m_2}{\partial t} = \frac{\partial^2 m_2}{\partial x_1^2} + \frac{\partial^2 m_2}{\partial x_2^2} - 2m_2 + 2m(x_1, t)\delta(x_1 - x_2),$$

$$E^P(\langle\beta_1, Y(t)\rangle\langle\beta_2, Y(t)\rangle) = \int_{-a}^{a} \int_{-a}^{a} \beta_1(x_1)\beta_2(x_2)m_2(x_1, x_2, t)dx_1dx_2,$$

with normal derivative $\partial m_2/\partial \nu = 0$ on the boundary of the square $S \times S$. Here δ is the Dirac function. We call $m_2(x_1, x_2, t)$ the *mean joint density* of (x_1, x_2). Similar equations are obtained for the mean joint density $m_K(x_1, \cdots, x_K, t)$ when $K > 2$, using (8.3).

In the ladder model of Ohta-Kimura [16] the type space consists of the integers, and mutation produces transitions to nearest neighbor types with equal frequency left and right. After a rescaling of variables one gets a continuous analogue of the ladder model, with $S = (-\infty, \infty)$, $\mathscr{L} = d^2/dx^2$. There does not appear to be any great difficulty in extending our results for compact S to include this case, although we shall not do so here. Instead, let us outline a procedure for calculating some quantities of interest. In the ladder model, the distribution of electrophoretic types tends to move erratically with time along the line S. For this reason Moran [15] used the term "wandering distribution". Instead of the type distribution itself, Ohta-Kimura and Moran considered the distribution of the differences $\xi = x_2 - x_1$ of the types x_1, x_2 of the two randomly selected individuals in the population. In the continuous-type version of the ladder model let us denote by $Z(t)$ the measure-valued process representing the distribution of differences in types. Then

$$\langle\beta, Z(t)\rangle = \langle\beta * Y(t), Y(t)\rangle$$

where * denotes convolution, and β is any continuous function on $(-\infty, \infty)$ with compact support. Let

$$I(\xi, t) = \int_{-\infty}^{\infty} m_2(x, x + \xi, t)dx.$$

We call $I(\xi, t)$ the mean density of types differing by ξ. Then

$$\frac{\partial I}{\partial t} = \frac{\partial^2 I}{\partial \xi^2} - 2I + 2\delta(\xi).$$

As $t \to \infty$, $I(\xi, t)$ tends to an equilibrium value

$$\lim_{t \to \infty} I(\xi, t) = 2^{-\frac{1}{2}} \exp\left(-2^{\frac{1}{2}}|\xi|\right).$$

See [16, p. 203] for the corresponding formula in the discrete state ladder model.

We also introduce

$$I_2(\xi_1, \xi_2, t) = \int m_3(x, x + \xi_1, x + \xi_2, t)dx$$

$$J(\xi_1, \xi_2, t) = \iint m_4(x_1, x_1 + \xi_1, x_2, x_2 + \xi_2, t)dx_1dx_2,$$

and call $J(\xi_1, \xi_2, t)$ the mean joint density of types differing by ξ_1, ξ_2. From (8.3) with $K = 2, 3, 4$ one can get a system of three partial differential equations for I, I_2, J. These can be studied with the aid of Fourier transforms in ξ, ξ_1, ξ_2, furnishing an alternative approach to the technique of Moran [15, Part II] for studying numerically mean joint densities.

A more general condition than diagonalizability is that $\mathscr{L}$ be triangulizable in the following sense. Let $C_c(S)$ denote the space of complex-valued, continuous functions on S, and $\mathscr{S}_c$ the dense subspace of functions with real and imaginary parts in $\mathscr{S}$. Triangulizable means that $\beta_1, \beta_2, \cdots$ in $\mathscr{S}_c$ exist, $\beta_1(x) \equiv 1$, such that

$$\mathscr{L}\beta_k = \sum_{\ell=1}^{k} c_{k\ell}\beta_\ell$$

and finite linear combinations of $\beta_1, \beta_2, \cdots$ are dense in $C_c(S)$.

By essentially the same method as for Theorem 4 one can show:

Theorem 5. *Property* (U) *holds if* $\mathscr{L}$ *is triangulizable.*

In particular, let $S \subset R^D$ be the closure of a bounded region with $C^{(2)}$ boundary ∂S and $\mathscr{L}$ is an elliptic second-order partial differential operator, with zero conormal derivative boundary conditions. Then $\mathscr{L}$ is triangulizable. This follows from results of Agmon [1], [2].

Pure random genetic drift. The operator $\mathscr{L} = 0$ is trivially diagonalizable (with all $\lambda_k = 0$). When $\mathscr{L} = 0$ and $H = 0$ we say that the process $Y(t)$ is *pure random genetic drift*. In this case, the semigroup $\{T_t\}$ can be constructed more directly. The generator $\mathscr{G}$ is the same as the discretized generator $\mathscr{G}^r$ (compare (8.1) with (6.2)). We can set $T_t\phi(\mu) = T_t^r\phi(\mu)$ for $\mu \in \mathscr{M}^r$, and then

$$T_t\phi(\mu) = \lim_{r \to \infty} T_t(\phi \circ G^r)(\mu), \qquad \mu \in \mathscr{M}.$$

Yet another way to construct the semigroup for pure random genetic drift is to begin with partitions of the type space S. Such a partition is a collection $\{A_1, \cdots, A_J\}$ of disjoint sets with union S. Consider first functions of the form

$$\phi(\mu) = f[\mu(A_1), \cdots, \mu(A_J)] \quad \text{with} \quad f \in C(\Sigma^J).$$

For each such ϕ set

$$T_t\phi(\mu) = \mathcal{T}_t^J f(p), \qquad p_i = \mu(A_i),\ i = 1, \cdots, J,$$

where $\mathcal{T}_t^J$ is generated by $\mathfrak{A}^J$(§2). Then $T_t\phi(\mu)$ is consistent under refinements of the partition $\{A_1, \cdots, A_J\}$, from which it follows that $\{T_t\}$ is a contracting semigroup of linear operators on the space $\mathscr{C}$ of such ϕ. Then T_t is extended to the uniform closure $\bar{\mathscr{C}}$, and $C(\mathscr{M}) \subset \bar{\mathscr{C}}$. To see that $C(\mathscr{M}) \subset \bar{\mathscr{C}}$, in the proof of Lemma 2 (§3) let $\alpha_1, \cdots, \alpha_J$ be the indicator functions for $A_1, \cdots, A_J$, where diam [spt A_i] $\leq \eta$ and η is small.

Let $e_1 = (1, 0, \cdots, 0), \cdots, e_J = (0, 0, \cdots, 1)$ denote the vertices of the simplex Σ^J in §2. If there is a (random) time τ_i such that the frequency vector satisfies $p(t) = e_i$ for $t \geq \tau_i$, then *fixation* of type x_i is said to occur. (As in §2, the possible types in the finite-dimensional case are denoted by $x_1, \cdots, x_J$.) For the finite-dimensional process of pure random genetic drift, fixation occurs with probability 1. Moreover, if τ denotes the fixation time, then $P(\tau = \tau_i) = p_i^0$, where $p^0 = p(0)$ is the initial frequency vector. See [4].

When the type space S is not a finite set, fixation of a single type may not be expected. However, for pure random genetic drift one can speak of a random "limiting" type in the following sense. Given a partition $\{A_1, \cdots, A_J\}$ of S, the finite-dimensional result implies that there is a random time τ such that

$$\sum_{i=1}^{J} P[Y(t)(A_i) = 1 \text{ for all } t \geq \tau] = 1.$$

We may call τ the time when the type x is fixed within one of the sets A_i. Now consider a sequence of refinements of a given partition, and let A_i^n denote the sets of the n^{th} refinement. Suppose that $\max_i \operatorname{diam} A_i^n \to 0$ as $n \to \infty$, and let $x_i^n \in A_i^n$. Let τ^n be the time of fixation within some set A_i^n of the n^{th} partition, $\tau^1 \leq \tau^2 \leq \cdots$. Define $X(t)$ by

$$X(t) = x_i^n, \qquad \tau^n \leq t < \tau^{n+1}$$

if fixation at time τ^n is within A_i^n. As $t \to \infty$, $X(t)$ tends P-almost surely to a limit X_∞, which we call the limiting type.

9. A model with natural selection. In §8 we took $H = 0$; *i.e.* all types $x \in S$ were considered equally fit. We now incorporate natural selection into the model, by appropriate choice of $H(\mu, \beta)$ according to (9.2) below.

For a finite number of types $x_1, \cdots, x_J$, a standard model for incorporating natural selection is to take in (2.6)

$$h_j(p) = p_j\left(\sum_{i=1}^{J} p_i m_{ij} - \sum_{i,k=1}^{J} p_i p_k m_{ik}\right). \tag{9.1}$$

The interpretation is that $x_1, \cdots, x_J$ are possible alleles carried by a gamete at some gene locus, and m_{ij} is a fitness coefficient of the genotype (x_i, x_j) such that $m_{ji} = m_{ij}$. If p_i is the frequency of type i, then $p_i p_j$ represents the frequency of (x_i, x_j). See [4].

Let m be a symmetric function in $C(S \times S)$. We interpret $m(x, x')$ as a fitness coefficient associated with (x, x'). Given a finite set $\tilde{S} = \{x_1, \cdots, x_J\}$ contained in S, we let $m_{ij} = m(x_i, x_j)$. From (5.1) and (9.1) we find that the correct choice for H is

$$H(\mu, \beta) = \langle m\beta, \mu \otimes \mu \rangle - \langle \beta, \mu \rangle \langle m, \mu \otimes \mu \rangle, \tag{9.2}$$

where $\mu \otimes \nu$ denotes the product measure on $S \times S$. If $m_{ij} = \hat{m}_i + \hat{m}_j$, then (9.1) simplifies to

$$h_j(p) = p_j\left(\hat{m}_j - \sum_{i=1}^{J} p_i \hat{m}_i\right). \tag{9.3}$$

Correspondingly, if $m(x, x') = \hat{m}(x) + \hat{m}(x')$, then (9.2) simplifies to

$$H(\mu, \beta) = \langle \hat{m}\beta, \mu \rangle - \langle \hat{m}, \mu \rangle \langle \beta, \mu \rangle. \tag{9.4}$$

When H has the form (9.2), Theorem 1 implies the existence of a solution to the martingale problem. The uniqueness proof in §8 can no longer be used. However, uniqueness of the solution to the martingale problem follows from a formula of Cameron-Martin-Girsanov type recently proved by Dawson [7, Theorem 5.1]. In order to apply Dawson's theorem we need a continuous mapping $\mu \to \alpha_\mu$ from $\mathcal{M}(S)$ into $C(S)$ such that

$$H(\mu, \beta) = \langle \alpha_\mu \beta, \mu \rangle - \langle \alpha_\mu, \mu \rangle \langle \beta, \mu \rangle.$$

See [7, formulas (5.9), (5.21)]. We take

$$\alpha_\mu(x) = \int_S m(x, x') d\mu(x').$$

In the special case (9.4) we may simply take $\alpha_\mu = m$. As in §8, the family $P = P(\mu^0)$ of solutions to the martingale problems defines a Markov, Feller process on $\mathcal{M}$, with associated semigroup $\{T_t\}$. Unfortunately, when natural selection acts one no longer has a closed system of differential equations for moments, or for mean joint densities, as was the case in §8 when selective neutrality holds. Without neutrality it is difficult to get information about the process $Y(t)$ which is useful for population genetics applications.

Appendix. In this Appendix we review a model of Moran type, from which a diffusion process of the kind in §2 is obtained as a limit after a suitable rescaling. For simplicity, we assume selective neutrality and hence $g_j = L_j$ in (2.6). We begin with the following more general model, and then specialize the transition coefficients b_{ij}. Consider a population with J possible types $x_1, \cdots, x_J$, and let τ denote time measured on some scale. Let $n_j(\tau)$ be the number of type x_j individuals at time τ. The total number N is assumed fixed:

$$N = n_1(\tau) + \cdots + n_J(\tau).$$

In the model, the vector

$$n(\tau) = (n_1(\tau), \cdots, n_J(\tau))$$

is a continuous time Markov chain, with the transition rule

$$P(n_i(\tau) \to n_i(\tau) - 1, n_j(\tau) \to n_j(\tau) + 1) = Nb_{ij}\Delta\tau + o(\Delta\tau),$$

and all other transitions have probability $o(\Delta\tau)$. If we let $p_j = N^{-1}n_j$, then $p = (p_1, \cdots, p_J)$ belongs to the simplex Σ^J. We suppose that $b_{ij} = b_{ij}(p)$. The generator of the Markov chain $p(\tau)$ is denoted by $\mathcal{B}$:

$$\mathcal{B}f(p) = N \sum_{i \neq j} b_{ij}(p)[f(p^{ij}) - f(p)],$$

where p^{ij} is the vector with components

$$p_k^{ij} = \begin{cases} p_k - 1/N & \text{if} \quad k = i \\ p_k + 1/N & \text{if} \quad k = j \\ p_k & \text{if} \quad k \neq i, j. \end{cases}$$

For the Moran type model, we take [14], [12]

$$b_{ij}(p) = a(p_i p_j + N^{-1} p_i \theta_{ij}), \qquad a > 0$$

$$\theta_{ij} \geq 0 \quad \text{for} \quad i \neq j, \; \theta_{ii} = -\sum_{j \neq i} \theta_{ij}.$$

The term ap_ip_j represents random mating effects, and $aN^{-1}\theta_{ij}$ a mutation rate per individual from type x_i to type x_j $(i \neq j)$. We introduce the new time scale $t = aN^{-1}\tau$, and let $\mathfrak{A}_N^J = Na^{-1}\mathcal{B}$ denote the generator on this new time scale. As in §2, let

$$\mathfrak{A}^J f(p) = \sum_{j=1}^{J} L_j(p) f_{p_j} + \sum_{i,j=1}^{J} (p_i\delta_{ij} - p_ip_j) f_{p_ip_j},$$

(A.1)

$$L_j(p) = \sum_{i=1}^{J} p_i\theta_{ij}.$$

Then

$$\mathfrak{A}^J f(p) = \lim_{N \to \infty} \mathfrak{A}_N^J f(p)$$

uniformly on Σ^J, for any $f \in C^2(\Sigma^J)$.

For each $N = 1, 2, \cdots$ one can pose a martingale problem corresponding to the generator $\mathfrak{A}_N^J$. A convenient canonical sample space is $\Omega = D([0, T]; \Sigma^J)$, consisting of piecewise continuous functions from $[0, T]$ into Σ^J. By standard methods [3] a Prokhorov compactness condition can be verified. The existence of a solution to the martingale problem in §2 is obtained as $N \to \infty$ through a sequence. (Alternatively, existence follows using stochastic differential equations as in [8].)

From the form (A.1) of the generator, we see that $\mathfrak{A}^J f(p)$ is a polynomial of degree less than or equal to d if f is a polynomial of degree less than or equal to d. From the equation

$$\frac{d}{dt} E^{p^J} f[p(t)] = E^{p^J} \mathfrak{A}^J f[p(t)]$$

one gets a closed system of linear differential equations for the moments of orders less than or equal to d. The same equations also hold for conditional moments. One can formulate the finite-dimensional analogue of property (U) in §8 and then prove uniqueness of the solution to the martingale problem by the same method as for Theorem 3.

References

1. S. Agmon, *On the eigenfunctions and on the eigenvalues of general elliptic boundary problems*, Comm. Pure Appl. Math. **15** (1962), 119–147.
2. S. Agmon, *Lectures on Elliptic Boundary Problems*, Van Nostrand, Princeton, N.J., 1965.
3. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
4. J. F. Crow & M. Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.
5. D. A. Dawson, *Stochastic evolution equations*, Math. Biosci. **15** (1972), 287–316.
6. D. A. Dawson, *Stochastic evolution equations and related measure processes*, J. Multivariate Anal. **5** (1975), 1–52.
7. D. A. Dawson, *Geostochastic calculus*, (preprint).
8. S. N. Ethier, *A class of degenerate diffusion processes occurring in population genetics*, Comm. Pure Appl. Math. **29** (1976), 483–493.
9. W. H. Fleming, *Diffusion processes in population biology*, Advances in Appl. Probability **7** (1975), 100–105.
10. W. H. Fleming, *Equilibrium distributions of continuous polygenic traits*, SIAM J. Appl. Math. **36** (1979), 148–168.
11. I. I. Gikhman & A. V. Skorokhod, *Introduction to the Theory of Random Processes*, Saunders, 1969.
12. S. Karlin & J. McGregor, *On some stochastic models in genetics*, in J. Gurland, ed., *Stochastic Models in Medicine and Biology*, University of Wisconsin Press, Madison, 1964, pp. 249–279.
13. P. A. Meyer, *Seminaire de Probabilités I*, Springer Lecture Notes in Math. No. 39, 1967.
14. P. A. P. Moran, *Random processes in genetics*, Proc. Camb. Phil. Soc. **54** (1958), 60–71.
15. P. A. P. Moran, *Wandering distributions and the electrophoretic profile I, II*, Theor. Popn. Biol. **8** (1975), 318–330; **10** (1976), 145–149.
16. T. Ohta & M. Kimura, *A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population*, Genet. Res. Camb. **22** (1973), 201–204.
17. I. G. Petrovsky, *Lectures on Partial Differential Equations*, Interscience, New York, 1957.
18. P. Priouret, *Ecole d'Été de Probabilités de Saint-Flour III*, Springer Lecture Notes in Math. No. 390, 1974.
19. K-I. Sato, *Diffusion processes and a class of Markov chains related to population genetics*, Osaka J. Math. **13** (1976), 631–659.
20. K-I. Sato, *Convergence to a diffusion of a multi-allelic model in population genetics*, Advances in Appl. Probability **10** (1978).

21. M. VIOT, *Méthodes de compacité et de monotonie compacité pour les equations aux dérivees partielles stochastiques*, Thèse, Univ. de Paris, 1975.

22. M. VIOT, *A stochastic partial differential equation arising in population genetics theory*, Brown University LCDS Technical Report 75-3, 1975.

Fleming's research was partially supported by the National Science Foundation under MCS 76-07247-A03 and by a Guggenheim Fellowship.

Viot's research was supported by C.N.R.S.

Received October 16, 1978

FLEMING: BROWN UNIVERSITY, PROVIDENCE, RI 02912

VIOT: INSTITUT DE RECHERCHE D'INFORMATIQUE ET D'AUTOMATIQUE, ROCQUENCOURT, 78150 LECHESNAY, FRANCE