The Annals of Probability
1987, Vol. 15, No. 2, 515–545

THE INFINITELY-MANY-SITES MODEL AS A MEASURE-VALUED DIFFUSION[1]

BY S. N. ETHIER AND R. C. GRIFFITHS

University of Utah and Monash University

The infinitely-many-sites model (with no recombination) is reformulated, with sites labelled by elements of $[0,1]$ and "type" space $E = [0,1]^{\mathbf{Z}_+}$. A gene is of type $\mathbf{x} = (x_0, x_1, \dots) \in E$ if $x_0, x_1, \dots$ is the sequence of sites at which mutations have occurred in the line of descent of that gene. The model is approximated by a diffusion process taking values in $\mathscr{P}_a^0(E)$, the set of purely atomic Borel probability measures μ on E with the property that the locations of every $n \ge 1$ atoms of μ form a family tree, and the diffusion is shown to have a unique stationary distribution $\tilde{\mu}$. The principal object of investigation is the $\tilde{\mu}(d\mu)$-expectation of the probability that a random sample from a population with types distributed according to μ has a given tree structure. Ewens' (1972) sampling formula and Watterson's (1975) segregating-sites distribution are obtained as corollaries.

1. Introduction. The infinitely-many-sites model (or infinite sites model) in molecular population genetics was introduced by Kimura (1969, 1971) and has been studied by Ewens (1974), Watterson (1975), Li (1977), Griffiths (1981, 1982), Strobeck (1983) and others. Our purpose here is to provide a rigorous mathematical foundation for a diffusion approximation to the model (with no recombination) and to investigate its properties. We do this by first reformulating the model in a more tractable way. To indicate our basic approach and main results, we begin by describing the infinitely-many-sites model as it might typically be formulated. It depends on $M \in \mathbb{N} \equiv \{1, 2, \dots\}$ and $u \in [0,1]$.

Consider a population consisting of M genes per generation. Each gene has an infinite number of nucleotide sites, and at each site there are two or more possible nucleotides, the original one and mutant ones. The "type" of a gene is given by a sequence of 1's and 0's indicating at which sites a mutation has occurred in the gene or one of its ancestors. The reproductive mechanism assumes in effect that each of the M genes in generation $\tau + 1$ selects a parent gene at random (with replacement) from generation τ. (A truly diploid model could be formulated, but it would lead to the same results.) The offspring gene is of the same type as the parent gene with probability $1 - u$; mutation occurs with probability u, in which case a 0 changes to a 1 at a site where no mutation has occurred in any gene. Which site is chosen for the mutation does not matter, as sites are unordered.

Perhaps the best-known result for this model is due to Watterson (1975). Let us say that a site is *segregating* with respect to a sample of genes if at least one

---

Received April 1985; revised April 1986.

[1] Research supported in part by the National Science Foundation.

AMS 1980 *subject classifications.* Primary 60G57, 60J70; secondary 92A10.

Key words and phrases. Measure-valued diffusion, population genetics, infinitely-many-sites model, infinitely-many-alleles model, segregating sites, family trees, sampling distributions.

1 and at least one 0 occur at that site in members of the sample. Fix $\theta > 0$. Watterson showed, assuming stationarity, that if $u = (\theta/2M) \wedge 1$ and $M \to \infty$, the number of segregating sites in a sample of size $n \geq 2$ converges in distribution to the convolution

$$\mathop{*}_{j=1}^{n-1} \text{geometric}\left(\frac{j}{j+\theta}\right), \tag{1.1}$$

where geometric (p) is the probability measure on $\mathbb{Z}_+ \equiv \{0, 1, \ldots\}$ with mass $(1-p)^k p$ at k.

We propose the following alternative version of the infinitely-many-sites model. Let us arbitrarily label the sites by elements of $[0, 1]$. The "type" of a gene is given by the element $\mathbf{x} = (x_0, x_1, \ldots)$ of $E \equiv [0, 1]^{\mathbb{Z}_+}$ such that $x_0, x_1, \ldots$ is the sequence of sites at which mutations have occurred in the line of descent of the gene in question. (In particular, x_0 is the site at which the most recent mutation in that line of descent has occurred.) Again, each of the M genes in generation $\tau + 1$ selects a parent gene at random (with replacement) from generation τ. If the parent gene is of type $\mathbf{x} \in E$, then the offspring gene is of type $\mathbf{x}$ with probability $1 - u$ and of type $(Y, \mathbf{x})$ with probability u, where the conditional distribution of Y is uniform on $[0, 1]$. At a time when the M genes are of types $\mathbf{x}_1, \ldots, \mathbf{x}_M$, the state of the model is given by the probability measure

$$\nu = M^{-1} \sum_{i=1}^{M} \delta_{\mathbf{x}_i},$$

where $\delta_{\mathbf{x}}$ denotes the unit mass at $\mathbf{x}$. Thus, our model is a Markov chain $\{\nu_\tau^{(M)}, \tau = 0, 1, \ldots\}$ in a subset of $\mathscr{P}(E)$, the set of Borel probability measures on E. In this context, we say that a site $z \in [0, 1]$ is *segregating* with respect to a sample of n genes of types $\mathbf{x}_1, \ldots, \mathbf{x}_n$ if z appears in at least one but not all of the sequences $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Before proceeding, we attempt to show that our model is essentially equivalent to the standard one. Consider a sample of three genes from the standard model consisting of types

$$\begin{aligned} &(\ldots, 1, 1, 0, 0, 0, 0, 0, \ldots),\\ &(\ldots, 1, 0, 1, 1, 1, 0, 0, \ldots),\\ &(\ldots, 1, 0, 1, 0, 0, 1, 0, \ldots) \end{aligned} \tag{1.2}$$

and a sample of three genes from our model consisting of types

$$\begin{aligned} &(y_1, x_0, x_1, x_2, \ldots),\\ &(y_4, y_3, y_2, x_0, x_1, x_2, \ldots),\\ &(y_5, y_2, x_0, x_1, x_2, \ldots), \end{aligned} \tag{1.3}$$

where $y_1, y_2, \ldots, y_5, x_0, x_1, \ldots$ are distinct. In both cases we can infer that the three genes have a common ancestor, five segregating sites and a (condensed) family tree of the form shown in Figure 1. Thus, ignoring ordering of columns in

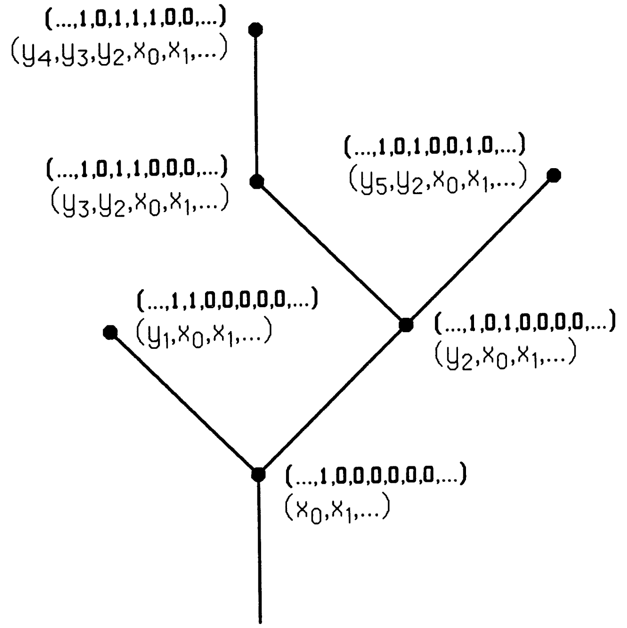

FIG. 1.

(1.2) and labelling of sites in (1.3), we see that (1.2) and (1.3) contain the same information in different forms. We emphasize that a branch in the tree of Figure 1 corresponds to the number of generations in that line of descent needed to produce a mutation, not simply to a single generation.

Now, letting $u = (\theta/2M) \wedge 1$, a result of Kurtz (1981) implies that

$$\left\{\nu^{(M)}_{[Mt]},\, t \geq 0\right\} \Rightarrow \{\mu_t,\, t \geq 0\} \text{ in } D_{\mathscr{P}(E)}[0,\infty) \text{ as } M \to \infty, \tag{1.4}$$

where $\{\mu_t, t \geq 0\}$ is a Markov process (in fact, a diffusion) in $\mathscr{P}(E)$ corresponding to the generator G defined on functions $\phi \in C(\mathscr{P}(E))$ of the form

$$\phi(\mu) = F(\langle f_1, \mu\rangle, \dots, \langle f_k, \mu\rangle) \tag{1.5}$$

by

$$\begin{aligned} G\phi(\mu) &= \tfrac{1}{2} \sum_{i,j=1}^{k} (\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle) F_{,ij}(\langle f_1, \mu\rangle, \dots, \langle f_k, \mu\rangle) \\ &\quad + \tfrac{1}{2}\theta \sum_{i=1}^{k} (\langle f_i, \lambda \times \mu\rangle - \langle f_i, \mu\rangle) F_{,i}(\langle f_1, \mu\rangle, \dots, \langle f_k, \mu\rangle), \end{aligned} \tag{1.6}$$

where $k \in \mathbb{N}$, F is a polynomial on $\mathbb{R}^k$, $f_1, \dots, f_k \in C(E)$, $\langle f, \mu\rangle$ denotes $\int f\, d\mu$, $F_{,i}$ and $F_{,ij}$ denote first- and second-order partial derivatives, and λ denotes Lebesgue measure on $[0,1]$. Of course, $\mathscr{P}(E)$ has the topology of weak convergence. This measure-valued diffusion limit of our reformulation of the infinitely-many-sites model is the subject of this paper.

It may be helpful to attempt to relate (1.5) and (1.6) to more-familiar material. We point out later that $f_1, \dots, f_k$ may be chosen bounded and Borel measurable (not necessarily continuous), so let $f_i = \chi_{\Lambda_i}$ (the indicator function of Λ_i), where $\Lambda_1, \dots, \Lambda_k$ is a partition of E into Borel measurable subsets, and put $p_i = \langle f_i, \mu \rangle = \mu(\Lambda_i)$. Then (1.5) and (1.6) become

$$\phi(\mu) = F(p_1, \dots, p_k)$$

and

$$G\phi(\mu) = \tfrac{1}{2} \sum_{i,j=1}^{k} p_i(\delta_{ij} - p_j) F_{,ij}(p_1, \dots, p_k)$$
$$+ \tfrac{1}{2}\theta \sum_{i=1}^{k} (q_i - p_i) F_{,i}(p_1, \dots, p_k),$$

where $q_i = \langle \chi_{\Lambda_i}, \lambda \times \mu \rangle = (\lambda \times \mu)(\Lambda_i)$ (which is not necessarily a function of $p_1, \dots, p_k$), and this is reminiscent of the generator (or backward operator) for the usual k-allele diffusion model. It may also help clarify the second sum in (1.6) to write it as

$$\sum_{i=1}^{k} \langle Bf_i, \mu \rangle F_{,i}(\langle f_1, \mu \rangle, \dots, \langle f_k, \mu \rangle),$$

where

$$Bf(\mathbf{x}) = \tfrac{1}{2}\theta \int_0^1 (f(y, \mathbf{x}) - f(\mathbf{x})) \lambda(dy).$$

The operator B is the generator of a pure jump Markov process in E that jumps from $\mathbf{x} \in E$ to $(Y, \mathbf{x}) \in E$ with rate $\theta/2$, where Y is uniformly distributed on $[0, 1]$. Thus, B describes the mutation process.

In Section 2 we give a more precise statement of the above result, i.e., (1.4), and show that $\{\mu_t, t \geq 0\}$ has a unique stationary distribution $\tilde{\mu}$ and is ergodic. Moreover, letting $\mathscr{P}_a(E) = \{\mu \in \mathscr{P}(E)\colon \mu \text{ is purely atomic}\}$, we show that $\tilde{\mu}(\mathscr{P}_a(E)) = 1$ and $\mathbb{P}\{\mu_t \in \mathscr{P}_a(E) \text{ for all } t > 0\} = 1$.

Section 3 indicates the relationships between $\{\mu_t, t \geq 0\}$ and the infinitely-many-alleles diffusion model, and between $\{\mu_t, t \geq 0\}$ and the finitely-many-sites diffusion model. In particular, it is shown that the $\mathscr{P}[0,1]$-valued process of zeroth marginals, $\{\mu_t(\cdot \times E), t \geq 0\}$, is the measure-valued infinitely-many-alleles diffusion model, and that $\{\mu_t, t \geq 0\}$ is the limit in distribution of a measure-valued m-site diffusion model as $m \to \infty$.

In Section 4 we examine the family-tree structure of a random sample from a population with types distributed according to the random measure μ having distribution $\tilde{\mu}$ and according to μ_t for each $t > 0$. If $n \in \mathbb{N}$ and $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in E^n$, let us say that $(\mathbf{x}_1,\dots,\mathbf{x}_n)$ is a *tree* if

(1.7) the coordinates x_{ij}, $j = 0,1,\dots,$ of $\mathbf{x}_i$ are distinct for fixed $i \in \{1,\dots,n\}$,

(1.8) if $i, i' \in \{1,\dots,n\}$, $j, j' \in \mathbb{Z}_+$ and $x_{ij} = x_{i'j'}$, then $x_{i,j+l} = x_{i',j'+l}$ for $l = 0,1,\dots,$

(1.9) there exist $j_1,\dots,j_n \in \mathbb{Z}_+$ such that $x_{1j_1} = \cdots = x_{nj_n}$.

Informally, (1.7) means that mutations never occur more than once at the same site; (1.8) means that if two genes have ancestors with the most recent mutations in their respective lines of descent at the same site, then the ancestors are of the same type; and (1.9) means that every n genes have a common ancestral type, hence a common ancestor.

Defining $\mathscr{P}_a^0(E) = \{\mu \in \mathscr{P}_a(E)\colon \mu^n(\mathscr{T}_n) = 1 \text{ for every } n \in \mathbb{N}\}$, where $\mathscr{T}_n = \{(\mathbf{x}_1,\dots,\mathbf{x}_n) \in E^n\colon (\mathbf{x}_1,\dots,\mathbf{x}_n) \text{ is a tree}\}$ and μ^n denotes the n-fold product measure $\mu \times \mu \times \cdots \times \mu$ and is the joint distribution of an ordered random sample of n types from a population with types distributed according to μ, we show that

$$\tilde{\mu}\big(\mathscr{P}_a^0(E)\big) = 1 \tag{1.10}$$

and that $\mathbb{P}\{\mu_0 \in \mathscr{P}_a^0(E)\} = 1$ implies

$$\mathbb{P}\big\{\mu_t \in \mathscr{P}_a^0(E) \text{ for all } t > 0\big\} = 1. \tag{1.11}$$

To prove (1.10) we show first that

$$\int \mu^n(\mathscr{T}_{s,n})\tilde{\mu}(d\mu) = \left[\mathop{*}_{j=1}^{n-1} \text{geometric}\left(\frac{j}{j+\theta}\right)\right](\{s\}) \tag{1.12}$$

for all $s \in \mathbb{Z}_+$ and $n \in \mathbb{N}$, where $\mathscr{T}_{s,n} = \{(\mathbf{x}_1,\dots,\mathbf{x}_n) \in E^n\colon (\mathbf{x}_1,\dots,\mathbf{x}_n)$ is a tree with s segregating sites$\}$ and the right-hand side is the mass at s of the probability measure (1.1), interpreted as δ_{s0} (the Kronecker delta) when $n = 1$, i.e., Watterson's result holds exactly.

Moreover, a recursive system of linear equations is obtained for the probabilities

$$\int \mu^n(T)\tilde{\mu}(d\mu), \tag{1.13}$$

where $n \in \mathbb{N}$ and $T \subset E^n$ is an equivalence class of trees belonging to a certain quotient set. Note that (1.13) is the $\tilde{\mu}(d\mu)$-expectation of the probability that an ordered random sample of size n from a population with types distributed according to μ forms a tree of class T. Here $(\mathbf{x}_1,\dots,\mathbf{x}_n)$ is equivalent to $(\mathbf{y}_1,\dots,\mathbf{y}_n)$ if the latter equals the former after relabelling sites, i.e., if there exists a bijection $\zeta\colon [0,1] \to [0,1]$ such that $y_{ij} = \zeta(x_{ij})$ for $i = 1,\dots,n$ and $j = 0,1,\dots$. The solution of these equations remains an open problem, but the relationship to (1.12) can be understood by considering the simple case $n = 2$. For $i, j \in \mathbb{Z}_+$ let $T_{i,j}$ be the equivalence class of trees of the form $((x_0,\dots,x_{i-1}, z_0, z_1,\dots), (y_0,\dots,y_{j-1}, z_0, z_1,\dots))$ where $x_0,\dots,x_{i-1},$

$y_0, \dots, y_{j-1}, z_0, z_1, \dots$ are distinct, and put

$$p_{i,j} = \int \mu^2(T_{i,j})\tilde{\mu}(d\mu).$$

Then it can be shown that

$$2(1+\theta)p_{i,j} = \theta p_{i,j-1} + \theta p_{i-1,j} + 2\delta_{i0}\delta_{j0} \tag{1.14}$$

for all $i, j \in \mathbb{Z}_+$, where $p_{i,-1} = p_{-1,j} = 0$, and hence that

$$p_{i,j} = \binom{i+j}{i}\left(\frac{\theta}{2(1+\theta)}\right)^{i+j}\frac{1}{1+\theta}. \tag{1.15}$$

It follows that the number of segregating sites ($= i + j$ for samples from $T_{i,j}$) is geometrically distributed under $\int \mu^2(\cdot)\tilde{\mu}(d\mu)$ with parameter $1/(1+\theta)$, and therefore (1.15) generalizes (1.12) when $n = 2$.

Strobeck (1983) derived a recursive system of linear equations for the probabilities of trees somewhat similar to ours in the case that the number of distinct types in the sample is at most three. See Remark 4.7.

While the Ewens (1972) sampling formula would follow from the results of Section 3, we give instead a combinatorial derivation based on the equations for the tree probabilities (1.13). It is clear, for example, how to do this when $n = 2$ using (1.14).

In Section 5 we give three probabilistic representations of the probabilities (1.13). The first involves a dual Markov process that in effect "chops down the tree." The second is based on a Markov chain that in effect "grows the tree." The third is related to Kingman's (1982) coalescent process. In lieu of explicit formulas such as (1.15), these representations may prove useful for purposes of simulation. Also included in this section is a measure-theoretic representation in terms of an analogue of what has been referred to as the frequency spectrum.

2. Characterization of the diffusion. Our first theorem summarizes some known results involving certain measure-valued Markov processes occurring in population genetics. To state it we need to introduce some notation and definitions.

Let E be a compact metric space, $\mathscr{B}(E)$ the Borel σ-algebra of E, $B(E)$ the space of real, bounded, Borel functions on E, $\mathscr{P}(E)$ the set of Borel probability measures on E with the topology of weak convergence, and $\mathscr{P}_a(E)$ the set of purely atomic Borel probability measures on E. For $f \in B(E)$ and $\mu \in \mathscr{P}(E)$ denote $\int f\,d\mu$ by $\langle f, \mu\rangle$. $C_E[0,\infty)$ is the set of continuous functions ω: $[0,\infty) \to E$ and is given the topology of uniform convergence on bounded intervals. $D_E[0,\infty)$ is the set of right continuous functions ω: $[0,\infty) \to E$ that have left limits and is given the Skorohod topology, and $\Rightarrow$ denotes convergence in distribution.

Given a set $A \subset B(E) \times B(E)$, a solution of the $C_E[0,\infty)$ *martingale problem for* A is a process $\{X(t),\ t \geq 0\}$ with sample paths in $C_E[0,\infty)$ such that $f(X(t)) - \int_0^t g(X(v))\,dv$ is a martingale with respect to $\sigma(X(v)\colon 0 \leq v \leq t)$ for every $(f, g) \in A$. The $C_E[0,\infty)$ martingale problem for A is *well posed* if, for each $\nu \in \mathscr{P}(E)$, there exists a solution of the $C_E[0,\infty)$ martingale problem for A

with initial distribution ν and every such solution induces the same distribution on $C_E[0, \infty)$. In this case the family of solutions (one for each initial distribution) satisfies the Markov property.

If $\{f_m\} \subset B(E)$ and $f \in B(E)$, we write $\text{bp-lim}_{m \to \infty} f_m = f$ ("bounded pointwise" limit) if $\sup_m \sup_{x \in E} |f_m(x)| < \infty$ and $\lim_{m \to \infty} f_m(x) = f(x)$ for every $x \in E$. A set $A \subset B(E) \times B(E)$ is *bp-closed* if whenever $\{(f_m, g_m)\} \subset A$, $(f, g) \in B(E) \times B(E)$, $\text{bp-lim}_{m \to \infty} f_m = f$, and $\text{bp-lim}_{m \to \infty} g_m = g$, we have $(f, g) \in A$. The *bp-closure* of A [denoted by $\text{bp-cl}(A)$] is the smallest bp-closed set containing A. If A is a subspace of $B(E) \times B(E)$, then so is $\text{bp-cl}(A)$.

THEOREM 2.1. *Let E be a compact metric space, let θ be a fixed positive constant, let $Q(x, dy)$ be a Feller transition function on $E \times \mathscr{B}(E)$, and define the bounded linear operator B on $B(E)$ by*

$$Bf(x) = \tfrac{1}{2}\theta \int (f(y) - f(x)) Q(x, dy). \tag{2.1}$$

[The Feller condition on Q implies that B maps $C(E)$ into $C(E)$.] Let

$$\mathscr{D} = \{\phi \in C(\mathscr{P}(E)) \colon \phi(\mu) \equiv F(\langle f_1, \mu \rangle, \dots, \langle f_k, \mu \rangle),$$
$$k \in \mathbb{N},\ F \text{ is a polynomial on } \mathbb{R}^k,\ f_1, \dots, f_k \in C(E)\}$$

and

$$\mathscr{D}^+ = \{\phi \in B(\mathscr{P}(E)) \colon \phi(\mu) \equiv \langle f, \mu^k \rangle,\ k \in \mathbb{N},\ f \in B(E^k)\},$$

where $\mu^k \in \mathscr{P}(E^k)$ denotes the k-fold product measure $\mu \times \mu \times \cdots \times \mu$. For $\phi \in \mathscr{D}^+$ of the form $\phi(\mu) = \langle f, \mu^k \rangle$, define

$$G\phi(\mu) = \sum_{1 \le i < j \le k} \left(\langle \Psi_{ij} f, \mu^{k-1} \rangle - \langle f, \mu^k \rangle \right) + \sum_{i=1}^{k} \langle B^{(i)} f, \mu^k \rangle, \tag{2.2}$$

where $\Psi_{ij} \colon B(E^k) \to B(E^{k-1})$ and $B^{(i)} \colon B(E^k) \to B(E^k)$ are given by

$$\Psi_{ij} f(x_1, x_2, \dots, x_{k-1}) = f(x_1, \dots, x_{j-1}, x_i, x_j, \dots, x_{k-1}) \tag{2.3}$$

and

$$\begin{aligned} B^{(i)} f(x_1, \dots, x_k) &= B[f(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_k)](x_i) \\ &= \tfrac{1}{2}\theta \int (f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_k) \\ &\qquad - f(x_1, \dots, x_k)) Q(x_i, dy). \end{aligned}$$

Note that $\mathscr{D} \subset \mathscr{D}^+$. Let

$$A = \{(\phi, G\phi) \colon \phi \in \mathscr{D}\} \quad \text{and} \quad A^+ = \{(\phi, G\phi) \colon \phi \in \mathscr{D}^+\}.$$

Then the following conclusions hold.

(a) *The $C_{\mathscr{P}(E)}[0, \infty)$ martingale problem for A is well posed.*

(b) *Let $\hat{A} = \{(\phi, \psi) \in B(\mathscr{P}(E)) \times B(\mathscr{P}(E))$: for each solution $\{\mu_t, t \ge 0\}$ of the $C_{\mathscr{P}(E)}[0, \infty)$ martingale problem for A, $\phi(\mu_t) - \int_0^t \psi(\mu_v)\, dv$ is a $\sigma(\mu_v \colon 0 \le$*

$v \leq t$)-martingale whose sample paths belong to $C_{\mathbf{R}}[0,\infty)$ *with probability* 1}. *Then* $A^+ \subset \text{bp-cl}(A) \subset \hat{A}$.

(c) *If* $\{\mu_t,\ t \geq 0\}$ *is a solution of the* $C_{\mathscr{P}(E)}[0,\infty)$ *martingale problem for* A, *then* $\mathbb{P}\{\mu_t \in \mathscr{P}_a(E) \text{ for all } t > 0\} = 1$, *regardless of the initial distribution.*

(d) *For each* $M \in \mathbb{N}$, *let* $P_M(x, dy)$ *be a transition function on* $E \times \mathscr{B}(E)$; *define the bounded linear operator* B_M *on* $B(E)$ *by*

$$B_M f(x) = M\int (f(y) - f(x))P_M(x, dy); \tag{2.4}$$

define the Markov chain $\{Y^M(\tau),\ \tau = 0,1,\dots\}$ *in* $E^M = E \times \cdots \times E$ *(M factors) by requiring that, conditionally on* $Y^M(\tau) = (x_1,\dots,x_M) \in E^M$, *the components of* $Y^M(\tau+1)$ *are independent with common distribution* $M^{-1}\sum_{i=1}^M P_M(x_i,\cdot)$; *define* $\rho_M\colon E^M \to \mathscr{P}(E)$ *by* $\rho_M(x_1,\dots,x_M) = M^{-1}\sum_{i=1}^M \delta_{x_i}$; *and define the Markov chain* $\{\nu_\tau^{(M)},\ \tau = 0,1,\dots\}$ *in a subset of* $\mathscr{P}(E)$ *by* $\nu_\tau^{(M)} = \rho_M(Y^M(\tau))$. *Suppose that*

$$\lim_{M\to\infty} \|B_M f - Bf\| = 0, \qquad f \in C(E). \tag{2.5}$$

Then, if $\{\mu_t,\ t \geq 0\}$ *is a solution of the* $C_{\mathscr{P}(E)}[0,\infty)$ *martingale problem for* A, *and if* $\nu_0^{(M)} \Rightarrow \mu_0$ *in* $\mathscr{P}(E)$, *we have* $\{\nu_{[Mt]}^{(M)},\ t \geq 0\} \Rightarrow \{\mu_t,\ t \geq 0\}$ *in* $D_{\mathscr{P}(E)}[0,\infty)$.

(e) *For each* $m \in \mathbb{N}$, *let* E_m *be a closed subset of* E, *let* $Q_m(x, dy)$ *be a Feller transition function on* $E_m \times \mathscr{B}(E_m)$, *define* B_m *and* A_m *in terms of* E_m *and* Q_m *in the same way that* B *and* A *were defined in terms of* E *and* Q, *and suppose that*

$$\lim_{m\to\infty} \sup_{x\in E_m} |B_m f(x) - Bf(x)| = 0, \qquad f \in C(E). \tag{2.6}$$

If $\{\mu_t^{(m)},\ t \geq 0\}$ *is a solution of the* $C_{\mathscr{P}(E_m)}[0,\infty)$ *martingale problem for* A_m *for each* $m \in \mathbb{N}$, *if* $\{\mu_t,\ t \geq 0\}$ *is a solution of the* $C_{\mathscr{P}(E)}[0,\infty)$ *martingale problem for* A, *and if* $\mu_0^{(m)} \Rightarrow \mu_0$ *in* $\mathscr{P}(E)$, *then* $\{\mu_t^{(m)},\ t \geq 0\} \Rightarrow \{\mu_t,\ t \geq 0\}$ *in* $C_{\mathscr{P}(E)}[0,\infty)$. [*Here we regard* $\mathscr{P}(E_m)$ *as a subset of* $\mathscr{P}(E)$.]

PROOF. (a) and (d) are due to Kurtz (1981), though in the case of particular interest [see (2.7) and (2.9) below], (a) follows from Fleming and Viot (1979). For (b), (c) and (e), see Ethier and Kurtz (1986b). □

In the sequel (except where the contrary is specifically indicated),

$$E = [0,1]^{\mathbf{Z}_+} \tag{2.7}$$

and θ *is a fixed positive constant.* The description of $\{\nu_\tau^{(M)},\ \tau = 0,1,\dots\}$ in Section 1 indicates that in Theorem 2.1(d),

$$P_M(\mathbf{x}, \Lambda) = (1-u)\delta_{\mathbf{x}}(\Lambda) + u\int_0^1 \delta_{(y,\mathbf{x})}(\Lambda)\lambda(dy) \tag{2.8}$$

for all $\mathbf{x} \in E$, $\Lambda \in \mathscr{B}(E)$, and $M \in \mathbb{N}$, where $u = (\theta/2M) \wedge 1$ and λ is Lebesgue measure on [0, 1]. Consequently, (2.5) holds with

$$Q(\mathbf{x}, \Lambda) = \int_0^1 \delta_{(y,\mathbf{x})}(\Lambda)\lambda(dy) \tag{2.9}$$

for all $\mathbf{x} \in E$ and $\Lambda \in \mathscr{B}(E)$, and it follows that the conclusions of Theorem 2.1 hold in this context. *Throughout the remainder of the paper,* B, $\mathscr{D}$, $\mathscr{D}^+$, G, A, A^+ *and* $\{\mu_t,\ t \geq 0\}$ *are as in Theorem* 2.1 [*assuming* (2.7) *and* (2.9)]. Notice that $G\phi$ is given by (1.6) if $\phi \in \mathscr{D}$.

Let us turn next to the question of stationary distributions. We begin with a lemma that will prove useful here as well as in Section 5.

LEMMA 2.2. *Let* $C \subset \text{bp-cl}(A)$ *and* $D = \{\phi\colon (\phi, \psi) \in C$ *for some* $\psi\}$, *and suppose that to each* $\phi \in D$ *there corresponds a nonnegative integer*, degree(ϕ), *such that* degree(ϕ) $= 0$ *implies* $\phi \equiv$ *constant. Assume that for each* $m \in \mathbb{N}$ *and* $(\phi, \psi) \in C$ *with* degree(ϕ) $= m$, *there exist* $\gamma > 0$ *and* $\eta \in \text{span}\{\xi \in D\colon$ degree(ξ) $\leq m - 1\}$ *such that*

$$\psi = -\gamma\phi + \eta. \tag{2.10}$$

If $\tilde{\mu}, \tilde{\nu} \in \mathscr{P}(\mathscr{P}(E))$ *are stationary distributions for* $\{\mu_t,\ t \geq 0\}$, *then*

$$\lim_{t \to \infty} \mathbb{E}[\phi(\mu_t)] = \int \phi \, d\tilde{\mu} = \int \phi \, d\tilde{\nu}, \qquad \phi \in D. \tag{2.11}$$

PROOF. With ϕ, ψ, γ and η as in (2.10),

$$\mathbb{E}[\phi(\mu_t)] - \mathbb{E}[\phi(\mu_0)] = \int_0^t \{-\gamma\mathbb{E}[\phi(\mu_v)] + \mathbb{E}[\eta(\mu_v)]\}\, dv \tag{2.12}$$

for all $t \geq 0$, and

$$0 = \int \{-\gamma\phi + \eta\}\, d\tilde{\mu} = \int \{-\gamma\phi + \eta\}\, d\tilde{\nu}. \tag{2.13}$$

By (2.12) $\mathbb{E}[\phi(\mu_t)]$ is continuous in $t \geq 0$ for all $\phi \in D$, and hence (since $\eta \in \text{span}\, D$) continuously differentiable. The second equality in (2.11) follows from (2.13) by induction on m. Also, (2.12) and the first equality in (2.13) imply [cf. Ethier and Kurtz (1981)]

$$\begin{aligned} \left|\mathbb{E}[\phi(\mu_t)] - \int \phi \, d\tilde{\mu}\right| &\leq e^{-\gamma t}\left|\mathbb{E}[\phi(\mu_0)] - \int \phi\, d\tilde{\mu}\right| \\ &\quad + e^{-\gamma t}\int_0^t e^{\gamma v}\left|\mathbb{E}[\eta(\mu_v)] - \int \eta\, d\tilde{\mu}\right| dv \end{aligned} \tag{2.14}$$

for all $t \geq 0$, so the first equality in (2.11) follows by induction on m. □

THEOREM 2.3. $\{\mu_t,\ t \geq 0\}$ *has a unique stationary distribution* $\tilde{\mu} \in \mathscr{P}(\mathscr{P}(E))$ *and is ergodic. Moreover,* $\tilde{\mu}(\mathscr{P}_a(E)) = 1$.

PROOF. Because $\{\mu_t,\ t \geq 0\}$ is a Feller process in a compact state space, the existence of a stationary distribution is automatic. Given $k \in \mathbb{N}$, $m_1, \ldots, m_k \in \mathbb{N}$, and $g_{ij} \in C[0,1]$ for $i = 1, \ldots, k$ and $j = 0, \ldots, m_i - 1$ with $g_{i, m_i - 1}$ non-constant, put

$$\phi(\mu) \equiv \prod_{i=1}^{k} \langle f_i, \mu\rangle, \quad \text{where} \quad f_i(\mathbf{x}) \equiv \prod_{j=0}^{m_i - 1} g_{ij}(x_j),$$

and define degree$(\phi) = m_1 + \cdots + m_k$. Then

$$G\phi(\mu) = -\left[\binom{k}{2} + \tfrac{1}{2}k\theta\right]\phi(\mu) + \tfrac{1}{2}\sum_{i\neq j}\langle f_i f_j, \mu\rangle \prod_{l:\, l\neq i, j}\langle f_l, \mu\rangle$$
$$+ \tfrac{1}{2}\theta\sum_i \langle f_i, \lambda \times \mu\rangle \prod_{j:\, j\neq i}\langle f_j, \mu\rangle,$$

and the (i, j)th term in the first sum on the right has degree at most $m_i \vee m_j + \sum_{l:\, l\neq i, j} m_l$, while the ith term in the second sum has degree $m_i - 1 + \sum_{j:\, j\neq i} m_j$. Thus, the class

$$C = \{(\phi, G\phi) \colon \phi \text{ as in (2.15)}\} \cup \{(\phi, 0) \colon \phi \text{ constant}\}$$

satisfies the conditions of Lemma 2.2, and since the linear span of D is dense in $C(\mathscr{P}(E))$, uniqueness of stationary distributions and ergodicity follow from the lemma. The assertion that $\tilde{\mu}(\mathscr{P}_a(E)) = 1$ is an immediate consequence of Theorem 2.1(c). □

In what follows, $\tilde{\mu}$ *denotes the unique stationary distribution of* $\{\mu_t, t \geq 0\}$.

3. Relationship to other models. We observe first that if $\{\nu_\tau^{(M)}, \tau = 0, 1, \dots\}$ is the $\mathscr{P}(E)$-valued infinitely-many-sites Markov chain model of Theorem 2.1(d) [assuming (2.7) and (2.9)], then the $\mathscr{P}[0,1]$-valued process $\{\nu_\tau^{(M)}(\cdot \times E), \tau = 0, 1, \dots\}$ is precisely the measure-valued infinitely-many-alleles Markov chain model. It is therefore not surprising that the $\mathscr{P}[0,1]$-valued process $\{\mu_t(\cdot \times E), t \geq 0\}$ is precisely the measure-valued infinitely-many-alleles diffusion model discussed by Ethier and Kurtz (1986a).

Let A_0 be the operator A of Theorem 2.1 with the roles of E and $Q(x, \cdot)$ played by $[0,1]$ and λ (Lebesgue measure on $[0,1]$) for every $x \in [0,1]$. Define $\xi_0 \colon E \to [0,1]$ by $\xi_0(\mathbf{x}) = x_0$ and $\tilde{\xi}_0 \colon \mathscr{P}(E) \to \mathscr{P}[0,1]$ by $\tilde{\xi}_0(\mu) = \mu\xi_0^{-1}$.

THEOREM 3.1. $\{\mu_t\xi_0^{-1}, t \geq 0\}$ *solves the* $C_{\mathscr{P}[0,1]}[0,\infty)$ *martingale problem for* A_0 *and has* $\tilde{\mu}\tilde{\xi}_0^{-1}$ *as a stationary distribution.*

PROOF. Let G_0 and $\mathscr{D}_0$ be related to G and $\mathscr{D}$ in the same way that A_0 is related to A. Noting that $\langle f \circ \xi_0, \mu\rangle = \langle f, \mu\xi_0^{-1}\rangle$ and $\langle f \circ \xi_0, \lambda \times \mu\rangle = \langle f, \lambda\rangle$ for all $f \in C[0,1]$ and $\mu \in \mathscr{P}(E)$, we find that $\phi \circ \tilde{\xi}_0 \in \mathscr{D}$ and

$$G(\phi \circ \tilde{\xi}_0) = (G_0\phi) \circ \tilde{\xi}_0 \tag{3.1}$$

for all $\phi \in \mathscr{D}_0$. The first conclusion is an immediate consequence of this, and the second follows from (3.1) and Echeverria's theorem [Ethier and Kurtz (1986a)]. □

Another way to view the infinitely-many-sites model is as a limit of m-site models as $m \to \infty$. Below we attempt to make this relationship precise. We begin by describing the m-site model as it might typically be formulated. It depends on $M \in \mathbb{N}$, $m \in \mathbb{N}$, $r \in \mathbb{N} - \{1\}$, $u \in [0,1]$, and a one-step transition matrix $(\pi_{kl})_{k, l=1,\dots,r}$ with $\pi_{kk} = 0$ for $k = 1, \dots, r$.

Consider a population consisting of M genes per generation. Each gene has m nucleotide sites, and at each site there are r possible nucleotides (usually, $r = 4$), labelled $1, 2, \dots, r$. Thus, the "type" of a gene is given by an element of

$$L_m = \{1, \dots, r\}^m.$$

The reproductive mechanism assumes in effect that each of the M genes in generation $\tau + 1$ selects a parent gene at random (with replacement) from generation τ. The offspring gene is of the same type as the parent gene with probability $1 - u$; mutation occurs with probability u, in which case a site is chosen at random, and given that the nucleotide appearing at that site in the parent gene is k, it is replaced at that site in the offspring gene by the nucleotide l with probability π_{kl}. Sites are unordered as in the infinite case, though for the purpose of analyzing the model it is more convenient to order them, and so we adopt this convention. Letting $\mathbf{p}^{m,M}(\tau)$ denote the vector of (relative) type frequencies in generation τ, the above description indicates that $\{\mathbf{p}^{m,M}(\tau), \tau = 0, 1, \dots\}$ is a Markov chain in

$$\Delta_{m,M} = \left\{\mathbf{p} \in \Delta_m \colon M\mathbf{p} \in \mathbb{Z}_+^{L_m}\right\},$$

where

$$\Delta_m = \left\{\mathbf{p} = (p_\alpha)_{\alpha \in L_m} \in [0,1]^{L_m} \colon \sum_{\alpha \in L_m} p_\alpha = 1\right\},$$

whose transition probabilities can easily be written out explicitly.

Moreover, if $u = (\theta/2M) \wedge 1$ and the initial distributions converge as $M \to \infty$, then it is easily shown that $\mathbf{p}^{m,M}([M\,\cdot\,]) \Rightarrow \mathbf{p}^m(\cdot)$ in $D_{\Delta_m}[0,\infty)$ as $M \to \infty$, where $\mathbf{p}^m(\cdot)$ is a diffusion process in Δ_m corresponding to $\mathscr{A}_m = \{(f, \mathscr{G}_m f)\colon f \in C^2(\Delta_m)\}$,

$$\begin{aligned}\mathscr{G}_m &= \tfrac{1}{2} \sum_{\alpha, \beta \in L_m} p_\alpha\left(\delta_{\alpha\beta} - p_\beta\right) \partial_\alpha \partial_\beta \\ &\quad + \tfrac{1}{2}\theta \sum_{\alpha \in L_m} \left[-p_\alpha + (1/m) \sum_{j=1}^{m} \sum_{k=1}^{r} \pi_{k,\alpha_j} p_{\alpha(j,k)}\right] \partial_\alpha,\end{aligned} \tag{3.2}$$

and $\alpha(j, k)$ is the element of L_m that differs from α only at the jth coordinate and equals k there. The $C_{\Delta_m}[0,\infty)$ martingale problem for $\mathscr{A}_m$ is well posed; see for example Ethier and Kurtz (1981), Proposition 2.3. The diffusion $\mathbf{p}^m(\cdot)$ has been studied by Notohara (1984) in the case $\pi_{kl} \equiv (r-1)^{-1}(1 - \delta_{kl})$ and $r = 4$.

To clarify the relationship between our measure-valued infinitely-many-sites diffusion model $\{\mu_t, t \geq 0\}$ and the limit in some sense of the above m-site diffusion model as $m \to \infty$, one might attempt to reformulate the discrete m-site model $\{\mathbf{p}^{n,M}(\tau), \tau = 0, 1, \dots\}$ in a way that is analogous to our reformulation of the discrete infinitely-many-sites model described in Section 1. Unfortunately, this is not possible for the following reason. In the infinitely-many-sites model, there is no back-mutation to the original nucleotide at a given site, and thus the sequence of sites at which mutations have occurred in the line of descent of a given gene is sufficient to determine the gene's nucleotides (or at least whether they are the original or mutant ones). This is not so in the finitely-many-sites model if $r \geq 3$ because of back-mutation. Consequently, we redefine the "type" of a gene to be the element $\boldsymbol{\alpha} = (\alpha^{(0)}, \alpha^{(1)}, \dots)$ of

$$E_m \equiv (L_m)^{\mathbb{Z}_+}$$

such that $\alpha^{(0)}, \alpha^{(1)}, \dots$ is the sequence of types (in the previous terminology) of the mutants in the line of descent of the gene in question.

With this in mind we let $\{\nu_\tau^{(m, M)},\ \tau = 0, 1, \dots\}$ be the Markov chain of Theorem 2.1(d) with the roles of E and $P_M(x, \Lambda)$ played by E_m and

$$P_{m,M}(\boldsymbol{\alpha}, \Lambda) = (1 - u)\delta_{\boldsymbol{\alpha}}(\Lambda) + u\frac{1}{m}\sum_{j=1}^{m}\sum_{k=1}^{r}\pi_{\alpha_j^{(0)}, k}\delta_{(\alpha^{(0)}(j,k), \boldsymbol{\alpha})}(\Lambda), \tag{3.3}$$

where $u = (\theta/2M) \wedge 1$. Defining

$$E_m^0 = \{\boldsymbol{\alpha} \in E_m \colon \alpha^{(i)} \text{ and } \alpha^{(i+1)} \text{ differ at one and only one coordinate for each } i \in \mathbb{Z}_+\},$$

we observe that $\mathbb{P}\{\nu_0^{(m, M)}(E_m^0) = 1\} = 1$ implies

$$\mathbb{P}\{\nu_\tau^{(m,M)}(E_m^0) = 1 \text{ for } \tau = 0, 1, \dots\} = 1. \tag{3.4}$$

Next, we let $\{\mu_t^{(m)},\ t \geq 0\}$ be the Markov process of Theorem 2.1(a) with the roles of E and $Q(x, \Lambda)$ played by E_m and

$$Q_m(\boldsymbol{\alpha}, \Lambda) = \frac{1}{m}\sum_{j=1}^{m}\sum_{k=1}^{r}\pi_{\alpha_j^{(0)}, k}\delta_{(\alpha^{(0)}(j,k), \boldsymbol{\alpha})}(\Lambda). \tag{3.5}$$

Let B_m, $\mathscr{D}_m^+$ and G_m be the corresponding B, $\mathscr{D}^+$ and G. By Theorem 2.1(d), if $\nu_0^{(m, M)} \Rightarrow \mu_0^{(m)}$ in $\mathscr{P}(E_m)$ as $M \to \infty$, then $\{\nu_{[Mt]}^{(m,M)},\ t \geq 0\} \Rightarrow \{\mu_t^{(m)},\ t \geq 0\}$ in $D_{\mathscr{P}(E_m)}[0, \infty)$ as $M \to \infty$. Also, arguing as in the proof of Theorem 2.3, $\{\mu_t^{(m)}, t \geq 0\}$ has a unique stationary distribution $\tilde{\mu}^{(m)} \in \mathscr{P}(\mathscr{P}(E_m))$. This measure-valued version of the m-site model is closely related to the diffusion $\mathbf{p}^m(\cdot)$ described earlier, as we now demonstrate.

Define $\xi_0\colon E_m \to L_m$ by $\xi_0(\boldsymbol{\alpha}) = \alpha^{(0)}$ and $\tilde{\xi}_0\colon \mathscr{P}(E_m) \to \Delta_m$ by $\tilde{\xi}_0(\mu) = (\mu\xi_0^{-1}(\{\alpha\}))_{\alpha \in L_m}$. (The dependence of ξ_0 and $\tilde{\xi}_0$ on m is implicit.)

PROPOSITION 3.2. *For each $m \in \mathbb{N}$, $\{\tilde{\xi}_0(\mu_t^{(m)}),\ t \geq 0\}$ solves the $C_{\Delta_m}[0, \infty)$ martingale problem for $\mathscr{A}_m$ and has $\tilde{\mu}^{(m)}\tilde{\xi}_0^{-1}$ as a stationary distribution.*

PROOF. It is enough to show that $f \circ \tilde{\xi}_0 \in \mathscr{D}_m^+$ and

$$G_m(f \circ \tilde{\xi}_0) = (\mathscr{G}_m f) \circ \tilde{\xi}_0 \tag{3.6}$$

for all $f \in C^2(\Delta_m)$. If $\mu \in \mathscr{P}(E_m)$ and $\tilde{\xi}_0(\mu) = \mathbf{p}$, then $f \circ \tilde{\xi}_0(\mu) = f((\langle \chi_{\xi_0^{-1}(\alpha)}, \mu\rangle)_{\alpha \in L_m})$, so

$$G_m(f \circ \tilde{\xi}_0)(\mu) = \tfrac{1}{2}\sum_{\alpha, \beta \in L_m} p_\alpha(\delta_{\alpha\beta} - p_\beta)f_{,\alpha\beta}(\mathbf{p}) + \sum_{\alpha \in L_m}\langle B_m\chi_{\xi_0^{-1}(\alpha)}, \mu\rangle f_{,\alpha}(\mathbf{p}),$$

where

$$\begin{aligned}\langle B_m \chi_{\xi_0^{-1}(\alpha)}, \mu\rangle &= \tfrac{1}{2}\theta(1/m)\sum_{j=1}^{m}\int\sum_{k=1}^{r}\pi_{\beta_j^{(0)},k}\left[\chi_{\{\alpha\}}\left(\beta^{(0)}(j,k)\right)-\chi_{\{\alpha\}}\left(\beta^{(0)}\right)\right]\mu(d\boldsymbol{\beta})\\ &= -\tfrac{1}{2}\theta p_\alpha + \tfrac{1}{2}\theta(1/m)\sum_{j=1}^{m}\int\sum_{k=1}^{r}\pi_{\beta_j^{(0)},k}\chi_{\{\alpha\}}\left(\beta^{(0)}(j,k)\right)\mu(d\boldsymbol{\beta}),\end{aligned}$$

the second term on the right-hand side of which is

$$\begin{aligned}&\tfrac{1}{2}\theta(1/m)\sum_{j=1}^{m}\int\sum_{k=1}^{r}\pi_{\beta_j^{(0)},k}\delta_{\alpha_j,k}\sum_{l=1}^{r}\chi_{\{\alpha(j,l)\}}\left(\beta^{(0)}\right)\mu(d\boldsymbol{\beta})\\ &\quad= \tfrac{1}{2}\theta(1/m)\sum_{j=1}^{m}\sum_{l=1}^{r}\int\pi_{l,\alpha_j}\chi_{\{\alpha(j,l)\}}\left(\beta^{(0)}\right)\mu(d\boldsymbol{\beta})\\ &\quad= \tfrac{1}{2}\theta(1/m)\sum_{j=1}^{m}\sum_{l=1}^{r}\pi_{l,\alpha_j}p_{\alpha(j,l)}.\end{aligned}$$

This proves (3.6). □

We now indicate the relationship between $\{\mu_t,\ t\ge 0\}$ and the limit in some sense of $\{\mu_t^{(m)},\ t\ge 0\}$. Observe first that the analogue of (3.4) holds: if $\mathbb{P}\{\mu_0^{(m)}(E_m^0)=1\}=1$, then

$$\mathbb{P}\left\{\mu_t^{(m)}\left(E_m^0\right)=1 \text{ for all } t\ge 0\right\}=1; \tag{3.7}$$

for if we define $\phi\in\mathscr{D}_m^+$ by $\phi(\mu)=\mu(E_m^0)$, then $G_m\phi=0$ since $B_m\chi_{E_m^0}=0$, so $\mu_t^{(m)}(E_m^0)$ is a martingale with almost all sample paths in $C_{[0,1]}[0,\infty)$ by Theorem 2.1(b). Next, define $\rho_m\colon E_m\to E$ by

$$\rho_m(\boldsymbol{\alpha})=\left(\frac{j_0}{m},\frac{j_1}{m},\dots\right),\qquad \boldsymbol{\alpha}\in E_m^0,$$

where $j_i\in\{1,\dots,m\}$ is the coordinate at which $\alpha^{(i)}$ and $\alpha^{(i+1)}$ differ, and $\rho_m\equiv(0,0,\dots)$ on $E_m-E_m^0$. Also, define $\tilde{\rho}_m\colon\mathscr{P}(E_m)\to\mathscr{P}(E)$ by $\tilde{\rho}_m(\mu)=\mu\rho_m^{-1}$.

Theorem 3.3. *If $\mu_0^{(m)}(E_m^0)=1$ for each $m\in\mathbb{N}$ and $\mu_0^{(m)}\rho_m^{-1}\Rightarrow\mu_0$ in $\mathscr{P}(E)$ as $m\to\infty$, then $\{\mu_t^{(m)}\rho_m^{-1},\ t\ge 0\}\Rightarrow\{\mu_t,\ t\ge 0\}$ in $C_{\mathscr{P}(E)}[0,\infty)$ as $m\to\infty$. Also, $\tilde{\mu}^{(m)}\tilde{\rho}_m^{-1}\Rightarrow\tilde{\mu}$ on $\mathscr{P}(E)$ as $m\to\infty$.*

Proof. Note first that $\{\mu_t^{(m)}\rho_m^{-1},\ t\ge 0\}$ is the process of Theorem 2.1(a) with the roles of E and $Q(x,\Lambda)$ played by $\{m^{-1},2m^{-1},\dots,1\}^{\mathbf{Z}_+}$ and

$$\hat{Q}_m(\mathbf{x},\Lambda)=\frac{1}{m}\sum_{j=1}^{m}\delta_{(j/m,\mathbf{x})}(\Lambda).$$

This follows from (3.7) and that fact that

$$\begin{aligned} B_m(f \circ \rho_m)(\boldsymbol{\alpha}) &= \tfrac{1}{2}\theta(1/m) \sum_{j=1}^{m} \sum_{k=1}^{r} \pi_{\alpha_j^{(0)},k}\big[f(j/m, \rho_m(\boldsymbol{\alpha})) - f(\rho_m(\boldsymbol{\alpha}))\big] \\ &= \big(\hat{B}_m f\big) \circ \rho_m(\boldsymbol{\alpha}) \end{aligned} \tag{3.8}$$

for all $\boldsymbol{\alpha} \in E_m^0$, where $\hat{B}_m$ corresponds to $\hat{Q}_m$ as in (2.1). By Theorem 2.1(e), it suffices for the first assertion to show that

$$\lim_{m\to\infty} \sup_{x \in E_m} \big|\hat{B}_m f(\mathbf{x}) - Bf(\mathbf{x})\big| = 0, \qquad f \in C(E). \tag{3.9}$$

But the supremum in (3.9) is bounded by

$$\begin{aligned} &\tfrac{1}{2}\theta \sup_{\mathbf{x}\in E} \left| (1/m) \sum_{j=1}^{m} f(j/m, \mathbf{x}) - \int_0^1 f(y, \mathbf{x}) \lambda(dy) \right| \\ &\quad \le \tfrac{1}{2}\theta \sup_{\mathbf{x}\in E} w_{f(\cdot,\mathbf{x})}(m^{-1}) \le \tfrac{1}{2}\theta \hat{w}_f(m^{-1}), \end{aligned}$$

where w_g is the modulus of continuity of $g \in C[0,1]$ and $\hat{w}_f$ is that of $f \in C(E)$ with respect to any metric d on E for which $d(\mathbf{x},\mathbf{y}) = |x_0 - y_0|$ for all $\mathbf{x}, \mathbf{y} \in E$ with $x_i = y_i$ for $i = 1, 2, \ldots$. The second assertion follows by a standard argument using the uniqueness in Theorem 2.3. □

4. Equations for tree probabilities.

Given $n \in \mathbb{N}$, let

$$\mathscr{T}_n = \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in E^n \colon (\mathbf{x}_1, \ldots, \mathbf{x}_n) \text{ is a tree, i.e., (1.7)–(1.9) hold}\}.$$

We define equivalence relations $\sim$ and $\approx$ on $\mathscr{T}_n$ as follows. Let us say that

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n) \sim (\mathbf{y}_1, \ldots, \mathbf{y}_n)$$

if there exists a bijection $\zeta\colon [0,1] \to [0,1]$ with $y_{ij} = \zeta(x_{ij})$ for $i = 1, \ldots, n$ and $j = 0, 1, \ldots,$ and that

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n) \approx (\mathbf{y}_1, \ldots, \mathbf{y}_n)$$

if there exists a bijection $\zeta\colon [0,1] \to [0,1]$ and a permutation σ of $(1, \ldots, n)$ such that $y_{\sigma(i),j} = \zeta(x_{ij})$ for $i = 1, \ldots, n$ and $j = 0, 1, \ldots$. Equivalence classes in the quotient set $\mathscr{T}_n/\!\sim$ might be referred to as unlabelled trees, while those in $\mathscr{T}_n/\!\approx$ might be called unlabelled, unordered trees. It is an important open problem to determine (1.13) for all $T \in \mathscr{T}_n/\!\approx$.

Given $T \in \mathscr{T}_n/\!\sim$, let $[T]$ denote the equivalence class in $\mathscr{T}_n/\!\approx$ containing T, and let $c(T)$ denote the number of equivalence classes $T' \in \mathscr{T}_n/\!\sim$ with $T' \subset [T]$. Then

$$\mu^n([T]) = c(T)\mu^n(T) \tag{4.1}$$

for each $\mu \in \mathscr{P}(E)$ since μ^n is invariant under permutation of coordinates of E^n. Consequently, it will suffice to consider (1.13) for $T \in \mathscr{T}_n/\!\sim$, and in fact we can restrict our attention to those $T \in \mathscr{T}_n/\!\sim$ with the property that repeated coordinates of $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in T$ are adjacent.

Fix $d \in \mathbb{N}$ until further notice. For $\mathbf{n} \in \mathbb{N}^d$ and $n = n_1 + \cdots + n_d$, define $\Phi_{\mathbf{n}}\colon E^d \to E^n$ by

$$\Phi_{\mathbf{n}}(\mathbf{x}_1, \ldots, \mathbf{x}_d) = (\mathbf{x}_1, \ldots, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_2, \ldots, \mathbf{x}_d, \ldots, \mathbf{x}_d),$$

where $\mathbf{x}_k$ appears n_k times for $k = 1, \ldots, d$, and note that $\Phi_{\mathbf{n}}$ induces a map from $\mathscr{T}_d/\sim$ into $\mathscr{T}_n/\sim$. Define the shift operator $\mathscr{S}\colon E \to E$ by

$$\mathscr{S}\mathbf{x} = \mathscr{S}(x_0, x_1, \ldots) = (x_1, x_2, \ldots),$$

and for $k = 1, \ldots, d$ define $\mathscr{S}_k\colon E^d \to E^d$ by

$$\mathscr{S}_k(\mathbf{x}_1, \ldots, \mathbf{x}_d) = (\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathscr{S}\mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_d),$$

and note that $\mathscr{S}_k$ induces a map from $\mathscr{T}_d/\sim$ into $\mathscr{T}_d/\sim$.

For $T \in \mathscr{T}_d/\sim$, $\mathbf{n} \in \mathbb{N}^d$, and $n = n_1 + \cdots + n_d$, define the function $\phi_{T,\mathbf{n}}$ on $\mathscr{P}(E)$ by

$$\phi_{T,\mathbf{n}}(\mu) = \mu^n(\Phi_{\mathbf{n}}(T)),$$

and define

$$\psi_{T,\mathbf{n}} = \frac{1}{2} \sum_{k\colon n_k \geq 2} n_k(n_k - 1)\phi_{T,\mathbf{n}-\mathbf{e}_k} + \frac{1}{2}\theta \sum_{\substack{k\colon n_k = 1,\\ x_{k0} \text{ distinct}}} \phi_{\mathscr{S}_k T,\mathbf{n}}$$
$$- \frac{1}{2}n(n - 1 + \theta)\phi_{T,\mathbf{n}},$$

where $\mathbf{e}_k = (\delta_{k1}, \ldots, \delta_{kd})$ and "x_{k0} distinct" means $x_{k0} \neq x_{ij}$ for all $(\mathbf{x}_1, \ldots, \mathbf{x}_d) \in T$ and $(i, j) \neq (k, 0)$. Note that $\phi_{T,\mathbf{n}}(\mu)$ is the probability that in an ordered random sample of size n from a population with types distributed according to μ, the sample forms a tree of class $\Phi_{\mathbf{n}}(T)$.

It will be convenient to define

$$(\mathscr{T}_d/\sim)_0 = \{T \in \mathscr{T}_d/\sim\ : \mathbf{x}_1, \ldots, \mathbf{x}_d \text{ are distinct for all } (\mathbf{x}_1, \ldots, \mathbf{x}_d) \in T\}.$$

THEOREM 4.1. *For each* $(T, \mathbf{n}) \in (\mathscr{T}_d/\sim)_0 \times \mathbb{N}^d$, $(\phi_{T,\mathbf{n}}, \psi_{T,\tilde{\mathbf{n}}}) \in A^+$.

PROOF. Let $n = n_1 + \cdots + n_d$. Note that if $f \in B(E^n)$ and if $\phi \in \mathscr{D}^+$ is given by $\phi(\mu) = \langle f, \mu^n \rangle$, then

$$G\phi(\mu) = \sum_{1 \leq i < j \leq n} \langle \Psi_{ij} f, \mu^{n-1} \rangle + \frac{1}{2}\theta \sum_{i=1}^{n} \langle \Lambda_i f, \mu^n \rangle \tag{4.2}$$
$$- \frac{1}{2}n(n - 1 + \theta)\langle f, \mu^n \rangle,$$

where Ψ_{ij} is as in (2.3) and

$$\Lambda_i f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \int_0^1 f(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, (y, \mathbf{x}_i), \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)\lambda(dy).$$

Let $f = \chi_{\Phi_{\mathbf{n}}(T)}$. Then $\phi = \phi_{T,\mathbf{n}}$,

$$\Psi_{ij} f = \begin{cases} \chi_{\Phi_{\mathbf{n}-\mathbf{e}_k}(T)}, & \text{if } n_1 + \cdots + n_{k-1} + 1 \leq i < j \leq n_1 + \cdots + n_k, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\Lambda_i f = \begin{cases} \chi_{\Phi_{\mathbf{n}}(\mathscr{S}_k T)}, & \text{if } i = n_1 + \cdots + n_k,\ n_k = 1 \text{ and } x_{k0} \text{ is distinct,} \\ 0, & \text{otherwise,} \end{cases}$$

the latter because, under the three stated conditions, for each $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in E^n$

$$\chi_{\Phi_{\mathbf{n}}(T)}(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, (y, \mathbf{x}_i), \mathbf{x}_{i+1}, \dots, \mathbf{x}_n) = \chi_{\Phi_{\mathbf{n}}(\mathscr{S}_k T)}(\mathbf{x}_1, \dots, \mathbf{x}_n)$$

for all but countably many $y \in [0,1]$, hence for λ-a.e. $y \in [0,1]$. Consequently, $G\phi_{T,\mathbf{n}} = \psi_{T,\mathbf{n}}$. □

If $d \geq 2$, define $\mathscr{R}_k$: $E^d \to E^{d-1}$ for $k = 1, \dots, d$ by

$$\mathscr{R}_k(\mathbf{x}_1, \dots, \mathbf{x}_d) = (\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \dots, \mathbf{x}_d)$$

and note that $\mathscr{R}_k$ induces a map from $\mathscr{T}_d/\sim$ into $\mathscr{T}_{d-1}/\sim$. Define $\mathscr{R}_k$: $\mathbb{N}^d \to \mathbb{N}^{d-1}$ similarly.

So far, $d \in \mathbb{N}$ has been fixed. We now allow d to vary with the understanding that the dependence of $\mathbf{e}_k$, $\mathscr{S}_k$, etc., on d is implicit. Let

$$\mathscr{T}^* = \bigcup_{d=1}^{\infty} \left[(\mathscr{T}_d/\sim)_0 \times \mathbb{N}^d \right], \tag{4.3}$$

and regard d and $n = n_1 + \cdots + n_d$ as functions of $\mathbf{n} = (n_1, \dots, n_d)$. $\mathscr{T}^*$ can be regarded as the set of all unlabelled ordered trees (with multiplicities) of the sort considered in Theorem 4.1.

COROLLARY 4.2. *The system of linear equations*

$$\begin{aligned} n(n-1+\theta)p(T,\mathbf{n}) = & \sum_{k:\, n_k \geq 2} n_k(n_k - 1)p(T, \mathbf{n} - \mathbf{e}_k) \\ & + \theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct} \\ \mathscr{S}\mathbf{x}_k \neq \mathbf{x}_j,\, \forall j}} p(\mathscr{S}_k T, \mathbf{n}) \\ & + \theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct}}} \sum_{j:\, \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} p\big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j)\big), \end{aligned} \tag{4.4}$$

where $(T, \mathbf{n}) \in \mathscr{T}^*$ *and* $(\mathbf{x}_1, \dots, \mathbf{x}_d)$ *denotes an arbitrary element of* T, *has a unique solution satisfying* $p(\mathscr{T}_1, (1)) = 1$. *It is given by*

$$p(T, \mathbf{n}) = \int \phi_{T,\mathbf{n}}\, d\tilde{\mu}. \tag{4.5}$$

PROOF. Let us define the degree of the quantity $p(T, \mathbf{n})$ to be $s + n$, where for every $(\mathbf{x}_1, \dots, \mathbf{x}_d) \in T$, s is the number of elements of $[0,1]$ appearing in at least one but not all of the sequences $\mathbf{x}_1, \dots, \mathbf{x}_d$. Then each of the terms on the right-hand side of (4.4) has degree $s + n - 1$ and uniqueness follows by induction. As for existence, Theorem 4.1 implies that $\int \psi_{T,\mathbf{n}}\, d\tilde{\mu} = 0$ for all $(T, \mathbf{n}) \in \mathscr{T}^*$.

Therefore, to show that (4.5) satisfies (4.4), it is enough to observe that if $(T, \mathbf{n}) \in \mathscr{T}^*$, $d \geq 2$, and j and k are as in the third term on the right-hand side of (4.4), then $\phi_{\mathscr{S}_k T, \mathbf{n}}(\mu) = \phi_{\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)}(\mu)$ for each $\mu \in \mathscr{P}(E)$ by the invariance of μ^n under permutation of coordinates of E^n.

It remains only to show that (4.5) satisfies the condition $p(\mathscr{T}_1, (1)) = 1$. Given $\Lambda \in \mathscr{B}(E)$, define $\phi \in \mathscr{D}^+$ by $\phi(\mu) = \mu(\Lambda)$ and observe that $G\phi(\mu) = \frac{1}{2}\theta\{(\lambda \times \mu)(\Lambda) - \mu(\Lambda)\}$ for all $\mu \in \mathscr{P}(E)$, hence

$$\int \mu(\Lambda)\tilde{\mu}(d\mu) = \int (\lambda \times \mu)(\Lambda)\tilde{\mu}(d\mu). \tag{4.6}$$

If $\Lambda = \Gamma_0 \times \cdots \times \Gamma_{r-1} \times E$, where $r \in \mathbb{N}$ and $\Gamma_0, \ldots, \Gamma_{r-1} \in \mathscr{B}[0,1]$, then repeated application of (4.6) gives

$$\begin{aligned} \int \mu(\Lambda)\tilde{\mu}(d\mu) &= \lambda(\Gamma_0)\int \mu(\mathscr{S}\Lambda)\tilde{\mu}(d\mu) \\ &= \lambda(\Gamma_0) \cdots \lambda(\Gamma_{r-1})\int \mu(\mathscr{S}^r\Lambda)\tilde{\mu}(d\mu) \\ &= \lambda(\Gamma_0) \cdots \lambda(\Gamma_{r-1}) \\ &= \lambda^\infty(\Lambda), \end{aligned} \tag{4.7}$$

where λ^∞ is the product measure $\lambda \times \lambda \times \cdots$. [Note that the product σ-algebra $\prod_0^\infty \mathscr{B}[0,1]$ coincides with $\mathscr{B}(E)$.] It follows that the left-hand side of (4.7) equals the right-hand side for all $\Lambda \in \mathscr{B}(E)$, in particular, for $\Lambda = \mathscr{T}_1$, and the proof is complete. □

We note that (1.14) is a special case of (4.4).

COROLLARY 4.3. *The linear system of differential equations*

$$\begin{aligned} 2\frac{d}{dt}p(T, \mathbf{n}; t) = &\sum_{k:\, n_k \geq 2} n_k(n_k - 1)p(T, \mathbf{n} - \mathbf{e}_k; t) \\ &+\theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct} \\ \mathscr{S}\mathbf{x}_k \neq \mathbf{x}_j,\, \forall j}} p(\mathscr{S}_k T, \mathbf{n}; t) \\ &+\theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct}}} \sum_{j:\, \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} p\big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j); t\big) \\ &-n(n - 1 + \theta)p(T, \mathbf{n}; t), \end{aligned} \tag{4.8}$$

where $(T, \mathbf{n}) \in \mathscr{T}^*$ *and* $t \geq 0$, *is satisfied by*

$$p(T, \mathbf{n}; t) = \mathbb{E}\left[\phi_{T,\mathbf{n}}(\mu_t)\right]. \tag{4.9}$$

[*Clearly,* (4.8) *has a unique solution for a given set of initial conditions.*]

Proof. Let $(T, \mathbf{n}) \in \mathscr{T}^*$. By Theorems 4.1 and 2.1(b),

$$\mathbb{E}[\phi_{T,\mathbf{n}}(\mu_t)] - \mathbb{E}[\phi_{T,\mathbf{n}}(\mu_0)] = \int_0^t \mathbb{E}[\psi_{T,\mathbf{n}}(\mu_v)]\, dv \tag{4.10}$$

for all $t \geq 0$. Using the observation about $\phi_{\mathscr{S}_k T, \mathbf{n}}$ in the proof of Corollary 4.2, we conclude (as in the proof of Lemma 2.2) that $\mathbb{E}[\phi_{T,\mathbf{n}}(\mu_t)]$ is continuously differentiable in $t \geq 0$. The result follows by differentiating (4.10). □

Let P_d denote the set of permutations of $(1, \ldots, d)$. Given $T \in \mathscr{T}_d/\sim$, $\mathbf{n} \in \mathbb{N}^d$, and $\sigma \in P_d$, define

$$T_\sigma = \{(\mathbf{x}_{\sigma(1)}, \ldots, \mathbf{x}_{\sigma(d)}) : (\mathbf{x}_1, \ldots, \mathbf{x}_d) \in T\}$$

and

$$\mathbf{n}_\sigma = (n_{\sigma(1)}, \ldots, n_{\sigma(d)}).$$

For each $(T, \mathbf{n}) \in \mathscr{T}^*$ let

$$a(T, \mathbf{n}) = \#\{\sigma \in P_d : T_\sigma = T, \mathbf{n}_\sigma = \mathbf{n}\}$$

and

$$a(\mathbf{n}) = \#\{\sigma \in P_d : \mathbf{n}_\sigma = \mathbf{n}\}.$$

Then $a(T, \mathbf{n})$ divides $a(\mathbf{n})$, and

$$a(\mathbf{n}) = a_1! a_2! \cdots a_n!, \tag{4.11}$$

where $a_i = \#\{k : n_k = i\}$ for $i = 1, \ldots, n$.

As we pointed out earlier in this section, it is actually $\int \hat{\phi}_{T,\mathbf{n}}\, d\tilde{\mu}$ and $\mathbb{E}[\hat{\phi}_{T,\mathbf{n}}(\mu_t)]$ that are of primary interest, where $\hat{\phi}_{T,\mathbf{n}}$ is defined on $\mathscr{P}(E)$ for each $(T, \mathbf{n}) \in \mathscr{T}^*$ by

$$\hat{\phi}_{T,\mathbf{n}}(\mu) = \mu^n([\Phi_{\mathbf{n}}(T)]);$$

recalling (4.1),

$$\hat{\phi}_{T,\mathbf{n}} = \frac{n!}{n_1! \cdots n_d!\, a(T, \mathbf{n})} \phi_{T,\mathbf{n}}, \tag{4.12}$$

because

$$\begin{aligned} c(\Phi_{\mathbf{n}}(T)) &= \#\{T' \in \mathscr{T}_n/\sim\ : T' \subset [\Phi_{\mathbf{n}}(T)]\} \\ &= \#\{\Phi_{\mathbf{n}}(T)_\sigma : \sigma \in P_n\} \\ &= n!/\#\{\sigma \in P_n : \Phi_{\mathbf{n}}(T)_\sigma = \Phi_{\mathbf{n}}(T)\} \\ &= \frac{n!}{n_1! \cdots n_d!\, a(T, \mathbf{n})}. \end{aligned}$$

Remark 4.4. Clearly, Theorem 4.1 and Corollaries 4.2 and 4.3 can be restated in terms of $\hat{\phi}_{T,\mathbf{n}}$ using (4.12). Indeed, it follows directly from Theorem

4.1 that for each $(T,\mathbf{n}) \in \mathscr{T}^*$, $(\hat{\phi}_{T,\mathbf{n}}, \hat{\psi}_{T,\mathbf{n}}) \in A^+$, where

$$
\begin{aligned}
\hat{\psi}_{T,\mathbf{n}} = {}& \tfrac{1}{2}n \sum_{k:\, n_k \ge 2} (n_k - 1) \frac{a(T, \mathbf{n} - \mathbf{e}_k)}{a(T,\mathbf{n})} \hat{\phi}_{T,\mathbf{n}-\mathbf{e}_k} \\
&+ \tfrac{1}{2}\theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct} \\ \mathscr{S}\mathbf{x}_k \ne \mathbf{x}_j,\, \forall j}} \frac{a(\mathscr{S}_k T, \mathbf{n})}{a(T,\mathbf{n})} \hat{\phi}_{\mathscr{S}_k T, \mathbf{n}} \\
&+ \tfrac{1}{2}\theta \sum_{\substack{k:\, n_k = 1 \\ x_{k0} \text{ distinct}}} \sum_{j:\, \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} (n_j + 1) \frac{a(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j))}{a(T,\mathbf{n})} \hat{\phi}_{\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)} \\
&- \tfrac{1}{2}n(n - 1 + \theta)\hat{\phi}_{T,\mathbf{n}}.
\end{aligned}
$$

Also of interest in a (ordered or unordered) random sample of size $n \in \mathbb{N}$ are the number of segregating sites and the sample configuration. Recall that a site $z \in [0,1]$ is segregating with respect to a sample of types $\mathbf{x}_1, \dots, \mathbf{x}_n$ if z appears in at least one but not all of the sequences $\mathbf{x}_1, \dots, \mathbf{x}_n$. Two trees that are equivalent under $\sim$ have the same number of segregating sites, so we can refer to the number of segregating sites of an unlabelled tree $T \in \mathscr{T}_n/\sim$. In fact it will be convenient to define for $(s, n) \in \mathbb{Z}_+ \times \mathbb{N}$,

$$
\mathscr{T}_{s,n} = \{(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathscr{T}_n \colon (\mathbf{x}_1, \dots, \mathbf{x}_n) \text{ has } s \text{ segregating sites}\}
$$

(implying

$$
\mathscr{T}_{s,n}/\sim \; = \{T \in \mathscr{T}_n/\sim \; : T \text{ has } s \text{ segregating sites}\})
$$

and

$$
(\mathscr{T}_{s,n}/\sim)_0 = \{T \in \mathscr{T}_{s,n}/\sim \; : \mathbf{x}_1, \dots, \mathbf{x}_n \text{ are distinct for all } (\mathbf{x}_1, \dots, \mathbf{x}_n) \in T\}.
$$

In addition, $\mathbf{x}_1, \dots, \mathbf{x}_n$ have sample configuration $\mathbf{n} \in \mathbb{N}^d$ if $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ has d distinct components with respective multiplicities $n_1, \dots, n_d$. (Here the order of $n_1, \dots, n_d$ is irrelevant.)

For $s \in \mathbb{Z}_+$, $n \in \mathbb{N}$, $d \in \mathbb{N}$ and $\mathbf{n} \in \mathbb{N}^d$, define the functions $\phi_{s,\mathbf{n}}$, $\hat{\phi}_{s,\mathbf{n}}$, $\phi_{\mathbf{n}}$, $\hat{\phi}_{\mathbf{n}}$ and $\phi_{s,n}$ on $\mathscr{P}(E)$ by

$$
\begin{aligned}
\phi_{s,\mathbf{n}}(\mu) &= \mu^n(\cup\{\Phi_{\mathbf{n}}(T) \colon T \in (\mathscr{T}_{s,d}/\sim)_0\}), \\
\hat{\phi}_{s,\mathbf{n}}(\mu) &= \mu^n(\cup\{[\Phi_{\mathbf{n}}(T)] \colon T \in (\mathscr{T}_{s,d}/\sim)_0\}), \\
\phi_{\mathbf{n}}(\mu) &= \mu^n(\cup\{\Phi_{\mathbf{n}}(T) \colon T \in (\mathscr{T}_d/\sim)_0\}), \\
\hat{\phi}_{\mathbf{n}}(\mu) &= \mu^n(\cup\{[\Phi_{\mathbf{n}}(T)] \colon T \in (\mathscr{T}_d/\sim)_0\}), \\
\phi_{s,n}(\mu) &= \mu^n(\mathscr{T}_{s,n}),
\end{aligned}
$$

and define

$$\psi_{s,\mathbf{n}} = \begin{cases} \frac{1}{2} \sum_{k:\, n_k \ge 2} n_k(n_k - 1)\phi_{s,\mathbf{n}-\mathbf{e}_k} + \frac{1}{2}\theta \#\{k\colon n_k = 1\}\phi_{s-1,\mathbf{n}} \\ \qquad + \frac{1}{2}\theta \sum_{k:\, n_k = 1} \sum_{j:\, j \ne k} \phi_{s-1,\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)} - \frac{1}{2}n(n - 1 + \theta)\phi_{s,\mathbf{n}}, & \text{if } \mathbf{n} \ne (1), \\ 0, & \text{if } \mathbf{n} = (1), \end{cases}$$

$$\psi_{\mathbf{n}} = \frac{1}{2} \sum_{k:\, n_k \ge 2} n_k(n_k - 1)\phi_{\mathbf{n}-\mathbf{e}_k} + \frac{1}{2}\theta\#\{k\colon n_k = 1\}\phi_{\mathbf{n}}$$
$$+ \frac{1}{2}\theta \sum_{k:\, n_k = 1} \sum_{j:\, j \ne k} \phi_{\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)} - \frac{1}{2}n(n - 1 + \theta)\phi_{\mathbf{n}},$$

$$\psi_{s,n} = \begin{cases} \frac{1}{2}n(n-1)\phi_{s,n-1} + \frac{1}{2}\theta n \phi_{s-1,n} - \frac{1}{2}n(n-1+\theta)\phi_{s,n}, & \text{if } n \ge 2, \\ 0, & \text{if } n = 1, \end{cases}$$

where $\phi_{-1,\mathbf{n}} \equiv \phi_{-1,n} \equiv 0$. Clearly, $\hat{\phi}_{s,\mathbf{n}}(\mu)$ is the joint probability of s segregating sites and sample configuration $\mathbf{n}$ in a random sample from a population with types distributed according to μ, and $\hat{\phi}_{\mathbf{n}}(\mu)$ and $\phi_{s,n}(\mu)$ are the marginal probabilities. Given $d \in \mathbb{N}$ and $(s, \mathbf{n}) \in \mathbb{Z}_+ \times \mathbb{N}^d$, observe that (4.12) implies

$$\mu^n([\Phi_{\mathbf{n}}(T)]) = \left(\frac{a(\mathbf{n})}{a(T,\mathbf{n})}\right)^{-1} \frac{n!}{n_1! \cdots n_d! a(T,\mathbf{n})} \mu^n\Bigl(\bigcup_{\sigma \in P_d:\, \mathbf{n}_\sigma = \mathbf{n}} \Phi_{\mathbf{n}}(T_\sigma)\Bigr)$$

for all $T \in (\mathscr{T}_{s,d}/\sim)_0$, and hence

$$\hat{\phi}_{s,\mathbf{n}} = \frac{n!}{n_1! \cdots n_d! a(\mathbf{n})} \phi_{s,\mathbf{n}} \tag{4.13}$$

and

$$\hat{\phi}_{\mathbf{n}} = \frac{n!}{n_1! \cdots n_d! a(\mathbf{n})} \phi_{\mathbf{n}}. \tag{4.14}$$

The next result is the analogue of Theorem 4.1 with $\phi_{s,\mathbf{n}}$, $\phi_{\mathbf{n}}$ and $\phi_{s,n}$ in place of $\phi_{T,\mathbf{n}}$. It can be proved analogously, but instead we derive it by appealing to Theorem 4.1.

THEOREM 4.5. (a) *For each* $d \in \mathbb{N}$ *and* $(s, \mathbf{n}) \in \mathbb{Z}_+ \times \mathbb{N}^d$, $(\phi_{s,\mathbf{n}}, \psi_{s,\mathbf{n}}) \in A^+$.
(b) *For each* $d \in \mathbb{N}$ *and* $\mathbf{n} \in \mathbb{N}^d$, $(\phi_{\mathbf{n}}, \psi_{\mathbf{n}}) \in A^+$.
(c) *For each* $(s, n) \in \mathbb{Z}_+ \times \mathbb{N}$, $(\phi_{s,n}, \psi_{s,n}) \in A^+$.

PROOF. (a) Fix $d \in \mathbb{N}$ and $(s, \mathbf{n}) \in \mathbb{Z}_+ \times \mathbb{N}^d$ with $n \ge 2$. Since

$$\phi_{s,\mathbf{n}} = \sum_{T \in (\mathscr{T}_{s,d}/\sim)_0} \phi_{T,\mathbf{n}}, \tag{4.15}$$

it will suffice by Theorem 4.1 to show that

$$\psi_{s,\mathbf{n}} = \sum_{T \in (\mathscr{T}_{s,d}/\sim)_0} \psi_{T,\mathbf{n}}. \tag{4.16}$$

By (4.15), the right-hand side of (4.16) is

$$\tag{4.17} \begin{aligned} &\tfrac{1}{2} \sum_{k:\, n_k \ge 2} n_k(n_k - 1)\phi_{s,\mathbf{n}-\mathbf{e}_k} + \tfrac{1}{2}\theta \sum_{T \in (\mathcal{T}_{s,d}/\sim)_0} \sum_{\substack{k:\, n_k = 1 \\ x_{k0}\text{ distinct}}} \phi_{\mathcal{S}_k T, \mathbf{n}} \\ &\quad - \tfrac{1}{2} n(n - 1 + \theta)\phi_{s,\mathbf{n}}, \end{aligned}$$

and the double sum in (4.17) equals

$$\sum_{k:\, n_k = 1} \sum_{\substack{T \in (\mathcal{T}_{s,d}/\sim)_0: \\ x_{k0}\text{ distinct}}} \phi_{\mathcal{S}_k T, \mathbf{n}}$$

$$\tag{4.18} \begin{aligned} &= \sum_{k:\, n_k = 1} \left[\sum_{T \in (\mathcal{T}_{s-1,d}/\sim)_0} \phi_{T,\mathbf{n}} + \sum_{j:\, j \ne k} \sum_{T \in (\mathcal{T}_{s-1,d-1}/\sim)_0} \phi_{T, \mathcal{R}_k(\mathbf{n}+\mathbf{e}_j)} \right] \\ &= \sum_{k:\, n_k = 1} \left[\phi_{s-1,\mathbf{n}} + \sum_{j:\, j \ne k} \phi_{s-1, \mathcal{R}_k(\mathbf{n}+\mathbf{e}_j)} \right], \end{aligned}$$

proving (4.16). If $n = 1$, (4.18) fails, but $\phi_{0,(1)} = \phi_{\mathcal{T}_1,(1)}$ and $\phi_{s,(1)} \equiv 0$ if $s \ge 1$, so Theorem 4.1 applies.

(b) It suffices to sum the result of (a) over $s \in \mathbb{Z}_+$.

(c) In this proof only, we temporarily drop the convention that $n = n_1 + \cdots + n_d$ whenever $d \in \mathbb{N}$ and $\mathbf{n} \in \mathbb{N}^d$. Fix $s \in \mathbb{Z}_+$ and $n \ge 2$. Using (4.13),

$$\tag{4.19} \begin{aligned} \phi_{s,n} &= \sum_{d \in \mathbb{N}} \sum_{\substack{\mathbf{n} \in \mathbb{N}^d:\, \Sigma n_i = n \\ n_1 \ge \cdots \ge n_d}} \hat{\phi}_{s,\mathbf{n}} \\ &= \sum_{d \in \mathbb{N}} \sum_{\mathbf{n} \in \mathbb{N}^d:\, \Sigma n_i = n} \frac{1}{d!} \frac{n!}{n_1! \cdots n_d!} \phi_{s,\mathbf{n}}, \end{aligned}$$

so it will suffice by Theorem 4.1 to show that

$$\psi_{s,n} = \sum_{d \in \mathbb{N}} \sum_{\mathbf{n} \in \mathbb{N}^d:\, \Sigma n_i = n} \frac{1}{d!} \frac{n!}{n_1! \cdots n_d!} \psi_{s,\mathbf{n}},$$

or that

$$\tag{4.20} \begin{aligned} n(n-1)\phi_{s,n-1} &= \sum_{d \in \mathbb{N}} \sum_{\mathbf{n} \in \mathbb{N}^d:\, \Sigma n_i = n} \frac{1}{d!} \frac{n!}{n_1! \cdots n_d!} \\ &\quad \times \sum_{k:\, n_k \ge 2} n_k(n_k - 1)\phi_{s,\mathbf{n}-\mathbf{e}_k} \end{aligned}$$

and

$$\tag{4.21} \begin{aligned} n\phi_{s-1,n} &= \sum_{d \in \mathbb{N}} \sum_{\mathbf{n} \in \mathbb{N}^d:\, \Sigma n_i = n} \frac{1}{d!} \frac{n!}{n_1! \cdots n_d!} \\ &\quad \times \left[\#\{k:\, n_k = 1\}\phi_{s-1,\mathbf{n}} + \sum_{k:\, n_k = 1} \sum_{j:\, j \ne k} \phi_{s-1, \mathcal{R}_k(\mathbf{n}+\mathbf{e}_j)} \right]. \end{aligned}$$

Denoting $(d!n_1!\cdots n_d!)^{-1}n!\phi_{s,\mathbf{n}}$ by $\tilde{\phi}_{s,\mathbf{n}}$, (4.20) and (4.21) become

$$\begin{aligned} n(n-1)\sum_{d\in\mathbb{N}}\ \sum_{\mathbf{n}\in\mathbb{N}^d:\,\Sigma n_i=n-1}\tilde{\phi}_{s,\mathbf{n}} \\ = n\sum_{d\in\mathbb{N}}\ \sum_{\mathbf{n}\in\mathbb{N}^d:\,\Sigma n_i=n}\ \sum_{k:\,n_k\geq 2}(n_k-1)\tilde{\phi}_{s,\mathbf{n}-\mathbf{e}_k}\end{aligned}\tag{4.22}$$

and

$$\begin{aligned} n\sum_{d\in\mathbb{N}}\ \sum_{\mathbf{n}\in\mathbb{N}^d:\,\Sigma n_i=n}\tilde{\phi}_{s-1,\mathbf{n}} \\ = \sum_{d\in\mathbb{N}}\ \sum_{\mathbf{n}\in\mathbb{N}^d:\,\Sigma n_i=n}\#\{k\colon n_k=1\}\tilde{\phi}_{s-1,\mathbf{n}} \\ + \sum_{d\in\mathbb{N}}\ \sum_{\mathbf{n}\in\mathbb{N}^d:\,\Sigma n_i=n}\frac{1}{d}\sum_{k:\,n_k=1}\ \sum_{j:\,j\neq k}(n_j+1)\tilde{\phi}_{s-1,\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)}.\end{aligned}\tag{4.23}$$

To prove (4.22), fix $d\in\mathbb{N}$ and $\mathbf{n}\in\mathbb{N}^d$ with $\Sigma n_i=n-1$, and compare coefficients of $\tilde{\phi}_{s,\mathbf{n}}$. The coefficient on the left is $n(n-1)$ and on the right is $n\Sigma n_k=n(n-1)$. To prove (4.23), fix $d\in\mathbb{N}$ and $\mathbf{n}\in\mathbb{N}^d$ with $\Sigma n_i=n$, and compare coefficients of $\tilde{\phi}_{s-1,\mathbf{n}}$. The coefficient on the left is n and on the right is

$$\#\{k\colon n_k=1\}+\frac{1}{d+1}\sum_{k=1}^{d+1}\ \sum_{j:\,n_j\geq 2}n_j=n,$$

because $\mathscr{R}_k(\mathbf{n}^*+\mathbf{e}_j)=\mathbf{n}$ implies $\mathbf{n}^*=(n_1,\dots,n_{k-1},1,n_k,\dots,n_d)-\mathbf{e}_j$ provided $j,k\in\{1,\dots,d+1\}$ and $n_j^*\geq 1$.

The case $n=1$ follows from part (a). □

We turn to the analogue of Corollary 4.2.

COROLLARY 4.6. (a) *The system of linear equations*

$$\begin{aligned} n(n-1+\theta)p(s,\mathbf{n}) = \sum_{k:\,n_k\geq 2}n_k(n_k-1)p(s,\mathbf{n}-\mathbf{e}_k) \\ +\theta\#\{k\colon n_k=1\}p(s-1,\mathbf{n}) \\ +\theta\sum_{k:\,n_k=1}\ \sum_{j:\,j\neq k}p\big(s-1,\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)\big),\end{aligned}\tag{4.24}$$

where $d\in\mathbb{N}$, $(s,\mathbf{n})\in\mathbb{Z}_+\times\mathbb{N}^d$, $n\geq 2$, $p(-1,\mathbf{n})\equiv 0$ *and* $p(s,(1))=\delta_{s0}$, *has a unique solution. It is given by*

$$p(s,\mathbf{n})=\int\phi_{s,\mathbf{n}}\,d\tilde{\mu}.\tag{4.25}$$

(b) *The system of linear equations*

$$\begin{aligned} n(n-1+\theta)p(\mathbf{n}) = \sum_{k:\,n_k\geq 2}n_k(n_k-1)p(\mathbf{n}-\mathbf{e}_k) \\ +\theta\#\{k\colon n_k=1\}p(\mathbf{n})+\theta\sum_{k:\,n_k=1}\ \sum_{j:\,j\neq k}p\big(\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)\big),\end{aligned}\tag{4.26}$$

where $d \in \mathbb{N}$ *and* $\mathbf{n} \in \mathbb{N}^d$, *has a unique solution satisfying* $p((1)) = 1$. *It is given by*

$$p(\mathbf{n}) = \int \phi_{\mathbf{n}}\, d\tilde{\mu}. \tag{4.27}$$

Consequently, Ewens' (1972) *sampling formula holds*: *For each* $d \in \mathbb{N}$ *and* $\mathbf{n} \in \mathbb{N}^d$,

$$\int \hat{\phi}_{\mathbf{n}}\, d\tilde{\mu} = \frac{n!}{n_1 \cdots n_d a(\mathbf{n})} \frac{\theta^{d-1}}{(1+\theta)\cdots(n-1+\theta)}, \tag{4.28}$$

where $a(\mathbf{n})$ *is as in* (4.11).

(c) *The system of linear equations*

$$(n-1+\theta)p(s,n) = (n-1)p(s,n-1) + \theta p(s-1,n), \tag{4.29}$$

where $s \in \mathbb{Z}_+$, $n \geq 2$, $p(-1,n) = 0$ *and* $p(s,1) = \delta_{s0}$, *has a unique solution. It is given by*

$$p(s,n) = \int \phi_{s,n}\, d\tilde{\mu}. \tag{4.30}$$

Consequently, Watterson's (1975) *result holds*: *For each* $(s,n) \in \mathbb{Z}_+ \times \mathbb{N}$,

$$\int \phi_{s,n}\, d\tilde{\mu} = \left[\mathop{*}_{j=1}^{n-1} \text{geometric}\left(\frac{j}{j+\theta}\right)\right](\{s\}), \tag{4.31}$$

where the right-hand side is interpreted as δ_{s0} *when* $n = 1$.

PROOF. With the degrees of $p(s,\mathbf{n})$, $p(\mathbf{n})$ and $p(s,n)$ defined to be $s+n$, $d+n$ and $s+n$, respectively, uniqueness follows as in the proof of Corollary 4.2. As for existence, Theorem 4.5 implies that (4.25), (4.27) and (4.30) satisfy (4.24), (4.26) and (4.29), respectively. The boundary conditions for $n = 1$ follow from Corollary 4.2. It remains only to verify (4.28) and (4.31).

For (4.28), it is enough by (4.14) to show that

$$p(\mathbf{n}) \equiv (n_1 - 1)! \cdots (n_d - 1)! \frac{\theta^{d-1}}{(1+\theta)\cdots(n-1+\theta)}$$

satisfies (4.26), and this follows by direct substitution. As for (4.31), consider the probability generating function

$$g(\xi, n) = \sum_{s=0}^{\infty} p(s,n)\xi^s,$$

where $n \in \mathbb{N}$. By (4.29),

$$(n-1+\theta)g(\xi,n) = (n-1)g(\xi,n-1) + \theta\xi g(\xi,n)$$

for each $n \geq 2$, hence

$$\begin{aligned} g(\xi, n) &= \prod_{j=1}^{n-1} \frac{j}{j + \theta(1-\xi)} g(\xi, 1) \\ &= \prod_{j=1}^{n-1} \frac{j/(j+\theta)}{1 - (1 - j/(j+\theta))\xi}. \end{aligned}$$

□

REMARK 4.7. By (4.13) and Corollary 4.6(a), $\hat{p}(s, \mathbf{n}) \equiv \int \hat{\phi}_{s,\mathbf{n}}\, d\tilde{\mu}$ uniquely solves the system

$$\begin{aligned} &n(n-1+\theta)\hat{p}(s,\mathbf{n}) \\ &\quad = n \sum_{k:\, n_k \geq 2} (n_k - 1) \frac{a(\mathbf{n} - \mathbf{e}_k)}{a(\mathbf{n})} \hat{p}(s, \mathbf{n} - \mathbf{e}_k) \\ &\qquad + \theta \# \{k: n_k = 1\} \hat{p}(s-1, \mathbf{n}) \\ &\qquad + \theta \sum_{k:\, n_k = 1} \sum_{j:\, j \neq k} (n_j + 1) \frac{a(\mathscr{R}_k(\mathbf{n} + \mathbf{e}_j))}{a(\mathbf{n})} \hat{p}(s-1, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j)), \end{aligned} \tag{4.32}$$

where $d \in \mathbb{N}$, $(s, \mathbf{n}) \in \mathbb{Z}_+ \times \mathbb{N}^d$, $n \geq 2$, $\hat{p}(-1, \mathbf{n}) \equiv 0$ and $\hat{p}(s, (1)) = \delta_{s0}$. When $d = 2$, (4.32) is precisely Equation (2) of Strobeck (1983); it is possible to obtain an explicit solution in this case, but we do not do so here.

It is straightforward to formulate and prove the analogue of Corollary 4.3 in the context of Theorem 4.5. See Griffiths (1979, 1981) for time-dependent versions of (4.28) and (4.31).

Recall that

$$\mathscr{P}_a^0(E) = \{\mu \in \mathscr{P}_a(E): \mu^n(\mathscr{T}_n) = 1 \text{ for each } n \in \mathbb{N}\}.$$

THEOREM 4.8. *$\tilde{\mu}(\mathscr{P}_a^0(E)) = 1$. Moreover, if $\mathbb{P}\{\mu_0 \in \mathscr{P}_a^0(E)\} = 1$, then*

$$\mathbb{P}\{\mu_t \in \mathscr{P}_a^0(E) \textit{ for all } t > 0\} = 1. \tag{4.33}$$

PROOF. For each $n \in \mathbb{N}$, define $\eta_n = \sum_{s=0}^{\infty} \phi_{s,n}$ and note that $\eta_n(\mu) = \mu^n(\mathscr{T}_n)$ for all $\mu \in \mathscr{P}(E)$. Consequently, the first assertion follows from Theorem 2.1(c) and (4.31) [Corollary 4.6(c)]. As for the second, Theorem 4.5(c) implies that

$$\left(\eta_n, \tfrac{1}{2}n(n-1)(\eta_{n-1} - \eta_n)\right) \in A^+$$

for each $n \in \mathbb{N}$, where $\eta_0 \equiv 0$. Thus, by Theorem 2.1(b),

$$\eta_n(\mu_t) - 1 - \int_0^t \tfrac{1}{2}n(n-1)\left(\eta_{n-1}(\mu_v) - \eta_n(\mu_v)\right) dv$$

is a zero-mean martingale with almost all sample paths in $C_{\mathbb{R}}[0, \infty)$ for each $n \in \mathbb{N}$. Taking expectations, we find that $\mathbb{E}[\eta_n(\mu_t)] = 1$ for all $n \in \mathbb{N}$ and $t \geq 0$ by induction on n. Thus, with probability 1, $\eta_n(\mu_t) = 1$ for each $n \in \mathbb{N}$ and all rational $t \geq 0$, hence all $t \geq 0$ by continuity. By Theorem 2.1(c), the proof is complete. □

We note that the assumption $\mathbb{P}\{\mu_0 \in \mathscr{P}_a^0(E)\} = 1$ is needed for (4.33), in contrast to the situation in Theorem 2.1(c).

5. Representations for tree probabilities. We begin this section with a representation of the probabilities (4.5) in terms of what has been referred to in other contexts [Ewens (1972) and Watterson (1976)] as the frequency spectrum. For $d \in \mathbb{N}$ put

$$\Delta_d^* = \left\{ \mathbf{z} = (z_1, \dots, z_d) \in (0,1]^d \colon \sum_{i=1}^{d} z_i \le 1 \right\}.$$

Given $T \in (\mathscr{T}_d/\sim)_0$ and $H \in \mathscr{B}(\Delta_d^*)$, define the function $\nu_T(H, \cdot)\colon \mathscr{P}(E) \to \mathbb{Z}_+ \cup \{\infty\}$ by

$$\nu_T(H, \mu) = \sum_{(\mathbf{x}_1, \dots, \mathbf{x}_d) \in T} \chi_H\big(\mu(\{\mathbf{x}_1\}), \dots, \mu(\{\mathbf{x}_d\})\big). \tag{5.1}$$

We claim that $\nu_T(H, \cdot)$ is Borel measurable. [This is not obvious because the countable set of nonzero terms in the sum in (5.1) depends on μ.] For each $\mu \in \mathscr{P}(E)$ and $i \in \mathbb{N}$, define $s_i(\mu)$ and $\mathbf{l}_i(\mu)$ to be the size and location of the ith largest atom of μ. To avoid ambiguities, we require that $\mathbf{l}_i(\mu) < \mathbf{l}_{i+1}(\mu)$ if $s_i(\mu) = s_{i+1}(\mu)$, where by definition $\mathbf{x} < \mathbf{y}$ if for some $j \in \mathbb{Z}_+$, $x_0 = y_0, \dots, x_{j-1} = y_{j-1}$ and $x_j < y_j$. We leave it to the reader to check that $s_1, s_2, \dots$ and $\mathbf{l}_1, \mathbf{l}_2, \dots$ are Borel measurable. The claim follows by observing that

$$\nu_T(H, \mu) = \sum_{(i_1, \dots, i_d) \in \mathbb{N}^d} \chi_H\big(s_{i_1}(\mu), \dots, s_{i_d}(\mu)\big) \chi_T\big(\mathbf{l}_{i_1}(\mu), \dots, \mathbf{l}_{i_d}(\mu)\big) \tag{5.2}$$

for all $\mu \in \mathscr{P}(E)$.

For $T \in (\mathscr{T}_d/\sim)_0$, define the positive Borel measure ν_T on Δ_d^* by

$$\nu_T(H) = \int \nu_T(H, \mu) \tilde{\mu}(d\mu). \tag{5.3}$$

Because the sum in (5.1) has only countably many nonzero terms for each $\mu \in \mathscr{P}(E)$, it is easy to check that ν_T is indeed a measure [(5.2) is not needed for this]. Note that $\nu_T(H)$ is the $\tilde{\mu}(d\mu)$-expected number of trees $(\mathbf{x}_1, \dots, \mathbf{x}_d)$ in the equivalence class T for which the vector of frequencies $(\mu(\{\mathbf{x}_1\}), \dots, \mu(\{\mathbf{x}_d\}))$ belongs to H.

THEOREM 5.1. *For each* $(T, \mathbf{n}) \in \mathscr{T}^*$,

$$\int \phi_{T, \mathbf{n}}\, d\tilde{\mu} = \int_{\Delta_d^*} z_1^{n_1} \cdots z_d^{n_d} \nu_T(d\mathbf{z}). \tag{5.4}$$

PROOF. For each $f \in B(\Delta_d^*)$, define $\bar{f} \in B(\mathbb{R}^d)$ by $\bar{f} = f$ on Δ_d^*, $\bar{f} = 0$ elsewhere. Then

$$\int_{\Delta_d^*} f\, d\nu_T = \int \sum_{(\mathbf{x}_1, \dots, \mathbf{x}_d) \in T} \bar{f}\big(\mu(\{\mathbf{x}_1\}), \dots, \mu(\{x_d\})\big) \tilde{\mu}(d\mu) \tag{5.5}$$

for simple functions $f \in B(\Delta_d^*)$ by (5.3) and (5.1), hence for all nonnegative $f \in B(\Delta_d^*)$ by the monotone convergence theorem [the justification of which is similar to the proof that (5.1) defines a measure]. Since

$$\phi_{T,\mathbf{n}}(\mu) = \sum_{(\mathbf{x}_1,\dots,\mathbf{x}_d) \in T} \mu(\{\mathbf{x}_1\})^{n_1} \cdots \mu(\{\mathbf{x}_d\})^{n_d} \tag{5.6}$$

for all $\mu \in \mathscr{P}_a(E)$, (5.4) follows from (5.5) with $f(\mathbf{z}) \equiv z_1^{n_d} \cdots z_d^{n_d}$ and from Theorem 2.3. □

It is easy to verify the well-known result that

$$\nu_{\mathscr{T}_1}(dz) = \theta z^{-1}(1-z)^{\theta-1}\,dz, \qquad 0 < z \le 1, \tag{5.7}$$

but the situation is much more complicated when $d \ge 2$.

The remaining three representations we consider involve certain stochastic processes that are considerably simpler than $\{\mu_t,\ t \ge 0\}$. In the first we show how to represent the probabilities in Corollaries 4.2 and 4.3 in terms of a certain dual Markov process in the countable set $\mathscr{T}^*$. Define $\{(T(t), \mathbf{n}(t)),\ t \ge 0\}$ to be a pure jump Markov process in $\mathscr{T}^*$ with transitions

$$(T,\mathbf{n}) \to \begin{cases} (T, \mathbf{n} - \mathbf{e}_k), & \text{rate } \tfrac{1}{2} n_k(n_k - 1),\ k\colon n_k \ge 2, \\ (\mathscr{S}_k T, \mathbf{n}), & \text{rate } \tfrac{1}{2}\theta,\ k\colon n_k = 1,\ x_{k0} \text{ distinct},\ \mathscr{S}\mathbf{x}_k \ne \mathbf{x}_j,\ \forall\, j, \\ \big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j)\big), & \text{rate } \tfrac{1}{2}\theta,\ \ k\colon n_k = 1,\ x_{k0} \text{ distinct},\ j\colon \mathscr{S}\mathbf{x}_k = \mathbf{x}_j, \end{cases}$$

and note that the process in effect "chops down the tree" $(T(0), \mathbf{n}(0))$. Also, define $c\colon \mathscr{T}^* \to (-\infty, 0]$ by

$$\begin{aligned} c(T,\mathbf{n}) = {} & \tfrac{1}{2} \sum_{k\colon n_k \ge 2} n_k(n_k - 1) + \tfrac{1}{2}\theta \#\{k\colon n_k = 1,\ x_{k0} \text{ distinct}\} \\ & - \tfrac{1}{2} n(n - 1 + \theta). \end{aligned}$$

THEOREM 5.2. *For all* $\mu \in \mathscr{P}(E)$, $(T,\mathbf{n}) \in \mathscr{T}^*$ *and* $t \ge 0$,

$$\mathbb{E}_\mu[\phi_{T,\mathbf{n}}(\mu_t)] = \mathbb{E}_{(T,\mathbf{n})}\left[\phi_{T(t),\mathbf{n}(t)}(\mu) \exp\left\{\int_0^t c(T(v), \mathbf{n}(v))\,dv\right\}\right], \tag{5.8}$$

where the subscripts on $\mathbb{E}_\mu$ *and* $\mathbb{E}_{(T,\mathbf{n})}$ *denote the starting points of the Markov processes.*

PROOF. $\{(T(t), \mathbf{n}(t)),\ t \ge 0\}$ has generator $\mathscr{L}$ on $B(\mathscr{T}^*)$ defined by

$$\begin{aligned} \mathscr{L}h(T,\mathbf{n}) = {} & \tfrac{1}{2} \sum_{k\colon n_k \ge 2} n_k(n_k - 1)\big(h(T, \mathbf{n} - \mathbf{e}_k) - h(T,\mathbf{n})\big) \\ & + \tfrac{1}{2}\theta \sum_{k\colon n_k = 1,\ x_{k0} \text{ distinct},\ \mathscr{S}\mathbf{x}_k \ne \mathbf{x}_j,\ \forall\, j} \big(h(\mathscr{S}_k T, \mathbf{n}) - h(T,\mathbf{n})\big) \\ & + \tfrac{1}{2}\theta \sum_{\substack{k\colon n_k = 1 \\ x_{k0} \text{ distinct}}} \ \sum_{j\colon \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} \big(h(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n} + \mathbf{e}_j)) - h(T,\mathbf{n})\big), \end{aligned}$$

and clearly

$$\psi_{T,\mathbf{n}}(\mu) = \mathscr{L}[\phi_{\cdot,\cdot}(\mu)](T,\mathbf{n}) + c(T,\mathbf{n})\phi_{T,\mathbf{n}}(\mu) \tag{5.9}$$

for all $\mu \in \mathscr{P}(E)$ and $(T,\mathbf{n}) \in \mathscr{T}^*$, where $\psi_{T,\mathbf{n}}$ is as in Theorem 4.1. The result follows from Dawson and Kurtz (1982). □

COROLLARY 5.3. *For all* $(T,\mathbf{n}) \in \mathscr{T}^*$,

$$\int \phi_{T,\mathbf{n}}\,d\tilde{\mu} = \mathbb{E}_{(T,\mathbf{n})}\left[\exp\left\{\int_0^\infty c(T(v),\mathbf{n}(v))\,dv\right\}\right]. \tag{5.10}$$

PROOF. Starting at $(T,\mathbf{n}) \in \mathscr{T}^*$, the Markov process absorbs at $(\mathscr{T}_1,(1))$ after $s + n - 1$ jumps, s being the number of segregating sites of T. For each $\mu \in \mathscr{P}_a^0(E)$, $\phi_{\mathscr{T}_1,(1)}(\mu) = 1$, so the right-hand side of (5.8) converges to the right-hand side of (5.10) as $t \to \infty$ by the dominated convergence theorem. On the other hand, Lemma 2.2 implies that the left-hand side of (5.8) converges to the left-hand side of (5.10) as $t \to \infty$. [This does not follow directly from the ergodicity proved in Theorem 2.3 because $\phi_{T,\mathbf{n}}$ is not continuous on $\mathscr{P}(E)$.] □

We turn next to a representation of the stationary tree probabilities in terms of a (discrete time) Markov chain. This was motivated in part by Watterson (1985) and Hoppe (1984). We define $\{T_\tau,\ \tau = 0,1,\dots\}$ to be a Markov chain in

$$\bigcup_{n=2}^{\infty} (\mathscr{T}_n/\approx)$$

with starting point $T_0 = \Phi_{(2)}(\mathscr{T}_1)$ and transitions

$$[(\mathbf{x}_1,\dots,\mathbf{x}_n)] \to \begin{cases} [(\mathbf{x}_1,\dots,\mathbf{x}_{m-1},(y,\mathbf{x}_m),\mathbf{x}_{m+1},\dots,\mathbf{x}_n)], \\ \qquad\qquad \text{prob. } \theta/n(n-1+\theta),\ m = 1,\dots,n, \\ [(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n,\mathbf{x}_m)], \\ \qquad\qquad \text{prob. } (n-1)/n(n-1+\theta),\ m = 1,\dots,n, \end{cases}$$

where $[(\mathbf{x}_1,\dots,\mathbf{x}_n)]$ denotes the equivalence class in $\mathscr{T}_n/\approx$ containing $(\mathbf{x}_1,\dots,\mathbf{x}_n)$ and $y \in [0,1]$ is distinct from each x_{ij}. We note that the Markov chain in effect "grows the tree" T_τ in τ steps starting from T_0.

THEOREM 5.4. *For each* $(T,\mathbf{n}) \in \mathscr{T}^*$ *except* $(\mathscr{T}_1,(1))$,

$$\int \hat{\phi}_{T,\mathbf{n}}\,d\tilde{\mu} = \mathbb{P}\{T_{s+n-2} = [\Phi_{\mathbf{n}}(T)],\ T_{s+n-1} \in \mathscr{T}_{n+1}/\approx\}, \tag{5.11}$$

where s is the number of segregating sites of T.

PROOF. For $(T,\mathbf{n}) \in \mathscr{T}^*$ with $s + n \ge 2$, let $q(T,\mathbf{n})$ denote the right-hand side of (5.11). Clearly,

$$q(T,\mathbf{n}) = \mathbb{P}\{T_{s+n-2} = [\Phi_{\mathbf{n}}(T)]\}\frac{n-1}{n-1+\theta} \tag{5.12}$$

for all $(T,\mathbf{n}) \in \mathscr{T}^*$ with $s+n \geq 2$, and one can check that

$$\begin{aligned}
&\mathbb{P}\{T_{s+n-2} = [\Phi_{\mathbf{n}}(T)]\}\\
&\quad= \sum_{k:\, n_k \geq 2} \mathbb{P}\{T_{s+n-3} = [\Phi_{\mathbf{n}-\mathbf{e}_k}(T)]\} \frac{(n_k-1)(n-2)}{(n-1)(n-2+\theta)} \frac{a(T,\mathbf{n}-\mathbf{e}_k)}{a(T,\mathbf{n})}\\
&\qquad+ \sum_{\substack{k:\, n_k=1\\ x_{k0}\text{ distinct}\\ \mathscr{S}\mathbf{x}_k \neq \mathbf{x}_j,\, \forall j}} \mathbb{P}\{T_{s+n-3} = [\Phi_{\mathbf{n}}(\mathscr{S}_k T)]\} \frac{\theta}{n(n-1+\theta)} \frac{a(\mathscr{S}_k T,\mathbf{n})}{a(T,\mathbf{n})}\\
&\qquad+ \sum_{\substack{k:\, n_k=1\\ x_{k0}\text{ distinct}}} \sum_{j:\, \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} \mathbb{P}\Big\{T_{s+n-3} = \Big[\Phi_{\mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)}(\mathscr{R}_k T)\Big]\Big\}\\
&\qquad\quad\times \frac{(n_j+1)\theta}{n(n-1+\theta)} \frac{a\big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)\big)}{a(T,\mathbf{n})}
\end{aligned} \tag{5.13}$$

for all $(T,\mathbf{n}) \in \mathscr{T}^*$ with $s+n \geq 3$. Substituting (5.12) in (5.13) gives

$$\begin{aligned}
&n(n-1+\theta)q(T,\mathbf{n})\\
&\quad= n \sum_{k:\, n_k \geq 2} (n_k - 1) \frac{a(T,\mathbf{n}-\mathbf{e}_k)}{a(T,\mathbf{n})} q(T,\mathbf{n}-\mathbf{e}_k)\\
&\qquad+\theta \sum_{\substack{k:\, n_k=1\\ x_{k0}\text{ distinct}\\ \mathscr{S}\mathbf{x}_k \neq \mathbf{x}_j,\, \forall j}} \frac{a(\mathscr{S}_k T,\mathbf{n})}{a(T,\mathbf{n})} q(\mathscr{S}_k T,\mathbf{n})\\
&\qquad+\theta \sum_{\substack{k:\, n_k=1\\ x_{k0}\text{ distinct}}} \sum_{j:\, \mathscr{S}\mathbf{x}_k = \mathbf{x}_j} (n_j+1) \frac{a\big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)\big)}{a(T,\mathbf{n})}\\
&\qquad\quad\times q\big(\mathscr{R}_k T, \mathscr{R}_k(\mathbf{n}+\mathbf{e}_j)\big)
\end{aligned} \tag{5.14}$$

for all $(T,\mathbf{n}) \in \mathscr{T}^*$ except $(\mathscr{T}_1,(1))$ and $(\mathscr{T}_1,(2))$. By (5.12) $q(\mathscr{T}_1,(2)) = 1/(1+\theta)$, so if we define $q(\mathscr{T}_1,(1)) = 1$, Corollary 4.2 and Remark 4.4 imply that

$$q(T,\mathbf{n}) = \int \hat{\phi}_{T,\mathbf{n}}\, d\bar{\mu}$$

for all $(T,\mathbf{n}) \in \mathscr{T}^*$, and hence (5.11) holds for all $(T,\mathbf{n}) \in \mathscr{T}^*$ except $(\mathscr{T}_1,(1))$. □

REMARK 5.5. It would be possible to obtain a numerical estimate of the left-hand side of (5.8) by simulating the Markov process $\{(T(t),\mathbf{n}(t)),\ t \geq 0\}$ a number of times, evaluating the integrand on the right-hand side of (5.8) for each realization, and taking the average over the simulations as an estimate of the $\mathbb{E}_{(T,\mathbf{n})}$ expectation.

Similarly, $\int \phi_{T,\mathbf{n}}\, d\bar{\mu}$ could be estimated using (5.10).

Noting that (5.11) can be written as

$$\int \hat{\phi}_{T,\mathbf{n}}\, d\bar{\mu} = \mathbb{P}\{T_{\tau(n+1)-1} = [\Phi_{\mathbf{n}}(T)]\}, \tag{5.15}$$

where $\tau(n+1)$ is the hitting time of the set $\mathscr{T}_{n+1}/\approx$, we see that the $\int \mu^n(\cdot)\tilde{\mu}(d\mu)$-distribution of unlabelled, unordered trees in $\mathscr{T}_n/\approx$ can be approximated by simulating $\{T_\tau,\ \tau = 0,1,\dots\}$ up to time $\tau(n+1)$ and observing $T_{\tau(n+1)-1}$.

The final representation of the stationary tree probabilities that we consider is in terms of Kingman's (1982) coalescent process. Fix $n \in \mathbb{N}$ and let $\mathscr{E}_n$ denote the set of partitions of $\{1,2,\dots,n\}$. For $\mathscr{A}, \mathscr{B} \in \mathscr{E}_n$ write $\mathscr{B} < \mathscr{A}$ if $\mathscr{B}$ is obtained from $\mathscr{A}$ by replacing two sets in $\mathscr{A}$ by their union, and note that $\#(\mathscr{B}) = \#(\mathscr{A}) - 1$ if $\mathscr{B} < \mathscr{A}$. The n-coalescent is a pure jump Markov process $\{\mathscr{A}_t,\ t \geq 0\}$ in $\mathscr{E}_n$ with starting point $\{\{1\},\{2\},\dots,\{n\}\}$, each of the $\frac{1}{2}\#(\mathscr{A})[\#(\mathscr{A}) - 1]$ transitions $\mathscr{A} \to \mathscr{B}$ $(\mathscr{B} < \mathscr{A})$ having rate 1, and absorbing point $\{\{1,2,\dots,n\}\}$. Let $\mathscr{B}_n,\dots,\mathscr{B}_2,\mathscr{B}_1$ be the sequence of partitions through which the process passes, labelled in reverse order so that $\#(\mathscr{B}_m) = m$ for $m = n,\dots,2,1$ and denote the sets in $\mathscr{B}_m$ by $B_{m1},\dots,B_{mm}$ for $m = n,\dots,2,1$. Let $\{Y_{mjk}\colon m = n,\dots,2,1;\ j = 1,\dots,m;\ k = 1,2,\dots\}$ be independent uniform $[0,1]$ random variables, and let $(K_{m1},\dots,K_{mm})$, $m = n,\dots,2$, be independent random vectors with

$$\mathbb{P}\{(K_{m1},\dots,K_{mm}) = (k_1,\dots,k_m)\} = \frac{k!}{k_1!\cdots k_m!}\left(\frac{\theta}{m}\right)^k \frac{m-1}{(m-1+\theta)^{k+1}}, \tag{5.16}$$

where $(k_1,\dots,k_m) \in \mathbb{Z}_+^m$ and $k = k_1 + \cdots + k_m$. Assume further that the sequences $\{\mathscr{B}_m\}$, $\{Y_{mjk}\}$ and $\{(K_{m1},\dots,K_{mm})\}$ are independent.

Informally, given a random sample of n genes, $K_{m1},\dots,K_{mm}$ are the numbers of mutations occurring to ancestors of the sample while there are m ancestors $(m = n,\dots,2)$. The length of time t_m^* during which there are m ancestors is exponentially distributed with parameter $\frac{1}{2}m(m-1)$, and mutations occur independently in each ancestral line according to a Poisson process with rate $\theta/2$. Conditioning on t_m^*, we should have

$$\mathbb{P}\{(K_{m1},\dots,K_{mm}) = (k_1,\dots,k_m)\} = \mathbb{E}\left[\prod_{j=1}^{m}\left\{\frac{(\theta t_m^*/2)^{k_j}}{k_j!}e^{-\theta t_m^*/2}\right\}\right].$$

This motivates the definition (5.16) as well as the following construction.

Let

$$\mathbf{Z}_{mj} = \left(Y_{mj1},\dots,Y_{mjK_{mj}}\right), \qquad m = n,\dots,2;\quad j = 1,\dots,m,$$
$$\mathbf{Z}_{11} = (Y_{111},Y_{112},\dots),$$

where the vector $\mathbf{Z}_{mj}$ is empty if $K_{mj} = 0$. For $m = n,\dots,2,1$ and $i = 1,\dots,n$ define $j(m,i) = l$ if $i \in B_{ml}$, and let

$$\mathbf{X}_i = \left(\mathbf{Z}_{n,j(n,i)},\mathbf{Z}_{n-1,j(n-1,i)},\dots,\mathbf{Z}_{1,j(1,i)}\right) \in E.$$

By properties of the n-coalescent, $(\mathbf{X}_1,\dots,\mathbf{X}_n) \in \mathscr{T}_n$ with probability 1, so we define U to be the (random) equivalence class in $\mathscr{T}_n/\sim$ containing $(\mathbf{X}_1,\dots,\mathbf{X}_n)$.

Theorem 5.6. *For each* $(T,\mathbf{n}) \in \mathscr{T}^*$,

$$\int \phi_{T,\mathbf{n}}\,d\tilde{\mu} = \mathbb{P}\{U = \Phi_{\mathbf{n}}(T)\}. \tag{5.17}$$

Proof. Let $W_{T,\mathbf{n}} = \{\omega\colon U(\omega) = \Phi_{\mathbf{n}}(T)\}$ and denote $(K_{n1},\dots,K_{nn})$ by $\mathbf{K}_n$. Then

$$\begin{aligned}\mathbb{P}(W_{T,\mathbf{n}}) &= \mathbb{P}\{\mathbf{K}_n = \mathbf{0}\}\mathbb{P}(W_{T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{0}\}) \\ &\quad + \sum_{\mathbf{k}\in\mathbb{Z}_+^n-\{\mathbf{0}\}} \mathbb{P}\{\mathbf{K}_n = \mathbf{k}\}\mathbb{P}(W_{T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{k}\}),\end{aligned} \tag{5.18}$$

and the first term on the right-hand side is

$$\frac{n-1}{n-1+\theta}\sum_{i:\,n_i\ge 2}\frac{n_i(n_i-1)}{n(n-1)}\mathbb{P}(W_{T,\mathbf{n}-\mathbf{e}_i}).$$

As for the sum, define $\alpha\colon \{1,\dots,n\}\to\{1,\dots,d\}$ by setting $\alpha(j) = i$ if $n_1 + \cdots + n_{i-1} + 1 \le j \le n_1 + \cdots + n_i$. If $i \in \{1,\dots,d\}$ and $n_i = 1$, define $\beta(i) = \alpha^{-1}(i) = n_1 + \cdots + n_i \in \{1,\dots,n\}$. Now

$$\mathbb{P}(W_{T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{k}\}) = \mathbb{P}\bigl(W_{\mathscr{S}_{\alpha(j)}T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{k} - \mathbf{e}_j\}\bigr) \tag{5.19}$$

for all $\mathbf{k} \in \mathbb{Z}_+^n$ and $j \in \{1,\dots,n\}$ with $k_j \ge 1$, provided $n_{\alpha(j)} = 1$ and $x_{\alpha(j),0}$ is distinct; if either of the latter two conditions is not satisfied, the left-hand side of (5.19) is zero. Also,

$$\mathbb{P}\{\mathbf{K}_n = \mathbf{k}\} = \frac{k_1 + \cdots + k_n}{k_j}\,\frac{\theta}{n(n-1+\theta)}\mathbb{P}\{\mathbf{K}_n = \mathbf{k} - \mathbf{e}_j\}$$

for all $\mathbf{k} \in \mathbb{Z}_+^n$ and $j \in \{1,\dots,n\}$ with $k_j \ge 1$. Consequently, the sum on the right-hand side of (5.18) becomes

$$\begin{aligned}&\sum_{\mathbf{k}\in\mathbb{Z}_+^n-\{\mathbf{0}\}}\ \sum_{j:\,k_j\ge 1}\frac{k_j}{k_1+\cdots+k_n}\mathbb{P}\{\mathbf{K}_n = \mathbf{k}\}\mathbb{P}(W_{T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{k}\}) \\ &= \frac{\theta}{n(n-1+\theta)}\sum_{\mathbf{k}\in\mathbb{Z}_+^n-\{\mathbf{0}\}}\ \sum_{\substack{j:\,k_j\ge 1\\ n_{\alpha(j)}=1\\ x_{\alpha(j),0}\text{ distinct}}}\mathbb{P}\{\mathbf{K}_n = \mathbf{k} - \mathbf{e}_j\} \\ &\qquad \times \mathbb{P}\bigl(W_{\mathscr{S}_{\alpha(j)}T,\mathbf{n}}|\{\mathbf{K}_n = \mathbf{k} - \mathbf{e}_j\}\bigr) \\ &= \frac{\theta}{n(n-1+\theta)}\sum_{\substack{i:\,n_i=1\\ x_{i0}\text{ distinct}}}\ \sum_{\mathbf{k}\in\mathbb{Z}_+^n:\,k_{\beta(i)}\ge 1}\mathbb{P}\bigl(W_{\mathscr{S}_iT,\mathbf{n}}\cap\{\mathbf{K}_n = \mathbf{k} - \mathbf{e}_{\beta(i)}\}\bigr) \\ &= \frac{\theta}{n(n-1+\theta)}\sum_{i:\,n_i=1,\ x_{i0}\text{ distinct}}\mathbb{P}(W_{\mathscr{S}_iT,\mathbf{n}}).\end{aligned}$$

Thus, (5.17) follows from Corollary 4.2. □

REFERENCES

DAWSON, D. A. and KURTZ, T. G. (1982). Applications of duality to measure-valued diffusion processes. *Advances in Filtering and Optimal Stochastic Control. Lecture Notes in Control and Information Sciences* **42** 91–105. Springer, Berlin.

ETHIER, S. N. and KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.* **13** 429–452.

ETHIER, S. N. and KURTZ, T. G. (1986a). *Markov Processes: Characterization and Convergence.* Wiley, New York.

ETHIER, S. N. and KURTZ, T. G. (1986b). The infinitely-many-alleles model with selection as a measure-valued diffusion. To appear in Proceedings of U.S.–Japan Seminar on Stochastic Methods in Biology.

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* **3** 87–112.

EWENS, W. J. (1974). A note on the sampling theory for infinite alleles and infinite sites models. *Theoret. Population Biol.* **6** 143–148.

FLEMING, W. H. and VIOT, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.* **28** 817–843.

GRIFFITHS, R. C. (1979). Exact sampling distributions from the infinite neutral alleles model. *Adv. in Appl. Probab.* **11** 326–354.

GRIFFITHS, R. C. (1981). Transient distribution of the number of segregating sites in a neutral infinite-sites model with no recombination. *J. Appl. Probab.* **18** 42–51.

GRIFFITHS, R. C. (1982). The number of alleles and segregating sites in a sample from the infinite-alleles model. *Adv. in Appl. Probab.* **14** 225–239.

HOPPE, F. M. (1984). Pólya-like urns and the Ewens sampling formula. *J. Math. Biol.* **20** 91–94.

KIMURA, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61** 893–903.

KIMURA, M. (1971). Theoretical foundations of population genetics at the molecular level. *Theoret. Population. Biol.* **2** 174–208.

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248.

KURTZ, T. G. (1981). *Approximation of Population Processes.* SIAM, Philadelphia.

LI, W.-H. (1977). Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85** 331–337.

NOTOHARA, M. (1984). A model for the evolution of cistrons in a finite population. Unpublished manuscript.

STROBECK, C. (1983). Estimation of the neutral mutation rate in a finite population from DNA sequence data. *Theoret. Population Biol.* **24** 160–172.

WATTERSON, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoret. Population Biol.* **7** 256–276.

WATTERSON, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13** 639–651.

WATTERSON, G. A. (1985). Estimating species divergence times using multilocus data. In *Population Genetics and Molecular Evolution* (T. Ohta and K. Aoki, eds.) 163–183. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112

DEPARTMENT OF MATHEMATICS
MONASH UNIVERSITY
CLAYTON, VICTORIA 3168
AUSTRALIA