The Annals of Applied Probability
2000, Vol. 10, No. 1, 123–162

A TRANSITION FUNCTION EXPANSION FOR A DIFFUSION MODEL WITH SELECTION

BY A. D. BARBOUR,[1] S. N. ETHIER AND R. C. GRIFFITHS

University of Zürich, University of Utah and Monash University

Using duality, an expansion is found for the transition function of the reversible K-allele diffusion model in population genetics. In the neutral case, the expansion is explicit but already known. When selection is present, it depends on the distribution at time t of a specified K-type birth-and-death process starting at "infinity." The latter process is constructed by means of a coupling argument and characterized as the Ray process corresponding to the Ray–Knight compactification of the K-dimensional nonnegative-integer lattice.

1. Introduction. Consider the K-allele diffusion model in population genetics, where $2 \le K < \infty$. It assumes values in the $(K-1)$-dimensional simplex

$$\Delta_K := \{x = (x_1, \ldots, x_K)\colon x_1 \ge 0, \ldots, x_K \ge 0,\ x_1 + \cdots + x_K = 1\} \tag{1.1}$$

and is characterized in terms of the generator

$$\begin{aligned} L := {}& \frac{1}{2} \sum_{i,\, j=1}^{K} x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i\, \partial x_j} \\ & + \sum_{i=1}^{K} \left[\sum_{j=1}^{K} \gamma_{ji} x_j + x_i \left(\sum_{j=1}^{K} \sigma_{ij} x_j - \sum_{k,\, l=1}^{K} \sigma_{kl} x_k x_l \right) \right] \frac{\partial}{\partial x_i}, \end{aligned} \tag{1.2}$$

where the infinitesimal matrix (γ_{ij}) describes the mutation structure and the real symmetric matrix (σ_{ij}) describes the diploid selection pattern. (By "infinitesimal matrix" we mean a square matrix with nonnegative off-diagonal entries and row sums equal to zero.) The domain of L is $\mathscr{D}(L) := \{f|_{\Delta_K}\colon f \in C^2(\mathbf{R}^K)\}$ and $(\partial/\partial x_i)(f|_{\Delta_K}) := (\partial f/\partial x_i)|_{\Delta_K}$. This formulation of the generator is due to Sato (1978).

Wright (1949) showed (essentially) that when

$$\gamma_{ij} = \tfrac{1}{2}\theta_j > 0, \qquad i, j \in \{1, \ldots, K\}, \qquad i \ne j, \tag{1.3}$$

(and hence $\gamma_{ii} = -(1/2)\sum_{j\colon j \ne i} \theta_j$) the diffusion has a unique stationary distribution Π in $\mathscr{P}(\Delta_K)$, the set of Borel probability measures on Δ_K, given by

$$\Pi(dx) = C x_1^{\theta_1 - 1} \cdots x_K^{\theta_K - 1} \exp\left\{ \sum_{i,\, j=1}^{K} \sigma_{ij} x_i x_j \right\} dx_1 \cdots dx_{K-1}. \tag{1.4}$$

Received May 1998; revised July 1999.
[1]Supported in part by Swiss Nationalfonds Grant 20-50686.97.
AMS 1991 *subject classifications*. Primary 60J35, 60J60, 60J27; secondary 92D10.
Key words and phrases. Finite-dimensional diffusion process, population genetics, duality, reversibility, multitype birth-and-death process, coupling, Ray–Knight compactification.

In the neutral case ($\sigma_{ij} = 0$ for $i, j = 1, \dots, K$), $C = \Gamma(\theta_1 + \cdots + \theta_K)/(\Gamma(\theta_1)\cdots\Gamma(\theta_K))$ and Π is the Dirichlet distribution with parameters $\theta_1, \dots, \theta_K$, which we denote hereafter by $\text{Dir}[\theta]$, where $\theta = (\theta_1, \dots, \theta_K)$.

In the neutral case more generally, it is known that the diffusion has a unique stationary distribution $\Pi \in \mathscr{P}(\Delta_K)$ assuming only that the infinitesimal matrix (γ_{ij}) is irreducible [Shiga (1981)], and in the latter case that the diffusion is reversible with respect to Π if and only if (1.3) holds [Overbeck and Röckner (1997) and Li, Shiga and Yao (1999)].

Explicit eigenfunction expansions for the transition density were found in the neutral case assuming (1.3) independently by Shimakura (1977) and Griffiths (1979). Shimakura's expansion was in terms of a pair of biorthogonal systems of eigenfunctions (the Appell polynomials), whereas Griffiths' was in terms of a single orthonormal system of eigenfunctions (not explicitly identified). Here orthogonality is in $L^2(\text{Dir}[\theta])$.

Griffiths and Li (1983) and Tavaré (1984) later found a simpler expansion for this transition function, which can be described as follows. Let $\{N(t), t \geq 0\}$ be the pure death process in $\mathbf{Z}_+ \cup \{\infty\}$ starting at the entrance boundary ∞ with death rates

$$q^\circ(n, n-1) := \tfrac{1}{2} n(n - 1 + |\theta|), \tag{1.5}$$

where $|\theta| := \theta_1 + \cdots + \theta_K$, and define

$$d_n^\circ(t) = \mathbf{P}[N(t) = n], \qquad n \geq 0,\ t > 0. \tag{1.6}$$

A complicated but explicit formula for $d_n^\circ(t)$ is known; see, for example, Tavaré [(1984), equation (5.5)]. In what follows, we use the vector notation

$$|\alpha| = \sum_{i=1}^K \alpha_i, \qquad \binom{|\alpha|}{\alpha} = \frac{|\alpha|!}{\alpha_1! \cdots \alpha_K!}, \qquad x^\alpha = \prod_{i=1}^K x_i^{\alpha_i}, \tag{1.7}$$

for $\alpha \in \mathbf{Z}_+^K$ and $x \in \Delta_K$, where $0^0 := 1$.

THEOREM 1.1. *If $2 \leq K < \infty$ and $\theta_1 > 0, \dots, \theta_K > 0$, then the neutral diffusion model in Δ_K with generator L as above, in which the infinitesimal matrix (γ_{ij}) satisfies (1.3) and $\sigma_{ij} = 0$ for $i, j = 1, \dots, K$, has transition function $P(t, x, dy)$ given for each $t > 0$ and $x \in \Delta_K$ by*

$$P(t, x, \cdot) = \sum_{n=0}^\infty d_n^\circ(t) \sum_{\alpha \in \mathbf{Z}_+^K:\ |\alpha| = n} \binom{n}{\alpha} x^\alpha \,\text{Dir}[\alpha + \theta](\cdot). \tag{1.8}$$

This expansion has several important consequences; in particular,

$$d_{TV}\big(P(t, x, \cdot), \text{Dir}[\theta](\cdot)\big) \leq 1 - d_0^\circ(t), \qquad t > 0,\ x \in \Delta_K, \tag{1.9}$$

where $d_{TV}(\mu, \nu) := \sup_{A \in \mathscr{B}(E)} |\mu(A) - \nu(A)|$ denotes the total variation distance between the Borel probability measures μ and ν on E. In fact, since $1 - d_0^\circ(t) \leq (1 + |\theta|)e^{-|\theta|t/2}$ for every $t > 0$ [see Tavaré (1984), equation (5.9)], we have an explicit estimate on the rate of convergence to equilibrium. It is thus of interest to extend the expansion to models of greater generality, especially those incorporating selection.

Before turning to such models, we point out an alternative form for (1.8). Let $\{\alpha(t),\ t \geq 0\}$ be the K-type pure death process starting at "infinity" with death rates

$$q(\alpha, \alpha - \varepsilon^i) := \tfrac{1}{2}\alpha_i(|\alpha| - 1 + |\theta|), \qquad \alpha \in \mathbf{Z}_+^K,\ 1 \leq i \leq K, \tag{1.10}$$

where $(\varepsilon^i)_j := \delta_{ij}$. Given $x \in \Delta_K$, the process is uniquely determined if we require that

$$|\alpha(t)| \to \infty \quad \text{and} \quad \alpha(t)/|\alpha(t)| \to x \qquad \text{as } t \to 0. \tag{1.11}$$

To see this, define the one-to-one map ρ: $\mathbf{Z}_+^K \mapsto \{0\} \cup (\mathbf{N} \times \Delta_K)$ by

$$\rho(0) = 0, \qquad \rho(\alpha) = (|\alpha|, \alpha/|\alpha|) \qquad \text{if } \alpha \neq 0 \tag{1.12}$$

and regard $\rho(\alpha)$ as α in radial form. Identifying $\mathbf{Z}_+^K$ with $\rho(\mathbf{Z}_+^K)$, we compactify the former by letting F be the closure of the latter in the compact space $\{0\} \cup \big((\mathbf{N} \cup \{\infty\}) \times \Delta_K\big)$; that is,

$$F = \{\rho(\alpha) \colon \alpha \in \mathbf{Z}_+^K\} \cup (\{\infty\} \times \Delta_K). \tag{1.13}$$

Since the transition probabilities of the K-type pure death process are given by

$$P_{\alpha\beta}(t) = P^\circ_{|\alpha|\,|\beta|}(t)\frac{\binom{\alpha_1}{\beta_1}\cdots\binom{\alpha_K}{\beta_K}}{\binom{|\alpha|}{|\beta|}}, \qquad \alpha \geq \beta, \tag{1.14}$$

$P^\circ_{mn}(t)$ being the transition probabilities of the one-dimensional pure death process with death rates (1.5), it is a simple matter to show that the formula

$$T(t)f(\rho(\alpha)) := \sum_{\beta \in \mathbf{Z}_+^K} f(\rho(\beta))P_{\alpha\beta}(t), \tag{1.15}$$

extended to F by continuity, defines a Feller semigroup $\{T(t)\}$ on $C(F)$. We can therefore define

$$d_\beta(t, x) = \lim_{\rho(\alpha) \to (\infty, x)} P_{\alpha\beta}(t) \tag{1.16}$$

and note by (1.14) that

$$d_\beta(t, x) = d^\circ_{|\beta|}(t)\binom{|\beta|}{\beta}x^\beta. \tag{1.17}$$

Consequently, we can rewrite the transition function (1.8) in the compact form

$$P(t, x, \cdot) = \sum_{\alpha \in \mathbf{Z}_+^K} d_\alpha(t, x)\,\mathrm{Dir}[\alpha + \theta](\cdot). \tag{1.18}$$

Let us now include haploid selection in the model [and retain (1.3)]. This amounts to assuming that there exists $\sigma = (\sigma_1, \dots, \sigma_K) \in \mathbf{R}^K$ such that

$$\sigma_{ij} = \sigma_i + \sigma_j, \qquad i, j = 1, \dots, K. \tag{1.19}$$

By (1.4), the unique stationary distribution in this case, which we denote by $\Pi[\theta]$, is absolutely continuous with respect to $\mathrm{Dir}[\theta]$ with Radon–Nikodym derivative equal to

$$(d\Pi[\theta]/d\,\mathrm{Dir}[\theta])(x) = c(\theta)^{-1}\, e^{2\sigma\cdot x}, \tag{1.20}$$

where $c(\theta)$ is a normalizing constant depending implicitly on σ, namely,

$$c(\theta) := \int_{\Delta_K} e^{2\sigma\cdot y}\,\mathrm{Dir}[\theta](dy) \in [\exp(2\min\sigma_i), \exp(2\max\sigma_i)]. \tag{1.21}$$

By analogy with (1.18), we might expect a transition function expansion of the form

$$P(t, x, \cdot) = \sum_{\alpha\in\mathbf{Z}_+^K} b_\alpha(t, x)\,\Pi[\alpha+\theta](\cdot), \tag{1.22}$$

where the coefficients $b_\alpha(t, x)$ remain to be determined.

In Section 3, we show that $b_{\bullet}(t, x)$ is the distribution at time t of a K-type birth-and-death process $\{\alpha(t),\ t \geq 0\}$ starting at "infinity" with death and birth rates

$$q(\alpha, \alpha - \varepsilon^i) := \frac{1}{2}\alpha_i(|\alpha| - 1 + |\theta|)\frac{c(\alpha - \varepsilon^i + \theta)}{c(\alpha+\theta)} \tag{1.23}$$

and

$$q(\alpha, \alpha + \varepsilon^i) := \sigma_i^- |\alpha| \frac{\alpha_i + \theta_i}{|\alpha| + |\theta|}\frac{c(\alpha + \varepsilon^i + \theta)}{c(\alpha+\theta)}, \tag{1.24}$$

where

$$\sigma_i^- := (\max\sigma_j) - \sigma_i. \tag{1.25}$$

The dependence on x enters by requiring that (1.11) should hold, for which it is necessary to show that a process with this initial behavior and these transition rates can be constructed. It appears that semigroup theory does not easily apply (except in the neutral case). Instead, we construct the process directly for each $x \in \Delta_K$ by means of a coupling argument, given in Sections 4 and 5, and then show in Section 7 that we have found the Ray process corresponding to the Ray–Knight compactification of $\mathbf{Z}_+^K$.

In the neutral model, the probabilities $d_n^\circ(t)$ determine the distribution of the number of nonmutant ancestors of the current infinite population at time t in the past [Griffiths (1980), Griffiths and Li (1983), Tavaré (1984), Donnelly and Kurtz (1996)] and are related to the coalescent process [Kingman (1982)]. Krone and Neuhauser (1997) have obtained an analogue of the coalescent in a large-population limit from a Moran model with selection, the same limit that leads to a diffusion process with generator L, while Donnelly and Kurtz (1999) have constructed an ancestral graph process for a Fleming–Viot model with selection, the latter being a generalization of the diffusion process with generator L; the relationship between these ancestral processes and the K-type birth-and-death process is as yet unclear.

A proof of Theorem 1.1 can be found in Ethier and Griffiths (1993). That proof uses the fact that the probabilities $d_n^\circ(t)$ satisfy the Kolmogorov forward equations for the one-dimensional pure death process with death rates (1.5). Here we take a different approach, basing the derivation on duality and reversibility. Duality was used in a similar context by Tavaré (1984), but our dual process is somewhat different. In the next section, we formulate the transition function expansion for a class of diffusions in Δ_K with polynomial coefficients. Section 3 then specializes these results to the K-allele diffusion model with generator L as in (1.2) and (1.3); in particular, diploid (not just haploid) selection is permitted.

2. Duality method. Let L [not necessarily as in (1.2)] with domain $\mathscr{D}(L) := C^2(\Delta_K)$ and range in $B(\Delta_K)$, the space of bounded Borel functions on Δ_K, be the generator for a Markov process in Δ_K with Feller transition function $P(t, x, dy)$ and stationary distribution Π. Assume that Π charges nonempty open subsets of Δ_K. Define $f_\alpha \in \mathscr{D}(L)$ for each $\alpha \in \mathbf{Z}_+^K$ by $f_\alpha(x) = x^\alpha$, and assume that

$$Lf_\alpha = \sum_{\beta\in\mathbf{Z}_+^K} r(\alpha, \beta) f_\beta, \qquad \alpha \in \mathbf{Z}_+^K, \tag{2.1}$$

where $r(\alpha, \beta) \geq 0$ for all $\alpha \neq \beta$ and $r(\alpha, \alpha) \leq 0$ for all α. Define

$$m(\alpha) = \int_{\Delta_K} f_\alpha \, d\Pi \tag{2.2}$$

and note that $m(\alpha) > 0$ for all $\alpha \in \mathbf{Z}_+^K$. Moreover,

$$0 = \int_{\Delta_K} Lf_\alpha \, d\Pi = \sum_{\beta\in\mathbf{Z}_+^K} r(\alpha, \beta) m(\beta), \qquad \alpha \in \mathbf{Z}_+^K, \tag{2.3}$$

where the interchange of sum and integral is justified by virtue of the fact that every summand in (2.1) save one is nonnegative. If we define

$$q(\alpha, \beta) = m(\alpha)^{-1} r(\alpha, \beta) m(\beta), \tag{2.4}$$

then $q(\alpha, \beta) \geq 0$ for all $\alpha \neq \beta$, $q(\alpha, \alpha) \leq 0$ for all α, and, by (2.3),

$$\sum_{\beta\in\mathbf{Z}_+^K} q(\alpha, \beta) = 0, \qquad \alpha \in \mathbf{Z}_+^K. \tag{2.5}$$

Further, defining

$$g_\alpha = m(\alpha)^{-1} f_\alpha \in \mathscr{D}(L), \tag{2.6}$$

(2.1) becomes

$$Lg_\alpha = \sum_{\beta\in\mathbf{Z}_+^K} q(\alpha, \beta) g_\beta, \qquad \alpha \in \mathbf{Z}_+^K. \tag{2.7}$$

We assume that $(q(\alpha, \beta))$ is the infinitesimal matrix for a nonexplosive pure jump Markov process $\{\alpha(t),\ t \geq 0\}$ in $\mathbf{Z}_+^K$ with transition probabilities $P_{\alpha\beta}(t)$, and we denote by $\{x(t),\ t \geq 0\}$ the Markov process in Δ_K with generator L. Then, in principle,

$$\mathbf{E}_x\{g_\alpha(x(t))\} = \mathbf{E}_\alpha\{g_{\alpha(t)}(x)\}, \qquad (x,\alpha) \in \Delta_K \times \mathbf{Z}_+^K, \quad t \geq 0, \tag{2.8}$$

where the subscripts on the expectations denote the starting points, or

$$m(\alpha)^{-1} \int_{\Delta_K} f_\alpha(z)\, P(t,x,dz) = \sum_{\beta \in \mathbf{Z}_+^K} P_{\alpha\beta}(t) m(\beta)^{-1} f_\beta(x), \qquad (x,\alpha) \in \Delta_K \times \mathbf{Z}_+^K, \qquad t \geq 0. \tag{2.9}$$

Since we have not assumed that $g_\alpha(x)$ as a function of α belongs to the domain of the generator of the jump process for arbitrary x, (2.8) requires additional justification. By Corollary 4.4.15 of Ethier and Kurtz (1986), it suffices to assume that there exists $H: \mathbf{Z}_+^K \mapsto [0,\infty)$ such that

$$g_\alpha(x) + |(Lg_\alpha)(x)| \leq H(\alpha), \qquad (x,\alpha) \in \Delta_K \times \mathbf{Z}_+^K, \tag{2.10}$$

and

$$\{H(\alpha(t \wedge \tau_N)),\ 0 \leq t \leq t_0,\ N \geq 1\} \text{ is uniformly integrable} \tag{2.11}$$

for each initial state $\alpha \in \mathbf{Z}_+^K$ and each $t_0 \geq 0$, where

$$\tau_N := \inf\{s \geq 0: |\alpha(s)| \geq N\}. \tag{2.12}$$

To express $P(t,x,\cdot)$ in terms of the transition probabilities $P_{\alpha\beta}(t)$ using (2.9) requires a method for recovering a distribution from its moments. For this we observe that the moments can be used to define a sampling distribution, which as the sample size tends to infinity converges weakly to the distribution we seek. More precisely, given $\mu \in \mathscr{P}(\Delta_K)$,

$$\Lambda_n \mu := \sum_{\alpha \in \mathbf{Z}_+^K:\ |\alpha|=n} \binom{n}{\alpha} \int_{\Delta_K} f_\alpha\, d\mu\, \delta_{\alpha/n} \Rightarrow \mu, \tag{2.13}$$

for if $G \subset \Delta_K$ is open, then, by Fatou's lemma and the weak law of large numbers for an i.i.d. sequence of multinomial$(1, y)$ random vectors,

$$\begin{aligned} \liminf_{n\to\infty} \Lambda_n \mu(G) &= \liminf_{n\to\infty} \int_{\Delta_K} \sum_{\alpha \in \mathbf{Z}_+^K:\ |\alpha|=n} \binom{n}{\alpha} f_\alpha(y) \delta_{\alpha/n}(G)\, \mu(dy) \\ &\geq \int_{\Delta_K} \liminf_{n\to\infty} \sum_{\alpha \in \mathbf{Z}_+^K:\ |\alpha|=n} \binom{n}{\alpha} f_\alpha(y) \delta_{\alpha/n}(G)\, \mu(dy) \\ &\geq \int_{\Delta_K} \delta_y(G)\, \mu(dy) \\ &= \mu(G). \end{aligned} \tag{2.14}$$

Incidentally, note that, since $\binom{|\beta|}{\beta} f_\beta \le 1$ for all $\beta \in \mathbf{Z}_+^K$,

$$\begin{aligned} g_\alpha(x) + |(Lg_\alpha)(x)| &\le \sum_{\beta\in\mathbf{Z}_+^K} (\delta_{\alpha\beta} + |q(\alpha,\beta)|) g_\beta(x) \\ &\le \sum_{\beta\in\mathbf{Z}_+^K} (\delta_{\alpha\beta} + |q(\alpha,\beta)|) m(\beta)^{-1} \binom{|\beta|}{\beta}^{-1} \\ &=: H(\alpha) \end{aligned} \tag{2.15}$$

for all $x \in \Delta_K$ and $\alpha \in \mathbf{Z}_+^K$, so we assume that (2.11) holds with H given by (2.15) and τ_N by (2.12).

LEMMA 2.1. *Let $\mu, \mu_0 \in \mathscr{P}(\Delta_K)$ and suppose there exists a Borel function h: $\Delta_K \mapsto [0,\infty)$ such that, with ρ as in* (1.12),

$$\liminf_{\rho(\alpha)\to(\infty,y)} \frac{\int_{\Delta_K} f_\alpha\, d\mu}{\int_{\Delta_K} f_\alpha\, d\mu_0} \ge h(y), \qquad y \in \Delta_K \tag{2.16}$$

and $\int_{\Delta_K} h\, d\mu_0 = 1$. Then $d\mu = h\, d\mu_0$.

PROOF. Define h_n on Δ_K by

$$h_n(x) = \begin{cases} \int_{\Delta_K} f_\alpha\, d\mu \Big/ \int_{\Delta_K} f_\alpha\, d\mu_0, & \text{if } x = \alpha/n, \text{ where } \alpha \in \mathbf{Z}_+^K \\ & \quad \text{and } |\alpha| = n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

Then, for each $g \in C(\Delta_K)$ with $g \ge 0$,

$$\begin{aligned} \int_{\Delta_K} g\, d\mu &= \lim_{n\to\infty} \int_{\Delta_K} g\, d\Lambda_n \mu \\ &= \lim_{n\to\infty} \sum_{\alpha\in\mathbf{Z}_+^K:\ |\alpha|=n} g(\alpha/n) \binom{n}{\alpha} \int_{\Delta_K} f_\alpha\, d\mu \\ &= \lim_{n\to\infty} \sum_{\alpha\in\mathbf{Z}_+^K:\ |\alpha|=n} g(\alpha/n) h_n(\alpha/n) \binom{n}{\alpha} \int_{\Delta_K} f_\alpha\, d\mu_0 \\ &= \lim_{n\to\infty} \int_{\Delta_K} g h_n\, d\Lambda_n \mu_0 \\ &\ge \int_{\Delta_K} gh\, d\mu_0, \end{aligned} \tag{2.18}$$

where the inequality uses (2.13) with μ_0 in place of μ, the Skorokhod representation theorem, Fatou's lemma and (2.16). This implies that $d\mu \ge h\, d\mu_0$, and since h is a probability density with respect to μ_0, equality must hold, thereby proving the lemma. □

To apply the lemma, we will need to assume that

$$b_\beta(t, y) := \liminf_{\rho(\alpha)\to(\infty, y)} P_{\alpha\beta}(t), \qquad \beta \in \mathbf{Z}_+^K, \tag{2.19}$$

defines a probability distribution on $\mathbf{Z}_+^K$ for each $t > 0$ and $y \in \Delta_K$. We will also assume that for each $t > 0$ the probability distribution (2.19) is weakly continuous in $y \in \Delta_K$.

Fix $t > 0$ and $x \in \Delta_K$ and let $\mu = P(t, x, \cdot)$ and $\mu_0 = \Pi$ in the lemma. Using (2.2), (2.9) and Fatou's lemma, we have, for each $y \in \Delta_K$,

$$\begin{aligned} \liminf_{\rho(\alpha)\to(\infty, y)} \frac{\int_{\Delta_K} f_\alpha(z)\, P(t, x, dz)}{\int_{\Delta_K} f_\alpha\, d\Pi} &= \liminf_{\rho(\alpha)\to(\infty, y)} \sum_{\beta\in\mathbf{Z}_+^K} P_{\alpha\beta}(t) m(\beta)^{-1} f_\beta(x) \\ &\geq \sum_{\beta\in\mathbf{Z}_+^K} b_\beta(t, y) m(\beta)^{-1} f_\beta(x) \\ &=: \phi(t, x, y). \end{aligned} \tag{2.20}$$

We must show that $\int_{\Delta_K} \phi(t, x, y)\,\Pi(dy) = 1$.

Recall that t and x are fixed and define h_n on Δ_K by (2.17) with $\mu = P(t, x, \cdot)$ and $\mu_0 = \Pi$, and define ϕ_n on Δ_K by

$$\phi_n(y) = \sum_{\alpha\in\mathbf{Z}_+^K:\ |\alpha|=n} h_n(\alpha/n)\binom{n}{\alpha} f_\alpha(y) = \int_{\Delta_K} h_n\, d\Lambda_n \delta_y. \tag{2.21}$$

Then $\int_{\Delta_K} \phi_n\, d\Pi = 1$ for each $n \geq 1$ and

$$\liminf_{n\to\infty} \phi_n(y) \geq \int_{\Delta_K} \phi(t, x, \cdot)\, d\delta_y = \phi(t, x, y), \qquad y \in \Delta_K, \tag{2.22}$$

using (2.20) and the argument cited for (2.18), so Fatou's lemma implies that

$$1 = \liminf_{n\to\infty} \int_{\Delta_K} \phi_n\, d\Pi \geq \int_{\Delta_K} \liminf_{n\to\infty} \phi_n\, d\Pi \geq \int_{\Delta_K} \phi(t, x, y)\,\Pi(dy), \tag{2.23}$$

and this holds for arbitrary $t > 0$ and $x \in \Delta_K$.

Keep t fixed. Then

$$\begin{aligned} \int_{\Delta_K}\int_{\Delta_K} \phi(t, x, y)\,\Pi(dy)\,\Pi(dx) &= \int_{\Delta_K}\int_{\Delta_K} \phi(t, x, y)\,\Pi(dx)\,\Pi(dy) \\ &= \int_{\Delta_K} \sum_{\beta\in\mathbf{Z}_+^K} b_\beta(t, y)\,\Pi(dy) \\ &= 1 \end{aligned} \tag{2.24}$$

by our assumption that (2.19) defines a probability distribution on $\mathbf{Z}_+^K$. This and (2.23) tell us that $\int_{\Delta_K} \phi(t, x, y)\,\Pi(dy) = 1$ for Π-a.e. $x \in \Delta_K$. For such x, Lemma 2.1 applies, and we conclude that

$$P(t, x, dy) = \phi(t, x, y)\,\Pi(dy), \qquad \Pi\text{-a.e. } x \in \Delta_K. \tag{2.25}$$

Now we assume reversibility with respect to Π, which means that

$$\Pi(dx)\,P(t,x,dy) = \Pi(dy)\,P(t,y,dx). \tag{2.26}$$

This and (2.25) then imply that, for $\Pi \times \Pi$-a.e. $(x, y) \in \Delta_K \times \Delta_K$,

$$\phi(t,x,y) = \phi(t,y,x) = \sum_{\beta\in\mathbf{Z}_+^K} b_\beta(t,x) m(\beta)^{-1} f_\beta(y). \tag{2.27}$$

Defining

$$d\Pi_\beta = m(\beta)^{-1} f_\beta\, d\Pi, \tag{2.28}$$

we conclude from (2.25) and (2.27) that

$$P(t,x,dy) = \sum_{\beta\in\mathbf{Z}_+^K} b_\beta(t,x)\,\Pi_\beta(dy) \tag{2.29}$$

for Π-a.e. $x \in \Delta_K$, hence for all $x \in \Delta_K$ by the Feller property of $P(t, x, dy)$ and by the assumption that the probability distribution (2.19) is weakly continuous in $y \in \Delta_K$. Since t was arbitrary, we have (2.29) for all $t > 0$ and $x \in \Delta_K$.

For the convenience of the reader, we provide a formal statement of the result we have just proved.

THEOREM 2.2. *Let L with domain $\mathscr{D}(L) := C^2(\Delta_K)$ and range in $B(\Delta_K)$ be the generator for a Markov process in Δ_K with Feller transition function $P(t, x, dy)$ and stationary distribution Π with respect to which the process is reversible. Assume that Π charges nonempty open subsets of Δ_K. Define $f_\alpha \in \mathscr{D}(L)$ for each $\alpha \in \mathbf{Z}_+^K$ by $f_\alpha(x) = x^\alpha$ and assume that the matrix $(r(\alpha, \beta))$ satisfies* (2.1), *where $r(\alpha,\beta) \geq 0$ for all $\alpha \neq \beta$ and $r(\alpha,\alpha) \leq 0$ for all α. Define $m(\alpha)$ by* (2.2), *$q(\alpha, \beta)$ by* (2.4) *and Π_β by* (2.28). *Assume that $(q(\alpha,\beta))$ is the infinitesimal matrix for a nonexplosive pure jump Markov process $\{\alpha(t),\ t \geq 0\}$ in $\mathbf{Z}_+^K$ with transition probabilities $P_{\alpha\beta}(t)$ and that the function $H\colon \mathbf{Z}_+^K \mapsto [0,\infty)$ defined by* (2.15) *satisfies* (2.11) *using* (2.12) *for each initial state $\alpha \in \mathbf{Z}_+^K$ and each $t_0 \geq 0$. Assume that* (2.19) *defines a probability distribution $b_{\boldsymbol{\cdot}}(t, y)$ on $\mathbf{Z}_+^K$ for each $t > 0$ and $y \in \Delta_K$, and that, for each $t > 0$, this probability distribution is weakly continuous in $y \in \Delta_K$. Then, for each $t > 0$ and $x \in \Delta_K$,*

$$P(t,x,\cdot) = \sum_{\alpha\in\mathbf{Z}_+^K} b_\alpha(t,x)\,\Pi_\alpha(\cdot). \tag{2.30}$$

The following corollary is immediate; it is useful if the dual process starting at "infinity" absorbs at 0 with probability 1.

COROLLARY 2.3. *Under the hypotheses of Theorem* 2.2,

$$d_{TV}\big(P(t,x,\cdot),\Pi(\cdot)\big) \leq 1 - b_0(t,x), \qquad t > 0,\ x \in \Delta_K. \tag{2.31}$$

3. Application to the *K*-allele model. In this section we apply Theorem 2.2 to obtain a transition function expansion for the K-allele diffusion model discussed in Section 1. The generator L is given by (1.2) and (1.3).

Let us treat the case of haploid selection first, because some simplifications occur in that case. Here (1.19) holds, so

$$\begin{aligned} Lf_\alpha &= \tfrac{1}{2}\sum_{i=1}^K \alpha_i(\alpha_i - 1 + \theta_i) f_{\alpha-\varepsilon^i} - \sum_{i=1}^K \sigma_i |\alpha| f_{\alpha+\varepsilon^i} \\ &\quad - \left\{ \tfrac{1}{2}|\alpha|(|\alpha| - 1 + |\theta|) - \sum_{i=1}^K \sigma_i \alpha_i \right\} f_\alpha. \end{aligned} \tag{3.1}$$

Using the notation (1.25) and the fact that $\sum_{i=1}^K f_{\alpha+\varepsilon^i} = f_\alpha$ for each $\alpha \in \mathbf{Z}_+^K$, this becomes

$$\begin{aligned} Lf_\alpha &= \tfrac{1}{2}\sum_{i=1}^K \alpha_i(\alpha_i - 1 + \theta_i) f_{\alpha-\varepsilon^i} + \sum_{i=1}^K \sigma_i^- |\alpha| f_{\alpha+\varepsilon^i} \\ &\quad - \left\{ \tfrac{1}{2}|\alpha|(|\alpha| - 1 + |\theta|) + \sum_{i=1}^K \sigma_i^- \alpha_i \right\} f_\alpha, \end{aligned} \tag{3.2}$$

and (2.1) holds with

$$r(\alpha, \alpha - \varepsilon^i) = \tfrac{1}{2}\alpha_i(\alpha_i - 1 + \theta_i), \qquad r(\alpha, \alpha + \varepsilon^i) = \sigma_i^- |\alpha|. \tag{3.3}$$

Let

$$\gamma(\alpha, \theta) = \frac{\Gamma(|\theta|)}{\Gamma(\theta_1)\cdots\Gamma(\theta_K)} \frac{\Gamma(\alpha_1 + \theta_1)\cdots\Gamma(\alpha_K + \theta_K)}{\Gamma(|\alpha| + |\theta|)}, \tag{3.4}$$

and observe that

$$d\,\mathrm{Dir}[\alpha + \theta]/d\,\mathrm{Dir}[\theta] = \gamma(\alpha, \theta)^{-1} f_\alpha. \tag{3.5}$$

Consequently, recalling (1.20) and (1.21),

$$\begin{aligned} m(\alpha) = \int_{\Delta_K} f_\alpha \, d\Pi[\theta] &= c(\theta)^{-1} \int_{\Delta_K} f_\alpha(x) e^{2\sigma\cdot x}\, \mathrm{Dir}[\theta](dx) \\ &= \gamma(\alpha, \theta) c(\theta)^{-1} \int_{\Delta_K} e^{2\sigma\cdot x}\, \mathrm{Dir}[\alpha + \theta](dx) \\ &= \gamma(\alpha, \theta) \frac{c(\alpha + \theta)}{c(\theta)}, \end{aligned} \tag{3.6}$$

and we conclude from (2.4) that

$$q(\alpha, \alpha - \varepsilon^i) = \frac{1}{2}\alpha_i(|\alpha| - 1 + |\theta|) \frac{c(\alpha - \varepsilon^i + \theta)}{c(\alpha + \theta)} \tag{3.7}$$

and

$$q(\alpha, \alpha + \varepsilon^i) = \sigma_i^- |\alpha| \frac{\alpha_i + \theta_i}{|\alpha| + |\theta|} \frac{c(\alpha + \varepsilon^i + \theta)}{c(\alpha + \theta)}. \tag{3.8}$$

These are the rates for a K-type birth-and-death process. Note that deaths occur at a quadratic rate, births at a linear rate.

We now turn to the general case (diploid selection), in which the dual process turns out to be more complicated. By (1.2) and (1.3),

$$\begin{aligned} Lf_\alpha = {}& \tfrac{1}{2}\sum_{i=1}^K \alpha_i(\alpha_i - 1 + \theta_i) f_{\alpha-\varepsilon^i} + \sum_{i,\,j=1}^K \sigma_{ij}\alpha_i f_{\alpha+\varepsilon^j} \\ & - \sum_{i,\,j=1}^K \sigma_{ij}|\alpha| f_{\alpha+\varepsilon^i+\varepsilon^j} - \tfrac{1}{2}|\alpha|(|\alpha| - 1 + |\theta|) f_\alpha. \end{aligned} \tag{3.9}$$

A successful reduction to a form in which (2.1) holds with the coefficients $r(\alpha, \beta)$ having the appropriate signs can be achieved by defining

$$\begin{gathered} \sigma_{ij}^+ = \sigma_{ij} - (\min \sigma_{kl}), \qquad \sigma_{ij}^- = (\max \sigma_{kl}) - \sigma_{ij}, \\ \overline{\sigma} = \max \sigma_{kl} - \min \sigma_{kl}, \end{gathered} \tag{3.10}$$

in which case

$$\begin{aligned} Lf_\alpha = {}& \tfrac{1}{2}\sum_{i=1}^K \alpha_i(\alpha_i - 1 + \theta_i) f_{\alpha-\varepsilon^i} + \sum_{j=1}^K \left(\sum_{i=1}^K \sigma_{ij}^+ \alpha_i \right) f_{\alpha+\varepsilon^j} \\ & + \sum_{i,\,j=1}^K \sigma_{ij}^- |\alpha| f_{\alpha+\varepsilon^i+\varepsilon^j} - \tfrac{1}{2}|\alpha| \big\{ |\alpha| - 1 + |\theta| + 2\overline{\sigma} \big\} f_\alpha. \end{aligned} \tag{3.11}$$

Arguing as in the haploid case and letting

$$\begin{aligned} c(\theta) &= \int_{\Delta_K} \exp\left\{ \sum_{i,\,j=1}^K \sigma_{ij} x_i x_j \right\} \mathrm{Dir}[\theta](dx) \\ &\in \big[\exp(\min \sigma_{ij}), \exp(\max \sigma_{ij})\big], \end{aligned} \tag{3.12}$$

we have

$$q(\alpha, \alpha - \varepsilon^i) = \frac{1}{2}\alpha_i(|\alpha| - 1 + |\theta|) \frac{c(\alpha - \varepsilon^i + \theta)}{c(\alpha + \theta)}, \tag{3.13}$$

$$q(\alpha, \alpha + \varepsilon^j) = \sum_{i=1}^K \sigma_{ij}^+ \alpha_i \frac{\alpha_j + \theta_j}{|\alpha| + |\theta|} \frac{c(\alpha + \varepsilon^j + \theta)}{c(\alpha + \theta)} \tag{3.14}$$

and

$$\begin{aligned} q\big(\alpha, \alpha + \varepsilon^i + \varepsilon^j\big) = {}& \sigma_{ij}^- |\alpha| \frac{(\alpha_i + \delta_{ij} + \theta_i)(\alpha_j + \theta_j)}{(|\alpha| + 1 + |\theta|)(|\alpha| + |\theta|)} \\ & \times \frac{c(\alpha + \varepsilon^i + \varepsilon^j + \theta)}{c(\alpha + \theta)}. \end{aligned} \tag{3.15}$$

These are the rates for a K-type birth-and-death process that permits the birth of twins. [Note that they do not reduce to (3.7) and (3.8) when (1.19) holds.] Again, deaths occur at a quadratic rate, births at a linear rate.

We now turn to the uniform integrability condition (2.11). Note first that, by (3.10) and (3.12)–(3.15),

$$\tfrac{1}{2}e^{-\overline{\sigma}}|\alpha|(|\alpha|-1+|\theta|) \le \sum_{i=1}^{K} q(\alpha,\alpha-\varepsilon^i) \le \tfrac{1}{2}e^{\overline{\sigma}}|\alpha|(|\alpha|-1+|\theta|), \tag{3.16}$$

$$0 \le \sum_{i=1}^{K} q(\alpha,\alpha+\varepsilon^i) \le \overline{\sigma}e^{\overline{\sigma}}|\alpha| \tag{3.17}$$

and

$$0 \le \sum_{i,\,j=1}^{K} q(\alpha,\alpha+\varepsilon^i+\varepsilon^j) \le \overline{\sigma}e^{\overline{\sigma}}|\alpha|. \tag{3.18}$$

[This is, of course, still true in the haploid case (1.19), where $\overline{\sigma} = 2(\max\sigma_i - \min\sigma_i)$, but in that case the bound on the right side of (3.17) can be halved, using (3.8).] The generator $L^{\#}$ of the K-type birth-and-death process has the form

$$\begin{aligned}(L^{\#}\varphi)(\alpha) = {} & \sum_{i=1}^{K} q(\alpha,\alpha-\varepsilon^i)[\varphi(\alpha-\varepsilon^i)-\varphi(\alpha)] \\ & + \sum_{i=1}^{K} q(\alpha,\alpha+\varepsilon^i)[\varphi(\alpha+\varepsilon^i)-\varphi(\alpha)] \\ & + \sum_{i,\,j=1}^{K} q(\alpha,\alpha+\varepsilon^i+\varepsilon^j)[\varphi(\alpha+\varepsilon^i+\varepsilon^j)-\varphi(\alpha)],\end{aligned} \tag{3.19}$$

so if $\varphi_l(\alpha) := (|\alpha|+1)^l$ for $l \in \mathbf{N}$, then

$$\begin{aligned}(L^{\#}\varphi_l)(\alpha) & \le \tfrac{1}{2}e^{-\overline{\sigma}}n(n-1+|\theta|)[n^l-(n+1)^l] \\ & \quad + \overline{\sigma}e^{\overline{\sigma}}n\big[(n+2)^l-(n+1)^l+(n+3)^l-(n+1)^l\big] \\ & \le C_l, \qquad \alpha \in \mathbf{Z}_+^K,\end{aligned} \tag{3.20}$$

where $n = |\alpha|$ and the existence of C_l is due to the fact that the penultimate expression is negative for n sufficiently large. We conclude that

$$\begin{aligned}\mathbf{E}_\alpha\{(|\alpha(t\wedge\tau_N)|+1)^l\} & = (|\alpha|+1)^l + \mathbf{E}_\alpha\Big\{\int_0^{t\wedge\tau_N}(L^{\#}\varphi_l)(\alpha(s))\,ds\Big\} \\ & \le (|\alpha|+1)^l + C_l t.\end{aligned} \tag{3.21}$$

Observe next, using (2.6), (3.6) and (3.4), that

$$
\begin{aligned}
g_\alpha(x) &= m(\alpha)^{-1} f_\alpha(x) \\
&\le m(\alpha)^{-1} \binom{|\alpha|}{\alpha}^{-1} \\
&= \gamma(\alpha, \theta)^{-1} \frac{c(\theta)}{c(\alpha+\theta)} \binom{|\alpha|}{\alpha}^{-1} \\
&\le \frac{e^{\bar{\sigma}}\Gamma(\theta_1)\cdots\Gamma(\theta_K)}{\Gamma(|\theta|)} \frac{\Gamma(|\alpha|+|\theta|)}{\Gamma(\alpha_1+\theta_1)\cdots\Gamma(\alpha_K+\theta_K)} \\
&\quad \times \frac{\Gamma(\alpha_1+1)\cdots\Gamma(\alpha_K+1)}{\Gamma(|\alpha|+1)} \\
&\le B_0(|\alpha|+1)^{l_0}, \qquad (x, \alpha) \in \Delta_K \times \mathbf{Z}_+^K,
\end{aligned}
\tag{3.22}
$$

for constants $B_0 \ge 0$ and $l_0 \in \mathbf{N}$ sufficiently large. It follows that H, defined by (2.15), satisfies

$$H(\alpha) \le B(|\alpha|+1)^{l_0+2}, \qquad \alpha \in \mathbf{Z}_+^K, \tag{3.23}$$

for a constant B sufficiently large. We therefore obtain the required uniform integrability from that fact that (3.21) holds for $l = l_0 + 3$.

We postpone to Sections 4–6 the verification of the hypothesis that, for each $t > 0$, $b_\bullet(t, y)$ is a probability distribution on $\mathbf{Z}_+^K$ for every $y \in \Delta_K$ that is weakly continuous in y. Modulo this step, we have verified the hypotheses of Theorem 2.2. We give below a precise statement of the resulting transition function expansion, but first we weaken our hypotheses a bit.

Just as Theorem 1.1 can be generalized to permit

$$\gamma_{ij} = \tfrac{1}{2}\theta_j \ge 0, \qquad i, j \in \{1, \dots, K\}, \qquad i \ne j, \tag{3.24}$$

in place of (1.3) [see Ethier and Griffiths (1993)], the transition function expansion for the K-allele diffusion model with selection permits a similar generalization. Let $\theta = (\theta_1, \dots, \theta_K)$, where $\theta_1 \ge 0, \dots, \theta_K \ge 0$, and let (σ_{ij}) be a real symmetric $K \times K$ matrix. Then the diffusion process with generator L as in (1.2) and (3.24) has a unique stationary distribution, provided only that $|\theta| > 0$. In the neutral case it is given by

$$\mathrm{Dir}[\theta](\cdot) := \mathbf{P}[(Y_1/|Y|, \dots, Y_K/|Y|) \in \cdot], \tag{3.25}$$

where $Y_1, \dots, Y_K$ are independent random variables with Y_i being gamma $(\theta_i, 1)$ distributed, $Y = (Y_1, \dots, Y_K)$ and $|Y| = Y_1 + \cdots + Y_K$. [By definition, gamma(0, 1) is the distribution of the zero random variable.] In general, the unique stationary distribution, which we again denote by $\Pi[\theta]$, is absolutely continuous with respect to $\mathrm{Dir}[\theta]$ with

$$\big(d\Pi[\theta]/d\mathrm{Dir}[\theta]\big)(x) = c(\theta)^{-1} \exp\left\{ \sum_{i,j=1}^K \sigma_{ij} x_i x_j \right\}, \tag{3.26}$$

where $c(\theta)$ depends implicitly on σ. The transition rates (3.13)–(3.15) [or (3.7) and (3.8) in the haploid case (1.19)] can now be defined (even if $\theta = 0$, because any undefined factors are multiplied by 0), and we denote the transition probabilities of the resulting birth-and-death process in $\mathbf{Z}_+^K$ by $P_{\alpha\beta}(t)$.

THEOREM 3.1. *Suppose* $2 \le K < \infty$, $\theta_1 \ge 0, \dots, \theta_K \ge 0$, *and* (σ_{ij}) *is a real symmetric* $K \times K$ *matrix. With* ρ *as in* (1.12), *assume that, for each* $t > 0$ *and* $y \in \Delta_K$,

$$b_\beta(t, y) := \liminf_{\rho(\alpha)\to(\infty, y)} P_{\alpha\beta}(t), \qquad \beta \in \mathbf{Z}_+^K \tag{3.27}$$

defines a probability distribution on $\mathbf{Z}_+^K$ *and that this probability distribution is weakly continuous in* y *as well as in* θ. *Then the diffusion model in* Δ_K *with generator* L *as in* (1.2) *and* (3.24) *has transition function* $P(t, x, dy)$ *given for each* $t > 0$ *and* $x \in \Delta_K$ *by*

$$P(t, x, \cdot) = \sum_{\alpha\in\mathbf{Z}_+^K} b_\alpha(t, x)\Pi[\alpha + \theta](\cdot). \tag{3.28}$$

REMARK. The hypotheses concerning (3.27) will be verified in Sections 4–6 and 8.

PROOF. We know from Theorem 2.2 that (3.28) holds if $\theta_1 > 0, \dots, \theta_K > 0$. The left side of (3.28) is weakly continuous in θ by Trotter's semigroup approximation theorem, and the right side is weakly continuous in θ by our assumption that $b_\bullet(t, x)$ is weakly continuous in θ and the fact that $\Pi[\theta]$ is weakly continuous in θ (except at $\theta = 0$ where it is undefined). If $\theta = 0$, the death rates take the form

$$q(\alpha, \alpha - \varepsilon^i) = \tfrac{1}{2}\alpha_i(|\alpha| - 1)\frac{c(\alpha - \varepsilon^i)}{c(\alpha)}. \tag{3.29}$$

In this case the $\alpha = 0$ term in (3.28) is absent, because $b_0(t, x) = 0$. Thus, (3.28) holds in general. □

The case of no mutation ($\theta = 0$) and two alleles ($K = 2$) was studied by Kimura (1955, 1957).

As in Corollary 2.3, if $|\theta| > 0$, then

$$d_{TV}\big(P(t, x, \cdot), \Pi[\theta](\cdot)\big) \le 1 - b_0(t, x), \qquad t > 0,\ x \in \Delta_K, \tag{3.30}$$

thereby generalizing (1.9).

4. The entrance process: preliminaries. Let X be a (continuous time) pure jump Markov process in $\mathbf{Z}_+^K$, whose transitions are specified by

$$\begin{aligned} &\alpha \to \alpha - \varepsilon^j && \text{at rate } \alpha_j|\alpha| + f_{1j}(\alpha),\\ &\alpha \to \alpha + \varepsilon^j && \text{at rate } f_{2j}(\alpha),\\ &\alpha \to \alpha + \varepsilon^i + \varepsilon^j && \text{at rate } f_{3ij}(\alpha), \end{aligned} \tag{4.1}$$

where the f-functions are Lipschitz and satisfy

$$
\begin{aligned}
f_{lj}(0) &= 0, \qquad |f_{lj}(\alpha) - f_{lj}(\alpha')| \le c_{lj}|\alpha - \alpha'|, \\
f_{3ij}(0) &= 0, \qquad \left|f_{3ij}(\alpha) - f_{3ij}(\alpha')\right| \le c_{3ij}|\alpha - \alpha'|,
\end{aligned} \tag{4.2}
$$

for all α, $\alpha' \in \mathbf{Z}_+^K$, $l = 1, 2$, and $i, j = 1, \dots, K$, for some nonnegative c's; assume also that there exists $C_1 \ge 0$ such that

$$f_{1j}(\alpha) \ge -C_1\alpha_j \tag{4.3}$$

for all $\alpha \in \mathbf{Z}_+^K$ and $j = 1, \dots, K$. Here and throughout, $|w| := \sum_{j=1}^K |w_j|$ for $w \in \mathbf{R}^K$. Such a process races "away from infinity" very fast, because of the downward rates which are quadratic in the components of X. In this section and the one that follows, given any $p \in \Delta_K$, we show that there is a Markov process $\{X^p(t),\ t > 0\}$ in $\mathbf{Z}_+^K$ with these transitions that satisfies $\lim_{t\to 0} |X^p(t)| = \infty$ and $\lim_{t\to 0} X^p(t)/|X^p(t)| = p$. We establish the existence of X^p as the weak limit of a sequence of processes $X^{(n)}$, the nth of which is defined for $t \ge 1/n$ and has $|X^{(n)}(1/n)| = n$ and $p^{(n)} := n^{-1}X^{(n)}(1/n)$ as close as possible to the specified p. We are interested in processes of this general form, which covers processes with rates given by (3.7) and (3.8), as well as those with rates given by (3.13)–(3.15) (see Section 8), provided that in these applications time is speeded up by a factor of 2; this latter modification of course has no effect on the existence of the limit process X^p.

The argument is based on two observations. The first is that the processes $|X^{(n)}(t)|$ do not get too far from the curve $x(t) = 1/t$, which is a crude approximation to the evolution of the expectation of $|X^{(n)}(t)|$ when the f-functions are identically zero. The second is that if two such processes X and X' start with $|X(1/n)| = |X'(1/n)| = n$ and with $X(1/n)$ not too different from $X'(1/n)$, then they can be constructed on a single probability space in such a way that $X(u) = X'(u)$ for all $u \ge u_n$ with probability at least $1 - q_n$, where $\lim_{n\to\infty} u_n = \lim_{n\to\infty} q_n = 0$.

These observations allow one to make explicit statements about the total variation distance between the distributions of the processes $X^{(m)}$ and $X^{(n)}$ on $t \ge t_0$ for any $t_0 > 0$, as follows. Supposing that $m > n$, let $X^{(m)}$ evolve until $|X^{(m)}|$ first hits n, which takes place at a time $\tau_n^{(m)}$ not very different from $1/n$, by the first observation. At this time, the proportions $X^{(m)}(\tau_n^{(m)})/|X^{(m)}(\tau_n^{(m)})|$ are little different from their initial values (in fact, uniformly in $m > n$), which are themselves not too different from those of $X^{(n)}$, because both are close to p. Now consider $X := X^{(n)}$ and $X'(\cdot) := X^{(m)}(\cdot + \tau_n^{(m)} - 1/n)$ and apply the second observation: the total variation distance between their distributions on $t \ge u_n$ is at most q_n (again, uniformly in $m > n$). Then it suffices to establish that the probability that $X^{(n)}$ has a transition in any interval $[t_0, t_0 + \eta)$ becomes small with η, uniformly in n, so that, using the Markov property, the coupling of X and X' can be modified to one of $X^{(n)}$ and $X^{(m)}$, without the small random time shift. This establishes the existence of the weak limit of $X^{(n)}$ as $n \to \infty$ on any interval $[t_0, \infty)$ for any $t_0 > 0$ and hence of the required process X^p.

Turning to the details of the argument, the simplest part is the last, showing that, for any given $t_0 > 0$, the probability that $X^{(n)}$ has a transition in $[t_0, t_0 + \eta)$ becomes small with η, uniformly in n. For this, it is enough to show that there is a stochastic upper bound for the distribution of $|X^{(n)}(t_0)|$ which is uniform in $n \geq n_0(t_0)$ and in $X^{(n)}(1/n)$, subject to $|X^{(n)}(1/n)| = n$. This is because the sum of all the transition rates of X is bounded, by B_M, say, in the set $\mathscr{X}_M := \{\alpha: |\alpha| \leq M\}$, for any choice of M; the chance of a transition in a time interval of length η when starting in any state of $\mathscr{X}_M$ is then no more than ηB_M, which can be made small by choice of η for any fixed M; and the probability that $X^{(n)}(t_0) \in \mathscr{X}_M$ can be made arbitrarily close to 1, by choice of M.

Before proceeding any further, we identify a class of martingales associated with a countable-state pure jump Markov process that will be needed in what follows. Let X be a nonexplosive pure jump Markov process in a countable state space $\mathscr{S}$ with infinitesimal matrix (q_{ij}) and with natural filtration $\mathscr{F}_t$, $t \geq 0$. Let $U: \mathscr{S} \times \mathbf{R}_+ \to \mathbf{R}$ be such that, for each $i \in \mathscr{S}$, $U(i, \cdot)$ is absolutely continuous and $\sum_{j \in \mathscr{S}} q_{ij} |U(j, t)|$ is bounded on bounded intervals. Define the action of the space–time generator $\mathscr{A}$ on U by

$$(\mathscr{A}U)(i, t) = \frac{\partial U}{\partial t}(i, t) + \sum_{j:\ j \neq i} q_{ij} \{U(j, t) - U(i, t)\}. \tag{4.4}$$

Let $\mathbf{E}_i$ denote expectation conditional on $X(0) = i$.

LEMMA 4.1. *Suppose that, for each $i \in \mathscr{S}$, $\mathbf{E}_i |(\mathscr{A}U)(X(u), u)|$ is integrable over finite intervals and that $\mathbf{E}_i \{\sup_{0 \leq u \leq t} |U(X(u), u)|\} < \infty$ for each t. Then*

$$M(t) := U(X(t), t) - \int_0^t (\mathscr{A}U)(X(u), u)\, du \tag{4.5}$$

is an $\mathscr{F}_t$-martingale.

PROOF. Let the sequence of jump times of X be denoted by τ_n, $n \geq 1$. Then a simple calculation using Fubini's theorem shows that, for each $i \in \mathscr{S}$ and $t > 0$,

$$\mathbf{E}_i U(X(\tau_1 \wedge t), \tau_1 \wedge t) = U(i, 0) + \mathbf{E}_i \left\{ \int_0^{\tau_1 \wedge t} (\mathscr{A}U)(X(u), u)\, du \right\}, \tag{4.6}$$

the expectations existing because of the assumptions on U, from which and the strong Markov property it follows that, for fixed $t > 0$, the sequence $M(\tau_n \wedge t)$ is an $\mathscr{F}_{\tau_n \wedge t}$-martingale and hence that

$$\begin{aligned} &\mathbf{E}_i \left\{ U(X(\tau_n), \tau_n) 1_{\{\tau_n < t\}} \right\} + \mathbf{E}_i \left\{ U(X(t), t) 1_{\{\tau_n \geq t\}} \right\} \\ &\quad = U(i, 0) + \mathbf{E}_i \left\{ \int_0^{\tau_n} (\mathscr{A}U)(X(u), u)\, du\, 1_{\{\tau_n < t\}} \right\} \\ &\qquad + \mathbf{E}_i \left\{ \int_0^t (\mathscr{A}U)(X(u), u)\, du\, 1_{\{\tau_n \geq t\}} \right\}. \end{aligned} \tag{4.7}$$

Letting $n \to \infty$ and using the conditions of the lemma to justify application of the dominated convergence theorem, it follows that $\mathbf{E}_i M(t) = M(0)$ for all i and t, and the Markov property concludes the proof. □

We now formalize the discussion that preceded Lemma 4.1 with the following lemma.

LEMMA 4.2. *There exists a constant C such that, for each n, $\mathbf{E}|X^{(n)}(t)| \le C(1 + 1/t)$ for all $t \ge 1/n$.*

PROOF. To establish such a bound, compare the process $|X^{(n)}|$ with a process $\widetilde{S}^{(n)}$ on $\mathbf{Z}_+ \cap [l, \infty)$, for suitably chosen l, which also has $\widetilde{S}^{(n)}(1/n) = n$, but which has transition rates given by

$$\begin{aligned} & i \to i-1 \quad \text{at rate } (1-\varepsilon)i^2, \qquad i \ge l+1; \\ & i \to i+1 \quad \text{at rate } \tfrac{1}{2}\varepsilon i^2, \qquad i \ge l; \\ & i \to i+2 \quad \text{at rate } \tfrac{1}{2}\varepsilon i^2, \qquad i \ge l; \end{aligned} \tag{4.8}$$

where $\varepsilon = \varepsilon_l$ is such that

$$\begin{gathered} |\alpha|^2 + \sum_{j=1}^{K} f_{1j}(\alpha) \ge (1-\varepsilon)(|\alpha|+1)^2, \qquad \sum_{j=1}^{K} f_{2j}(\alpha) \le \tfrac{1}{2}\varepsilon|\alpha|^2, \\ \sum_{i,\, j=1}^{K} f_{3ij}(\alpha) \le \tfrac{1}{2}\varepsilon|\alpha|^2, \end{gathered} \tag{4.9}$$

whenever $|\alpha| \ge l$, and l is chosen large enough to ensure that $\varepsilon_l < 1/3$. This process can be simply coupled with $|X^{(n)}|$ in such a way that $\widetilde{S}^{(n)}(t) \ge |X^{(n)}(t)|$ for all $t \ge 1/n$, because both are integer-valued processes that make downward jumps only of size 1 and upward jumps of sizes only 1 and 2. So run the two processes independently whenever $\widetilde{S}^{(n)}(t) > |X^{(n)}(t)| + 1$; whenever $\widetilde{S}^{(n)}(t) \in \{|X^{(n)}(t)|, |X^{(n)}(t)| + 1\}$, use Bernoulli thinnings of the upward jumps of $\widetilde{S}^{(n)}$ for the upward jumps of $|X^{(n)}|$, and a Bernoulli thinning of the downward jumps of $|X^{(n)}|$ for those of $\widetilde{S}^{(n)}$, so that $|X^{(n)}|$ cannot jump past $\widetilde{S}^{(n)}$. Thus it suffices to prove the bound in the lemma for $\mathbf{E}\widetilde{S}^{(n)}(t)$.

Any process $\widetilde{S}$ with transition rates given by (4.8) has a limiting stationary distribution μ which has all its moments, as is seen by comparison with the process obtained by replacing the factors i^2 by l^2 in (4.8), which is a reflecting random walk in continuous time with negative drift. What is more, versions of $\widetilde{S}$ starting in equilibrium and at l can be coupled so that the former is always at least as large as the latter, much as in Lindvall (1992), Chapter V.4, since the double upward jumps of the latter can be obtained as a thinning of those of the former. Hence it follows that

$$\mathbf{E}\widetilde{S}(t)1_{\{\tau_l \le t\}} = \int_0^t f_{\tau_l}(u)\mathbf{E}_l\widetilde{S}(t-u)\,du \le \mathbf{E}_\mu\widetilde{S}; \tag{4.10}$$

here, τ_l is the time of first hitting l and f_{τ_l} its probability density. Hence, to control $\mathbf{E}\widetilde{S}(t)$, we merely need a bound for the transient component $\mathbf{E}\{\widetilde{S}(t) 1_{\{\tau_l>t\}}\}$, for which a bound on the larger quantity

$$\tilde{m}_l(t) := \mathbf{E}\{\widetilde{S}(t \wedge \tau_l)\} \tag{4.11}$$

suffices.

Using the fact that $\widetilde{S}^* := \sup_{0\le t\le\tau_l} \widetilde{S}(t)$ is a proper random variable with all moments finite, again from the random walk comparison, it follows from Lemma 4.1 that

$$\widetilde{S}(t \wedge \tau_l) + \int_0^{t\wedge\tau_l} (1 - \tfrac{5}{2}\varepsilon)\widetilde{S}^2(u)\,du \tag{4.12}$$

is a martingale, from which we deduce that

$$\begin{aligned} h^{-1}\{\tilde{m}_l(t+h) - \tilde{m}_l(t)\} &= -h^{-1}(1-\tfrac{5}{2}\varepsilon)\mathbf{E}\left\{\int_{t\wedge\tau_l}^{(t+h)\wedge\tau_l} \widetilde{S}^2(u)\,du\right\} \\ &= -h^{-1}(1-\tfrac{5}{2}\varepsilon)\mathbf{E}\left\{\int_t^{t+h} \widetilde{S}^2(u)1_{\{u<\tau_l\}}du\right\}. \end{aligned} \tag{4.13}$$

Letting $h \to 0$ and using dominated convergence, it follows that $\tilde{m}_l$ is right-differentiable; a similar argument shows that it is left-differentiable also, and, furthermore, it follows that

$$\begin{aligned} \frac{d\tilde{m}_l}{dt}(t) &= -\left(1 - \frac{5}{2}\varepsilon\right)\mathbf{E}[\widetilde{S}^2(t)1_{\{t<\tau_l\}}] \\ &\le -\left(1 - \frac{5}{2}\varepsilon\right)[\mathbf{E}(\widetilde{S}(t)1_{\{t<\tau_l\}})]^2 \le -\left(1 - \frac{5}{2}\varepsilon\right)(\tilde{m}_l(t) - l)^2. \end{aligned} \tag{4.14}$$

Now $\widetilde{S}^{(n)}$ has transition rates given by (4.8), and $\tilde{m}_l^{(n)}(1/n) = \mathbf{E}\{\widetilde{S}^{(n)}(1/n)\} = n$; integrating from $1/n$ to t, it thus easily follows that

$$\tilde{m}_l^{(n)}(t) \le l + \frac{1}{t\left(1 - \frac{5}{2}\varepsilon\right)} \tag{4.15}$$

for all $t \ge 1/n$. Hence, using (4.10) and (4.11), we have shown that

$$\mathbf{E}\widetilde{S}^{(n)}(t) \le \mathbf{E}_\mu \widetilde{S} + l + \frac{1}{t\left(1 - \frac{5}{2}\varepsilon\right)} \tag{4.16}$$

for all $n > l$ and $t \ge 1/n$, which is enough. □

The next step is to establish that, for small values of t, $|X^{(n)}(t)|$ remains close to $1/t$. We do this in the following form.

LEMMA 4.3. *Fix any* $0 < \varphi < 1$. *Then, for any* $m \geq 5$ *and for all* $1/m \leq T \leq 1/4$,

$$\mathbf{P}\left[\sup_{1/m \leq t \leq T} \frac{1}{t}\left|\frac{1}{|X^{(m)}(t)|} - t\right| > \varphi\right] \leq c(1 + 1/\varphi)^2 T, \tag{4.17}$$

where $c_l = \sum_{j=1}^{K} c_{lj}$ *for* $l = 1, 2$, $c_3 = \sum_{i,\, j=1}^{K} c_{3ij}$ *and* $c = 30(1 + c_1 + c_2 + c_3)^2$.

PROOF. The idea behind the proof is that $t - 1/|X(t)|$ is almost a martingale and has variance $O(t^3)$; this is combined with a stopping argument to obtain the desired result.

Define a stopping time $\sigma = \sigma(\varphi)$ for an X-process started at initial time v by

$$\sigma = \inf\left\{u \geq v\colon \left|\frac{1}{|X(u)|} - u\right| \geq \varphi u\right\}, \tag{4.18}$$

and note that $|X(t)| \geq 2$ for all $t \leq \min\{\sigma, 1/4\}$. Let $\mathscr{A}$ denote the space–time generator of X as in (4.4), and take $U(\alpha, t) = (t - 1/|\alpha|)^2$, for which, if $n = |\alpha| \geq 2$,

$$\begin{aligned} |(\mathscr{A}U)(\alpha, u)| &= \Bigg| 2\left(u - \frac{1}{n}\right) + \left(n^2 + \sum_{j=1}^{K} f_{1j}(\alpha)\right) \\ &\quad \times \left\{\frac{1}{(n-1)^2} - \frac{1}{n^2} - 2u\left(\frac{1}{n-1} - \frac{1}{n}\right)\right\} \\ &\quad + \left(\sum_{j=1}^{K} f_{2j}(\alpha)\right)\left\{\frac{1}{(n+1)^2} - \frac{1}{n^2} - 2u\left(\frac{1}{n+1} - \frac{1}{n}\right)\right\} \\ &\quad + \left(\sum_{i,\, j=1}^{K} f_{3ij}(\alpha)\right)\left\{\frac{1}{(n+2)^2} - \frac{1}{n^2} - 2u\left(\frac{1}{n+2} - \frac{1}{n}\right)\right\}\Bigg| \\ &\leq 4(1 + c_1 + c_2 + c_3)\{n^{-2} + n^{-1}|u - n^{-1}|\}, \end{aligned} \tag{4.19}$$

where, in obtaining the last line, cancellation between the terms not involving f-functions is exploited. Applying Lemma 4.1, it thus follows that

$$U(X(t \wedge \sigma), t \wedge \sigma) - \int_{v}^{t \wedge \sigma} (\mathscr{A}U)(X(u), u)\, du \tag{4.20}$$

is a martingale in $v \le t \le 1/4$. Now, if $U(X(v), v) = 0$, which is the case for $X^{(m)}$ if $v = 1/m$, it follows by the martingale property that, for any $v \le t \le 1/4$,

$$
\begin{aligned}
&\mathbf{E}\{[(t \wedge \sigma) - 1/|X(t \wedge \sigma)|]^2\} \\
&= \left| \mathbf{E} \int_v^{t\wedge\sigma} (\mathscr{A}U)(X(u), u)\, du \right| \\
&\quad \le 4(1 + c_1 + c_2 + c_3) \\
&\qquad \times \mathbf{E}\left\{ \int_v^{t\wedge\sigma} (1+\varphi)^2 u^2\, du + \int_v^{t\wedge\sigma} (1+\varphi) u \left| u - 1/|X(u)| \right| du \right\} \\
&\quad \le 4(1 + c_1 + c_2 + c_3) \\
&\qquad \times \left\{ (1+\varphi)^2 \int_v^t u^2\, du + (1+\varphi) \int_v^t u \sqrt{\mathbf{E}\big[(u - 1/|X(u)|)^2 1_{\{u<\sigma\}}\big]}\, du \right\}.
\end{aligned} \tag{4.21}
$$

Letting

$$
m_t^* = \sup_{v \le s \le t} \mathbf{E}\{[(s \wedge \sigma) - 1/|X(s \wedge \sigma)|]^2\}, \tag{4.22}
$$

it now follows that

$$
m_t^* \le 4(1 + c_1 + c_2 + c_3) \left\{ (1+\varphi)^2 \int_v^t u^2\, du + (1+\varphi)\sqrt{m_t^*} \int_v^t u\, du \right\}, \tag{4.23}
$$

and hence that, for all $v \le t \le 1/4$,

$$
\begin{aligned}
m_t^* &\le \left[4(1 + c_1 + c_2 + c_3)(1+\varphi) \int_v^t u\, du \right]^2 \\
&\qquad + 8(1 + c_1 + c_2 + c_3)(1+\varphi)^2 \int_v^t u^2\, du \\
&\le \tfrac{64}{9}(1 + c_1 + c_2 + c_3)^2 (1+\varphi)^2 (t^{3/2} - v^{3/2})^2 \\
&\qquad + \tfrac{8}{3}(1 + c_1 + c_2 + c_3)(1+\varphi)^2 (t^3 - v^3) \\
&\le 30(1 + c_1 + c_2 + c_3)^2 (1+\varphi)^2 \int_v^t u^2\, du,
\end{aligned} \tag{4.24}
$$

since, if x, k_1 and k_2 are all nonnegative and $x^2 \le k_1 + k_2 x$, then $x^2 \le 2k_1 + k_2^2$.

However, if F_σ denotes the distribution function of σ for $X^{(m)}$, which has initial time point $v = 1/m$, it follows from the definition of m_t^* that

$$
\begin{aligned}
m_t^* &\ge \mathbf{E}\big\{[\sigma - 1/|X^{(m)}(\sigma)|]^2 1_{\{\sigma \le t\}}\big\} \\
&\ge \mathbf{E}\big\{(\varphi\sigma)^2 1_{\{\sigma \le t\}}\big\} = \int_{1/m}^t (\varphi u)^2 F_\sigma(du)
\end{aligned} \tag{4.25}
$$

for all $t \geq 1/m$. However, by Fubini's theorem,

$$\mathbf{P}\left[\sup_{1/m\leq t\leq T} \frac{1}{t}\left|\frac{1}{|X^{(m)}(t)|} - t\right| > \varphi\right] \leq \int_{1/m}^{T} F_\sigma(du) \tag{4.26}$$
$$= T^{-2}\int_{1/m}^{T} u^2 F_\sigma(du) + 2\int_{1/m}^{T} t^{-3}\int_{1/m}^{t} u^2 F_\sigma(du)\,dt$$

and the lemma follows from (4.24)–(4.26). □

COROLLARY 4.4. *For any $m > n$, let $\tau_n^{(m)}$ denote the time at which $|X^{(m)}|$ first takes the value n. Then*

$$\mathbf{P}\left[\left|\tau_n^{(m)} - n^{-1}\right| \leq \left(\frac{\varphi}{1-\varphi}\right)\frac{1}{n}\right] \geq 1 - c(1+1/\varphi)^2\frac{1}{n(1-\varphi)}, \tag{4.27}$$

for all $m > n$, where c is as in (4.17).

PROOF. Take $T = \{n(1-\varphi)\}^{-1}$ in Lemma 4.3. If $\sup_{1/m\leq t\leq T} t^{-1}\left|t - 1/|X^{(m)}(t)|\right| \leq \varphi$, which is the case with probability at least $1 - c(1 + 1/\varphi)^2\{n(1-\varphi)\}^{-1}$, it follows that $\{n(1+\varphi)\}^{-1} \leq \tau_n^{(m)} \leq \{n(1-\varphi)\}^{-1}$, and the corollary is immediate. □

LEMMA 4.5. *Suppose that $q_j = X_j(v)/|X(v)|$ for $1 \leq j \leq K$. Then, for all $v \leq T \leq 1/4$ and for any $\eta > 0$,*

$$\mathbf{P}\left[\sup_{v\leq s\leq T}\max_{1\leq j\leq K}\left|q_j - X_j(s)/|X(s)|\right| > \eta\right] \tag{4.28}$$
$$\leq \{c(1+1/\varphi)^2 + c'\eta^{-2}\}T,$$

where c is as in (4.17), *$c' = 2k_1 + k_2^2$, and k_1 and k_2 are given in* (4.37) *below.*

PROOF. Start with $U(\alpha, t) = U(\alpha) = \sum_{j=1}^{K}(q_j - \alpha_j/|\alpha|)^2$, and observe that now, using Δ in this proof to denote differences, and once again writing $n = |\alpha|$, we have

$$(\Delta_i^- U)(\alpha) := U(\alpha - \varepsilon^i) - U(\alpha)$$
$$= -2\sum_{j=1}^{K}\frac{\alpha_j}{n(n-1)}\left(q_j - \frac{\alpha_j}{n} - \frac{\alpha_j}{2n(n-1)}\right) \tag{4.29}$$
$$+\frac{2}{n-1}\left(q_i - \frac{\alpha_i}{n} - \frac{\alpha_i}{2n(n-1)}\right) + \frac{n-\alpha_i}{n(n-1)^2},$$

so that $\sum_{i=1}^{K}\alpha_i n(\Delta_i^- U)(\alpha) = \sum_{i=1}^{K}\alpha_i(n-\alpha_i)(n-1)^{-2}$ and

$$\left|(\Delta_i^- U)(\alpha)\right| \leq \frac{2}{n(n-1)}\sum_{j=1}^{K}\left\{\alpha_j\left|q_j - n^{-1}\alpha_j\right| + \frac{\alpha_j^2}{2n(n-1)}\right\} \tag{4.30}$$
$$+ \frac{2}{n-1}|q_i - n^{-1}\alpha_i| + (n-1)^{-2};$$

similarly, it follows that

$$\begin{aligned}\left|(\Delta_i^+ U)(\alpha)\right| &:= \left|U(\alpha+\varepsilon^i)-U(\alpha)\right| \\ &\le \frac{2}{n(n+1)}\sum_{j=1}^K\left\{\alpha_j\left|q_j-n^{-1}\alpha_j\right|+\frac{\alpha_j^2}{2n(n+1)}\right\} \\ &\quad+\frac{2}{n+1}\left|q_i-n^{-1}\alpha_i\right|+(n+1)^{-2}\end{aligned}\tag{4.31}$$

and that

$$\begin{aligned}\left|(\Delta_{ij}^{++} U)(\alpha)\right| &:= \left|U(\alpha+\varepsilon^i+\varepsilon^j)-U(\alpha)\right| \\ &\le \frac{4}{n(n+2)}\sum_{l=1}^K\left\{\alpha_l\left|q_l-n^{-1}\alpha_l\right|+\frac{\alpha_l^2}{n(n+2)}\right\} \\ &\quad+\frac{2}{n+2}\left\{\left|q_i-n^{-1}\alpha_i\right|+\left|q_j-n^{-1}\alpha_j\right|\right\} \\ &\quad+2(1+\delta_{ij})(n+2)^{-2}.\end{aligned}\tag{4.32}$$

Hence, in $n \ge 2$, invoking (4.2), we obtain

$$\begin{aligned}|(\mathscr{A}U)(\alpha,u)| &\le \sum_{i=1}^K\frac{\alpha_i(n-\alpha_i)}{(n-1)^2}+\sum_{i,\,j=1}^K f_{3ij}(\alpha)\left|(\Delta_{ij}^{++}U)(\alpha)\right| \\ &\quad+\sum_{i=1}^K\left\{|f_{1i}(\alpha)|\,|(\Delta_i^- U)(\alpha)|+f_{2i}(\alpha)|(\Delta_i^+ U)(\alpha)|\right\},\end{aligned}\tag{4.33}$$

with the various terms estimated in $n \ge 2$ as follows: first, $\sum_{i=1}^K \alpha_i(n-\alpha_i)(n-1)^{-2} \le 2$; for the second term, we have

$$\begin{aligned}&\sum_{i,\,j=1}^K f_{3ij}(\alpha)\left|(\Delta_{ij}^{++}U)(\alpha)\right| \\ &\quad\le 4c_3 n^{-1}\sum_{l=1}^K\left\{\alpha_l\left|q_l-n^{-1}\alpha_l\right|+n^{-2}\alpha_l^2\right\} \\ &\qquad+4\sum_{i,\,j=1}^K c_{3ij}\left|q_i-n^{-1}\alpha_i\right|+4c_3n^{-1} \\ &\quad\le 8c_3\left\{\left(\sum_{i=1}^K(q_i-n^{-1}\alpha_i)^2\right)^{1/2}+n^{-1}\right\},\end{aligned}\tag{4.34}$$

where $c_{3ij} = c_{3ji}$ is assumed without loss of generality; for the remainder, in similar fashion, the bound

$$\begin{aligned}&\sum_{i=1}^K\left\{|f_{1i}(\alpha)|\,|(\Delta_i^- U)(\alpha)|+f_{2i}(\alpha)|(\Delta_i^+ U)(\alpha)|\right\} \\ &\quad\le 8(c_1+c_2)\left\{\left(\sum_{i=1}^K(q_i-n^{-1}\alpha_i)^2\right)^{1/2}+n^{-1}\right\}\end{aligned}\tag{4.35}$$

is obtained. Combining these results, it follows that

$$|(\mathscr{A}U)(\alpha, u)| \le k_1 + k_2\left(\sum_{i=1}^{K}(q_i - n^{-1}\alpha_i)^2\right)^{1/2}, \tag{4.36}$$

where

$$k_1 = 2 + 4(c_1 + c_2 + c_3); \qquad k_2 = 8(c_1 + c_2 + c_3). \tag{4.37}$$

Now, constructing a martingale using Lemma 4.1 and noting that $U(X(v), v) = 0$, it follows much as in the proof of Lemma 4.3 that

$$n_t^* \le k_1(t - v) + k_2\sqrt{n_t^*}(t - v) \le k_1 t + k_2\sqrt{t n_t^*}, \tag{4.38}$$

where

$$n_t^* := \sup_{v \le s \le t} \mathbf{E}\left\{\sum_{j=1}^{K}\left(q_j - \frac{X_j(s \wedge \sigma \wedge \sigma^*)}{|X(s \wedge \sigma \wedge \sigma^*)|}\right)^2\right\} \tag{4.39}$$

and

$$\sigma^* = \sigma^*(\eta) := \inf\left\{u \ge v : \max_{1 \le j \le K}\left|q_j - X_j(u)/|X(u)|\right| > \eta\right\}; \tag{4.40}$$

from this, we find that $n_t^* \le c't$, where $c' = 2k_1 + k_2^2$, in $v \le t \le 1/4$. On the other hand,

$$n_t^* \ge \eta^2 \mathbf{P}[\sigma^* \le (\sigma \wedge t)] \ge \eta^2\{\mathbf{P}[\sigma^* \le t] - \mathbf{P}[\sigma < t]\}, \tag{4.41}$$

implying that

$$\mathbf{P}[\sigma^* \le t] \le \mathbf{P}[\sigma \le t] + \eta^{-2} n_t^*, \tag{4.42}$$

from which, using Lemma 4.3, the lemma follows. □

COROLLARY 4.6. *For any* $1 \le \psi(n) \le n^{1/2}$,

$$\mathbf{P}\left[\sum_{j=1}^{K}\left|n^{-1}X_j^{(m)}(\tau_n^{(m)}) - p_j^{(m)}\right| > Kn^{-1/2}\psi(n)\right] = O(\psi^{-2}(n)), \tag{4.43}$$

uniformly in $m > n$.

PROOF. Apply Lemmas 4.3 and 4.5 with $T = \{n(1 - \varphi)\}^{-1}$, φ fixed and $\eta = n^{-1/2}\psi(n)$. Then $\mathbf{P}[\sigma < T] = O(n^{-1})$, by Lemma 4.3; if $\sigma \ge T$, then $\tau_n^{(m)} \le T$, and Lemma 4.5 completes the proof. □

5. The entrance process: coupling construction. As a consequence of the lemmas of the previous section, it can be seen that for each $m > n$ the process $X^{(m)}$ hits $|\alpha| = n$ close to the time $1/n$, with $X^{(m)}/|X^{(m)}|$ almost unchanged from its initial value. The next stage involves showing that two X-processes, both starting with the same value n of $|X|$ and having similar initial values of X/n, can be constructed on the same probability space in such a way that they very soon coincide with high probability. To do this, we construct a bivariate process (Y, Y') such that $|Y(t)| = |Y'(t)|$ for all $t \geq 0$ and such that, usually, $Y(t) = Y'(t)$ for all $t \geq t_1$, for some small t_1; we then show that, with high probability, (Y, Y') coincides with the required pair of X-processes.

The bivariate process (Y, Y') is a pure jump Markov process on $\mathbf{Z}_+^K \times \mathbf{Z}_+^K$ which starts with $|Y(0)| = |Y'(0)|$, and has transition rates given in $|\alpha| \geq C_1$ by

$$
\begin{aligned}
&(1)\ (\alpha, \alpha') \to (\alpha, \alpha') - (\varepsilon^j, \varepsilon^j) \\
&\qquad\qquad \text{at rate } (\alpha_j \wedge \alpha'_j)(|\alpha| - C_1) + \tilde{f}_{1j}(\alpha) \wedge \tilde{f}_{1j}(\alpha'), \\
&\qquad\qquad\qquad 1 \leq j \leq K, \\
&(2)\ (\alpha, \alpha') \to (\alpha, \alpha') + (\varepsilon^j, \varepsilon^j) \quad \text{at rate } f_{2j}(\alpha) \wedge f_{2j}(\alpha'), \quad 1 \leq j \leq K, \\
&(3)\ (\alpha, \alpha') \to (\alpha, \alpha') + (\varepsilon^i + \varepsilon^j, \varepsilon^i + \varepsilon^j) \quad \text{at rate } f_{3ij}(\alpha) \wedge f_{3ij}(\alpha'), \\
&\qquad\qquad\qquad 1 \leq i, j \leq K, \\
&(4)\ (\alpha, \alpha') \to (\alpha, \alpha') - (\varepsilon^j, \varepsilon^l) \quad \text{at rate } \pi_{jl}(\alpha, \alpha')(|\alpha| - C_1), \\
&\qquad\qquad\qquad 1 \leq j \neq l \leq K,
\end{aligned} \tag{5.1}
$$

where $\tilde{f}_{1j}(\alpha) := f_{1j}(\alpha) + C_1\alpha_j \geq 0$ from (4.3) and the π_{jl} are chosen to satisfy the "total variation" matching conditions

$$
\begin{aligned}
\sum_{l:\ l \neq j} \pi_{jl}(\alpha, \alpha') &= \alpha_j - (\alpha_j \wedge \alpha'_j), \qquad 1 \leq j \leq K, \\
\sum_{j:\ j \neq l} \pi_{jl}(\alpha, \alpha') &= \alpha'_l - (\alpha_l \wedge \alpha'_l), \qquad 1 \leq l \leq K;
\end{aligned} \tag{5.2}
$$

for example, take

$$
\pi_{jl}(\alpha, \alpha') = (\alpha_j - \alpha'_j)_+ (\alpha'_l - \alpha_l)_+ \Big/ \sum_{i=1}^{K} (\alpha_i - \alpha'_i)_+. \tag{5.3}
$$

In $|\alpha| < C_1$, all transition rates of (Y, Y') are taken to be 0; that is, (Y, Y') is stopped once $|Y(t)| < C_1$.

With this construction, while $|Y(t)| \geq C_1$, the marginal processes Y and Y' are arranged to have exactly the same quadratic elements in their transition rates as do X-processes, but the additional perturbing elements, involving the various f-functions, are a little different; nonetheless, if $|Y(1/m)| = m$, then the conclusion of Lemma 4.3 still holds for the process Y.

LEMMA 5.1. *If (Y, Y') satisfies $|Y(1/m)| = m$, then, for any $m \geq 5$ and for any $1/m \leq T \leq \left(\frac{1}{4} \wedge \{C_1(1+\varphi)\}^{-1}\right)$, it follows that*

$$\mathbf{P}\left[\sup_{1/m \leq t \leq T} \frac{1}{t}\left|\frac{1}{|Y(t)|} - t\right| > \varphi\right] \leq c''(1 + 1/\varphi)^2 T, \tag{5.4}$$

where c'' is defined in the same way as c in (4.17), *but with $c_1'' = c_1 \vee C_1$ in place of c_1.*

PROOF. The proof is similar to that of Lemma 4.3. Although the process (Y, Y') is bivariate, taking $U(\alpha, \alpha', t) = (t - 1/|\alpha|)^2$ still yields essentially the same bound for $|(\mathscr{A}U)(\alpha, \alpha', u)|$ as previously for $|(\mathscr{A}U)(\alpha, u)|$, because

$$0 \leq f_{2j}(\alpha) \wedge f_{2j}(\alpha') \leq f_{2j}(\alpha), \qquad 0 \leq f_{3ij}(\alpha) \wedge f_{3ij}(\alpha') \leq f_{3ij}(\alpha) \tag{5.5}$$

and totalling the rates for transitions in which $\alpha \to \alpha - \varepsilon^j$ gives

$$\begin{aligned} &\alpha_j(|\alpha| - C_1) + \tilde{f}_{1j}(\alpha) \wedge \tilde{f}_{1j}(\alpha') \\ &\qquad \in \left[\alpha_j(|\alpha| - C_1), \alpha_j|\alpha| + f_{1j}(\alpha)\right], \end{aligned} \tag{5.6}$$

so that $\sum_{j=1}^K f_{1j}(\alpha)$ in (4.19) is replaced by a quantity Σ' satisfying

$$-C_1|\alpha| \leq \Sigma' \leq \sum_{j=1}^K f_{1j}(\alpha) \leq c_1|\alpha|. \tag{5.7}$$

The remainder of the proof is identical; the restriction $T \leq \{C_1(1+\varphi)\}^{-1}$ ensures that $|Y(t)| \geq C_1$ for all $t \leq T$ such that $t^{-1}\big|t - 1/|Y(t)|\big| \leq \varphi$. □

The important feature of the construction is that $|Y|$ and $|Y'|$ change in the same way at each transition, and hence remain equal forever; however,

$$D(\alpha, \alpha') := \tfrac{1}{2}\sum_{j=1}^K |\alpha_j - \alpha'_j| \tag{5.8}$$

remains constant at transitions of the form (1), (2) and (3), but decreases by 1 at any transition of the form (4), the total rate of such transitions being $D(\alpha, \alpha')(|\alpha| - C_1)$. Thus $D(Y, Y')$ evolves according to a pure death process and is eventually absorbed in 0, after which time Y and Y' have identical paths, except when

$$T_2 := \inf\{t > 0\colon |Y(t)| \le C_1\} \tag{5.9}$$

occurs earlier.

LEMMA 5.2. *Define* $t_0 = 1/n$, *and suppose that* (Y, Y') *satisfies* $|Y(t_0)| = |Y'(t_0)| = n$ *and* $D(Y(t_0), Y'(t_0)) \le n\varepsilon$; *then it follows that* $D(Y(t), Y'(t)) = 0$ *for all* $t \ge t_1 := (\log n)^{-1}\sqrt{\varepsilon}$ *with probability at least* $1 - O(\log n\sqrt{\varepsilon})$, *uniformly in* $\varepsilon \le (\log n)^{-2}$ *and* $n \ge \exp(2(1 \vee C_1))$.

PROOF. Consider the process (Y, Y') evolving over $[t_0, t_1]$. The chance that the event $A = \bigcup_{t_0 \le t \le t_1}\{|Y(t)| < \{(1+\varphi)t\}^{-1}\}$ occurs is at most $O(t_1\varphi^{-2})$, by Lemma 5.1, and we take $\varphi = 1/\log n$ and t_0 and t_1 as defined above, so that $\mathbf{P}[A] = O(\log n\sqrt{\varepsilon})$ and, if A does not occur, then $|Y(t)| \ge C_1$ for all $t_0 \le t \le t_1$. So couple the process $D(Y, Y')$ with a pure death process D' starting with $D'(t_0) = D(Y(t_0), Y'(t_0))$ and having time dependent *per capita* death rate $\{(1+\varphi)t\}^{-1} - C_1$ in the following way. If $D' > D$, let both evolve independently of one another. Whenever $D' = D$, sample the jumps of D' as a Bernoulli thinning of those of D; the construction can break down only if A occurs, and, if A does not occur, $D'(t) \ge D(Y(t), Y'(t))$ for all $t_0 \le t \le t_1$. Hence

$$\mathbf{P}[D(Y(t_1), Y'(t_1)) > 0] \le \mathbf{P}[A] + \mathbf{E}D'(t_1). \tag{5.10}$$

However, D' is just a deterministic time change of a linear pure death process $\hat{D}$ with unit *per capita* death rate, specifically,

$$\begin{aligned} D'(t) &= \hat{D}\biggl(\int_{t_0}^{t}\{u(1+\varphi)\}^{-1}\,du - C_1(t - t_0)\biggr) \\ &= \hat{D}\bigl((1+\varphi)^{-1}\log(t/t_0) - C_1(t - t_0)\bigr) \end{aligned} \tag{5.11}$$

for all $t_0 \le t \le \{C_1(1+\varphi)\}^{-1}$, from which it follows that

$$\begin{aligned} \mathbf{E}D'(t_1) &= \{t_0/t_1\}^{1/(1+\varphi)}\exp(C_1(t_1 - t_0))D'(t_0) \\ &= O\bigl((nt_1)^{-1}n\varepsilon\bigr) = O(\log n\sqrt{\varepsilon}), \end{aligned} \tag{5.12}$$

uniformly in $\varepsilon \le (\log n)^{-2}$, completing the proof. □

We now consider a bivariate process (X, X') which, for $|X(t)| \ge C_1$, has the same transitions (1)–(4) of (5.1) as does (Y, Y'), and, in addition, has the transitions

$$
(5.13)\quad
\begin{aligned}
&(5)\ (\alpha,\alpha') \to (\alpha,\alpha') - (\varepsilon^j, 0) \quad \text{at rate } \tilde{f}_{1j}(\alpha) - (\tilde{f}_{1j}(\alpha) \wedge \tilde{f}_{1j}(\alpha')), \\
&\qquad\qquad 1 \le j \le K; \\
&(6)\ (\alpha,\alpha') \to (\alpha,\alpha') - (0, \varepsilon^j) \quad \text{at rate } \tilde{f}_{1j}(\alpha') - (\tilde{f}_{1j}(\alpha) \wedge \tilde{f}_{1j}(\alpha')), \\
&\qquad\qquad 1 \le j \le K; \\
&(7)\ (\alpha,\alpha') \to (\alpha,\alpha') + (\varepsilon^j, 0) \quad \text{at rate } f_{2j}(\alpha) - (f_{2j}(\alpha) \wedge f_{2j}(\alpha')), \\
&\qquad\qquad 1 \le j \le K; \\
&(8)\ (\alpha,\alpha') \to (\alpha,\alpha') + (0, \varepsilon^j) \quad \text{at rate } f_{2j}(\alpha') - (f_{2j}(\alpha) \wedge f_{2j}(\alpha')), \\
&\qquad\qquad 1 \le j \le K; \\
&(9)\ (\alpha,\alpha') \to (\alpha,\alpha') + (\varepsilon^i + \varepsilon^j, 0) \\
&\qquad\qquad \text{at rate } f_{3ij}(\alpha) - (f_{3ij}(\alpha) \wedge f_{3ij}(\alpha')), \quad 1 \le i, j \le K; \\
&(10)\ (\alpha,\alpha') \to (\alpha,\alpha') + (0, \varepsilon^i + \varepsilon^j) \\
&\qquad\qquad \text{at rate } f_{3ij}(\alpha') - (f_{3ij}(\alpha) \wedge f_{3ij}(\alpha')), \quad 1 \le i, j \le K.
\end{aligned}
$$

However, if $|X(t)| < C_1$ and $X(t) = X'(t)$, then X and X' are continued as identical X-processes, whereas if $|X(t)| < C_1$ and $X(t) \neq X'(t)$ they are continued as independent X-processes. Then the first coordinate of the pair is an X-process, and they are both X-processes up to the first time at which one of the transitions (5)–(10) takes place; at later times, it is possible that $|X'(t)| \neq |X(t)|$, so that the second coordinate has the wrong quadratic rates for an X-process. However, if X and X' are equal at any time, they have identical paths thereafter.

LEMMA 5.3. *Let $t_0 = 1/n$, (Y, Y') with $D(Y(t_0), Y'(t_0)) \le n\varepsilon$ and $t_1 = (\log n)^{-1}\sqrt{\varepsilon}$ be as for Lemma* 5.2, *and construct a bivariate process (X, X') on the same probability space as (Y, Y'), in such a way that $(X(t_0), X'(t_0)) = (Y(t_0), Y'(t_0))$ and that (X, X') shares all the jumps of (Y, Y') up to the first time ν at which a transition from* (5)–(10) *of* (5.13) *takes place*; *after that time, they evolve independently*. *Then*, *with this construction*,

$$
\mathbf{P}[\nu \le t_1] = O(\log n \sqrt{\varepsilon}), \tag{5.14}
$$

uniformly in $\varepsilon \le (\log n)^{-2}$ and $n \ge \exp(2(1 \vee C_1))$.

PROOF. From (4.2), the sum of the transition rates for jumps of types (5)–(10) is at most $2(c_1+C_1+c_2+c_3)D(Y,Y')$ before time ν. Hence

$$\mathbf{P}[\nu \le t_1] \le \mathbf{E}\left\{1 \wedge 2(c_1+C_1+c_2+c_3)\int_{t_0}^{t_1} D(Y(u),Y'(u))\,du\right\}$$

$$\le \mathbf{P}[B] + \mathbf{E}\left[\left\{1 \wedge 2(c_1+C_1+c_2+c_3) \times \int_{t_0}^{t_1} D(Y(u),Y'(u))\,du\right\} 1_{B^c}\right], \tag{5.15}$$

where

$$B := \bigcup_{t_0 \le t \le t_1} \left\{\frac{1}{t}\left|\frac{1}{|Y(t)|} - t\right| > \varphi\right\}, \tag{5.16}$$

and $\varphi = 1/\log n$ as for Lemma 5.2. Now $\mathbf{P}[B] = O(\varphi^{-2}t_1) = O(\log n\sqrt{\varepsilon})$ by Lemma 5.1, and on B^c we have $D(Y(u),Y'(u)) \le D'(u)$ for all $t_0 \le u \le t_1$, where D' is as constructed in the proof of Lemma 5.2. Hence the second term in (5.15) is no larger than

$$2(c_1+C_1+c_2+c_3)D\big(Y(t_0),Y'(t_0)\big) \times \int_{t_0}^{t_1} \{t_0/u\}^{1/(1+\varphi)}\,du\,\exp(C_1(t_1-t_0)) = O(\varepsilon\log n). \tag{5.17}$$

This completes the proof. □

COROLLARY 5.4. *If $\alpha,\alpha' \in \mathbf{Z}_+^K$ are such that $|\alpha| = |\alpha'| = n$ and $D(\alpha,\alpha') \le n\varepsilon$, then X-processes $\hat{X}$ and $\hat{X}'$ with $\hat{X}(1/n) = \alpha$ and $\hat{X}'(1/n) = \alpha'$ can be constructed on the same probability space, in such a way that $\hat{X}(t) = \hat{X}'(t)$ for all $t \ge t_1 = (\log n)^{-1}\sqrt{\varepsilon}$, with probability at least $1 - O(\log n\sqrt{\varepsilon})$, uniformly in $\varepsilon \le (\log n)^{-2}$ and $n \ge \exp(2(1 \vee C_1))$.*

PROOF. Combine Lemmas 5.2 and 5.3. Defining the event $E = \{\nu > t_1\} \cap \{Y(t) = Y'(t)$ for all $t \ge t_1\}$, a coupling of Y and Y' can be achieved in such a way that $\mathbf{P}[E] = 1 - O(\log n\sqrt{\varepsilon})$, and, on E, we have $\nu = \infty$ and $(X,X') = (Y,Y')$ for all $t \ge 1/n$. As soon as ν occurs, or at time t_1 if $Y(t_1) \ne Y'(t_1)$, continue both components independently as X-processes. □

Given $p \in \Delta_K$, these results can be collected to prove the existence of an X-process X^p which satisfies $\lim_{t\to 0}|X^p(t)| = \infty$ and $\lim_{t\to 0} X^p(t)/|X^p(t)| = p$. For each $m \ge 1$, let $X^{(m)}$ be an X-process started with $|X^{(m)}(1/m)| = m$ and with $|p^{(m)} - p| \le Km^{-1}$, where $p^{(m)} = m^{-1}X^{(m)}(1/m)$, and set $\psi(n) = n^{1/12}$,

$$\varepsilon(n) = K(2n^{-1} + n^{-1/2}\psi(n)). \tag{5.18}$$

THEOREM 5.5. *Define* $t_1 = (\log n)^{-1}\sqrt{\varepsilon(n)}$, $t_2 = \max\{t_1, n^{-1/6}\}$ *and* $u_n = t_2 + 1/(2n)$. *Then*

$$d_{TV}\{\mathscr{L}(X^{(m)}(s),\ s \geq u_n),\ \mathscr{L}(X^{(n)}(s),\ s \geq u_n)\} = O(n^{-1/6}), \tag{5.19}$$

uniformly in $m \geq n$. *Furthermore, as* $n \to \infty$, $X^{(n)}$ *converges weakly in* $D((0,\infty), \mathbf{Z}_+^K)$ *to a limit* X^p, *which satisfies* $\lim_{t\to 0} \rho(X^p(t)) = (\infty, p)$ *a.s.*

PROOF. From Corollary 4.6, for any $m > n$, it follows that

$$\left|p^{(m)} - n^{-1}X^{(m)}(\tau_n^{(m)})\right| \leq Kn^{-1/2}\psi(n), \tag{5.20}$$

except on a set of probability $O(\psi^{-2}(n))$. Define $\hat{X}^{(m)}$ on $s \geq 1/n$ by $\hat{X}^{(m)}(s) = X^{(m)}(s+\tau_n^{(m)}-1/n)$. Then, since also $|p^{(m)}-p^{(n)}| \leq 2Kn^{-1}$ uniformly in $m > n$, Corollary 5.4 shows that $\hat{X}^{(m)}$ and $X^{(n)}$ can be realized together in such a way that $\hat{X}^{(m)}(s) = X^{(n)}(s)$ for all $s \geq t_1$, except possibly on a set having probability $O(\log n\sqrt{\varepsilon(n)})$, and in this case we have $X^{(m)}(u_n) = X^{(n)}(u_n - \tau_n^{(m)} + 1/n)$, provided that $u_n - \tau_n^{(m)} + 1/n \geq t_1$. Now Corollary 4.4 shows that $|\tau_n^{(m)} - 1/n| \leq 1/(2n)$, which in its turn implies that

$$t_1 \leq t_2 \leq u_n - \tau_n^{(m)} + 1/n \leq t_2 + 1/n, \tag{5.21}$$

except on an event having probability $O(n^{-1})$, and Lemma 4.2 applied with $t = t_2$ combined with the Markov inequality shows that $\mathbf{P}[|X^{(n)}(t_2)| > n^{1/3}] = O(n^{-1/6})$; from this, it follows that

$$\begin{aligned} &\mathbf{P}\big[X^{(n)} \text{ does not remain constant on } [t_2, t_2 + 1/n]\big] \\ &\qquad = O\big(n^{-1/6} + n^{-1}n^{2/3}\big) = O(n^{-1/6}). \end{aligned} \tag{5.22}$$

Thus, using the joint realizations above,

$$\begin{aligned} &\mathbf{P}\big[X^{(m)}(u_n) \neq X^{(n)}(u_n)\big] \\ &\qquad = O\Big(n^{-1/6} + \psi^{-2}(n) + \log n\sqrt{\varepsilon(n)}\Big) = O(n^{-1/6}), \end{aligned} \tag{5.23}$$

and hence

$$d_{TV}\big(\mathscr{L}(X^{(m)}(u_n)),\ \mathscr{L}(X^{(n)}(u_n))\big) = O(n^{-1/6}). \tag{5.24}$$

However, for two Markov processes Z, Z' on a Polish space which have the same generator and satisfy $d_{TV}\big(\mathscr{L}(Z(s)), \mathscr{L}(Z'(s))\big) \leq \delta$, a simple coupling argument is enough to show that $d_{TV}\big(\mathscr{L}(Z(t), t \geq s), \mathscr{L}(Z'(t),\ t \geq s)\big) \leq \delta$ also, giving (5.19).

Now, for any $0 < t < T < \infty$, the space $D([t,T], \mathbf{Z}_+^K)$ is Polish, and hence it follows from (5.19) that $(X^{(n)}(s),\ t \leq s \leq T)$ converges weakly; by consistency, we can extend the limit X^p to the whole of $t > 0$, and (5.19) implies that

$$d_{TV}\{\mathscr{L}(X^p(s),\ s \geq u_n),\ \mathscr{L}(X^{(n)}(s),\ s \geq u_n)\} = O(n^{-1/6}). \tag{5.25}$$

Furthermore, given any $\varepsilon, \eta > 0$, it follows from Lemmas 4.3 and 4.5 that T can be chosen small enough to ensure that, for any $0 < v < T$ and for all n large enough,

$$\mathbf{P}\left[\sup_{v\le s\le T}\max_{1\le j\le K}\left|p_j^{(n)} - X_j^{(n)}(s)\ /|X^{(n)}(s)|\right| > \eta\right] < \varepsilon \tag{5.26}$$

and

$$\mathbf{P}\left[\sup_{v\le s\le T} 1/|X^{(n)}(s)| > \eta\right] < \varepsilon. \tag{5.27}$$

From this, it follows also that, for the same choice of T,

$$\mathbf{P}\left[\sup_{0<s\le T}\max_{1\le j\le K}\left|p_j - X_j^p(s)/|X^p(s)|\right| > \eta\right] \le \varepsilon \tag{5.28}$$

and

$$\mathbf{P}\left[\sup_{0<s\le T} 1/|X^p(s)| > \eta\right] \le \varepsilon, \tag{5.29}$$

and indeed the same choice of T can be used uniformly for all $p \in \Delta_K$. Hence, in particular, $\lim_{s\to 0}\rho(X^p(s)) = (\infty, p)$ a.s. as claimed. □

THEOREM 5.6. *Suppose that a sequence of X-processes $\widetilde{X}^{(m)}$ satisfy $|\widetilde{X}^{(m)}(0)| = s_m \to \infty$ and $|\tilde{p}^{(m)} - p| \to 0$ as $m\to\infty$, where $\tilde{p}^{(m)} = s_m^{-1}\widetilde{X}^{(m)}(0)$. Then, with $t_1 = (\log n)^{-1}\sqrt{\varepsilon(n)}$, $t_2 = \max\{t_1, n^{-1/6}\}$ and $u_n = t_2 + 1/(2n)$ as before,*

$$d_{TV}\{\mathscr{L}(\widetilde{X}^{(m)}(s),\ s\ge u_n),\ \mathscr{L}(X^{(n)}(s),\ s\ge u_n)\} = O(n^{-1/6}), \tag{5.30}$$

uniformly for all m such that $s_m > n$ and $|\tilde{p}^{(m)} - p| \le Kn^{-1}$. In particular, $\rho(\widetilde{X}^{(m)})$ converges weakly in $D([0,\infty), F)$ to Y^p as $m\to\infty$, where $Y^p(0) = (\infty, p)$ and $Y^p(s) = \rho(X^p(s))$ for all $s > 0$.

PROOF. The proof is much the same as for Theorem 5.5. Define $\bar{X}^{(m)}(t) = \widetilde{X}^{(m)}(t - 1/s_m)$ on $t \ge 1/s_m$, so that $|\bar{X}^{(m)}(1/s_m)| = s_m$ and, from Corollary 4.6, if $s_m > n$, then

$$\mathbf{P}\left[\left|\tilde{p}^{(m)} - n^{-1}\bar{X}^{(m)}(\bar{\tau}_n^{(m)})\right| \le Kn^{-1/2}\psi(n)\right] = 1 - O(\psi^{-2}(n)), \tag{5.31}$$

where $\bar{\tau}_n^{(m)}$ denotes the time at which $|\bar{X}^{(m)}|$ first takes the value n. Then, because $|\tilde{p}^{(m)} - p| \le Kn^{-1}$, it follows that $|\tilde{p}^{(m)} - p^{(n)}| \le 2Kn^{-1}$ also, and Corollary 5.4 can be applied to show that $\bar{X}^{(m)}$ and $X^{(n)}$ can be realized together in such a way that

$$\begin{aligned}&\mathbf{P}\left[\bar{X}^{(m)}\left(s + \bar{\tau}_n^{(m)} - 1/n\right) = X^{(n)}(s) \text{ for all } s \ge t_2\right]\\ &\qquad = 1 - O\left(\log n\sqrt{\varepsilon(n)}\right).\end{aligned} \tag{5.32}$$

However, $\bar{X}^{(m)}(s + \bar{\tau}_n^{(m)} - 1/n) = \widetilde{X}^{(m)}(s + \bar{\tau}_n^{(m)} - 1/n - 1/s_m)$, so that, from Corollary 4.4,

$$\begin{aligned} \mathbf{P}\big[\bar{X}^{(m)}(u_n) = X^{(n)}(u'_n) \text{ for some } u'_n \in [t_2 + 1/s_m, t_2 + 1/n + 1/s_m]\big] \\ = 1 - O(n^{-1/6}) \end{aligned} \tag{5.33}$$

and $\mathbf{P}\big[X^{(n)}$ is constant on $[t_2, t_2 + 1/n + 1/s_m]\big] = 1 - O(n^{-1/6})$ when $s_m > n$, as before. Hence (5.30) is satisfied, and $\widetilde{X}^{(m)}$ converges to X^p weakly in $D((0,\infty), \mathbf{Z}_+^K)$ as in Theorem 5.5. The convergence of $\rho(\widetilde{X}^{(m)})$ to Y^p in $D([0,\infty), F)$ now follows, in view of (5.28) and (5.29). □

THEOREM 5.7. *If $p^{[k]} \to p$ as $k \to \infty$, then $X^{p^{[k]}}$ converges weakly to X^p in $D((0,\infty), \mathbf{Z}_+^K)$.*

PROOF. We show that

$$d_{TV}\big\{\mathscr{L}(X^{p^{[k]}}(s),\, t \le s \le T),\, \mathscr{L}(X^p(s),\, t \le s \le T)\big\} \to 0 \tag{5.34}$$

as $k \to \infty$ for each $0 < t < T < \infty$, which is enough, in view of (5.28) and (5.29). For any $n > 0$, choose k_n so large that $|p^{[k]} - p| \le Kn^{-1/2}\psi(n)$ for all $k \ge k_n$, with $\psi(n) = n^{1/12}$ and $\varepsilon(n)$ as for Theorem 5.5. Then, if $\tau_n^{[k]}$ denotes the first time that $|X^{p^{[k]}}| = n$ and τ_n the first time that $|X^p| = n$, it follows from Corollary 4.6 and Theorem 5.5 that, for $k \ge k_n$,

$$\mathbf{P}\big[n^{-1}|X^{p^{[k]}}(\tau_n^{[k]}) - X^p(\tau_n)| > Kn^{-1/2}\psi(n)\big] = O(n^{-1/6}). \tag{5.35}$$

Thus, from Corollary 5.4, whenever $k \ge k_n$, $X^{p^{[k]}}$ and X^p can be realized together in such a way that they are identical from time $\max\{(\log n)^{-1}\sqrt{\varepsilon(n)}, n^{-1/6}\} + n^{-1}$ onward, with probability at least $1 - O(n^{-1/6})$, proving (5.34). □

6. The entrance process: θ-continuity. Now we suppose that we have a family of X-processes, indexed by $\theta \in \mathbf{R}_+^K$, defined as in (4.1) and (4.2). Denote their transition rates $q(\alpha, \beta; \theta)$ by

$$\begin{aligned} q(\alpha, \alpha - \varepsilon^j; \theta) &= \alpha_j|\alpha| + f_{1j}^{(\theta)}(\alpha), \\ q(\alpha, \alpha + \varepsilon^j; \theta) &= f_{2j}^{(\theta)}(\alpha), \\ q(\alpha, \alpha + \varepsilon^i + \varepsilon^j; \theta) &= f_{3ij}^{(\theta)}(\alpha), \end{aligned} \tag{6.1}$$

and $q(\alpha, \beta; \theta) = 0$ for all other $\beta \ne \alpha$, and let the corresponding probabilities be denoted by $\mathbf{P}^{(\theta)}$; set

$$\begin{aligned} P_{\alpha\beta}(t;\theta) &= \mathbf{P}^{(\theta)}[X(t) = \beta \mid X(0) = \alpha]; \\ b_\beta(t, p; \theta) &= \mathbf{P}^{(\theta)}[X^p(t) = \beta]. \end{aligned} \tag{6.2}$$

Fix any $\theta_0 \in \mathbf{R}_+^K$ and assume that:

1. The functions $q(\alpha, \beta; \cdot)$ are continuous at θ_0 for each $\alpha, \beta \in \mathbf{Z}_+^K$.
2. There exists a neighborhood N_0 of θ_0 such that $\bar{c}_l := \sup_{\theta \in N_0} c_l^{(\theta)} < \infty$ for $l = 1, 2, 3$ and $\bar{C}_1 := \sup_{\theta \in N_0} C_1^{(\theta)} < \infty$.

LEMMA 6.1. *Under Assumption* 1, $P_{\alpha\beta}(t; \cdot)$ *is continuous at* θ_0 *for any* $t > 0$ *and* $\alpha, \beta \in \mathbf{Z}_+^K$.

PROOF. By Example 1.1 of Xia (1994), $\theta \mapsto \mathbf{P}^{(\theta)}[X \in \cdot \mid X(0) = \alpha]$ is continuous at θ_0 for each $\alpha \in \mathbf{Z}_+^K$, so if we can show that

$$\mathbf{P}^{(\theta_0)}[X(t) \neq X(t-) \mid X(0) = \alpha] = 0, \qquad \alpha \in \mathbf{Z}_+^K, t > 0, \tag{6.3}$$

then the conclusion of the lemma will follow. To see (6.3), let $0 < \tau_1 < \tau_2 < \cdots$ denote the jump times of X. Then

$$\begin{aligned} &\mathbf{P}^{(\theta_0)}\big[\tau_n = t \mid X(0) = \alpha\big] \\ &\quad = \sum_{\alpha_1, \dots, \alpha_n} \mathbf{P}^{(\theta_0)}\big[X(\tau_1) = \alpha_1, \dots, X(\tau_n) = \alpha_n \mid X(0) = \alpha\big] \\ &\qquad\qquad \times \mathbf{P}^{(\theta_0)}\big[\tau_n = t \mid X(0) = \alpha, X(\tau_1) = \alpha_1, \dots, X(\tau_n) = \alpha_n\big] \\ &\quad = 0, \end{aligned} \tag{6.4}$$

using the fact that the conditional distribution of τ_n, given $X(0) = \alpha$, $X(\tau_1) = \alpha_1, \dots, X(\tau_n) = \alpha_n$, is the distribution of the sum of n independent exponential random variables. Finally, the sum of the probabilities on the left side of (6.4) is equal to the probability on the left side of (6.3). □

THEOREM 6.2. *Under assumptions* 1 *and* 2, $b_\beta(t, p; \cdot)$ *is continuous at* θ_0, *for any* $t > 0$, $p \in \Delta_K$ *and* $\beta \in \mathbf{Z}_+^K$.

PROOF. Fix $t > 0$, $p \in \Delta_K$ and $\beta \in \mathbf{Z}_+^K$. First note that, by assumption 2, all the order estimates in Sections 4 and 5 are uniform in $\theta \in N_0$. Thus in particular, from (5.19), given any $\eta > 0$, we can fix $n = n(\eta)$ such that

$$d_{TV}\big(\mathscr{L}^{(\theta)}(X^p(t)), \mathscr{L}^{(\theta)}(X^{(n)}(t))\big) \le \eta/3, \tag{6.5}$$

uniformly in $\theta \in N_0$, from which it follows that

$$|b_\beta(t, p; \theta) - P_{np^{(n)}, \beta}(t - n^{-1}; \theta)| \le \eta/3 \quad \text{for all } \theta \in N_0. \tag{6.6}$$

Then, by Lemma 6.1, there exists a neighborhood N_η of θ_0 such that

$$|P_{np^{(n)}, \beta}(t - n^{-1}; \theta) - P_{np^{(n)}, \beta}(t - n^{-1}; \theta_0)| \le \eta/3 \tag{6.7}$$

for all $\theta \in N_\eta$. Hence, for all $\theta \in N_0 \cap N_\eta$, we have $|b_\beta(t, p; \theta) - b_\beta(t, p; \theta_0)| \leq \eta$, proving that $b_\beta(t, p; \cdot)$ is continuous at θ_0. □

7. Ray–Knight theory. The following sketch of Ray–Knight theory in the context of countable-state Markov pure jump processes is based on Sections 81 and 57 of Williams (1979). Let I be a countably infinite set representing the state space of the Markov process. Let $p_{ij}(t)$ be the transition probabilities and

$$R_\lambda(i, j) = \int_0^\infty e^{-\lambda t} p_{ij}(t)\, dt \tag{7.1}$$

the resolvent. Note that $R_1(i, \cdot)$ belongs to $l^1(I)$ for each $i \in I$. Indeed,

$$\|R_1(i, \cdot)\|_1 = \sum_{j \in I} \int_0^\infty e^{-t} p_{ij}(t)\, dt = \int_0^\infty e^{-t}\, dt = 1. \tag{7.2}$$

Let E be the closure in $l^1(I)$ of the set of functions $\{R_1(i, \cdot): i \in I\}$. Then E is the (relevant part of the) Ray–Knight compactification of I, and the map $i \mapsto R_1(i, \cdot)$ embeds I into E.

Now, suppose we specify a particular compactification F of I. How do we show that F is the Ray–Knight compactification of I, in the sense of being homeomorphic to E?

LEMMA 7.1. *Define E as above to be the closure in $l^1(I)$ of $\{R_1(i, \cdot): i \in I\}$. Let F be a compact space and suppose that ρ: $I \mapsto F$ is a one-to-one map with the property that $\rho(I)$ is dense in F and its relative topology is the discrete topology. Assume the following conditions*:

(i) *For each $x \in F - \rho(I)$, there exists $R(x, \cdot) \in E \subset l^1(I)$ such that, if $\{i_n\} \subset I$ and $\rho(i_n) \to x$ in F, then $R_1(i_n, \cdot) \to R(x, \cdot)$ in E.*

(ii) *$x \mapsto R(x, \cdot)$ is a continuous map from $F - \rho(I)$ into E.*

(iii) *The map Λ: $F \mapsto E$ defined by $\Lambda(\rho(i)) = R_1(i, \cdot)$ for $i \in I$ and $\Lambda(x) = R(x, \cdot)$ for $x \in F - \rho(I)$ is one-to-one.*

Then Λ is a homeomorphism from F onto E.

PROOF. Conditions (i) and (ii) show that Λ is continuous. It is one-to-one by (iii) and onto by the definition of E, the compactness of F and (i). Finally, the continuity of Λ^{-1} is now automatic from the compactness of F, and the conclusion follows. □

We wish to apply Lemma 7.1 to the case of the X-process with transition rates satisfying (4.1)–(4.3). Denote its transition probabilities by $P_{\alpha\beta}(t)$, put $I = \mathbf{Z}_+^K$ and let ρ and F be as in (1.12) and (1.13).

LEMMA 7.2. *Suppose that, for each* $p \in \Delta_K$, *there exists a Markov process* $\{Y(t),\ t \ge 0\}$ *in* F *starting at* (∞, p), *and assume the following properties*:

(a) $\mathbf{P}_{(\infty, p)}[Y(t) \in \rho(\mathbf{Z}_+^K)] = 1$ *for all* $p \in \Delta_K$ *and* $t > 0$.

(b) $P_{\alpha\beta}(t) \to \mathbf{P}_{(\infty, p)}[Y(t) = \rho(\beta)]$ *as* $\rho(\alpha) \to (\infty, p)$ *for all* $\beta \in \mathbf{Z}_+^K$, $t > 0$ *and* $p \in \Delta_K$.

(c) $\mathbf{P}_{(\infty, p)}[Y(t) = \rho(\beta)]$ *is continuous in* $p \in \Delta_K$ *for all* $t > 0$ *and* $\beta \in \mathbf{Z}_+^K$.

Then the Ray–Knight compactification of I *with respect to the* X-*process is homeomorphic to* F, *and the process* $\{Y(t),\ t \ge 0\}$ *above is the corresponding Ray process*.

PROOF. Define

$$R((\infty, p), \beta) = \int_0^\infty e^{-t} \mathbf{P}_{(\infty, p)}[Y(t) = \rho(\beta)]\, dt. \tag{7.3}$$

Then properties (a)–(c) imply conditions (i) and (ii) of Lemma 7.1, using the fact that if $\{f_n\} \subset l^1(I)$, $f \in l^1(I)$, $\|f_n\|_1 = 1$ for all n and $\|f\|_1 = 1$, then $f_n \to f$ pointwise on I implies $f_n \to f$ in $l^1(I)$.

Condition (iii) can be seen as follows. First, let $\alpha, \alpha' \in \mathbf{Z}_+^K$ and suppose $R_1(\alpha, \cdot) = R_1(\alpha', \cdot)$. Then, for every bounded function g on $\mathbf{Z}_+^K$,

$$\sum_{\beta \in \mathbf{Z}_+^K} \int_0^\infty e^{-t} g(\beta) P_{\alpha\beta}(t)\, dt = \sum_{\beta \in \mathbf{Z}_+^K} \int_0^\infty e^{-t} g(\beta) P_{\alpha'\beta}(t)\, dt, \tag{7.4}$$

or, letting $\mathscr{A}_0$ denote the generator of the process,

$$(1 - \mathscr{A}_0)^{-1} g(\alpha) = (1 - \mathscr{A}_0)^{-1} g(\alpha'). \tag{7.5}$$

Taking $g = (1 - \mathscr{A}_0) f$ for $f \in \mathscr{D}(\mathscr{A}_0)$, we conclude that

$$f(\alpha) = f(\alpha'), \qquad f \in \mathscr{D}(\mathscr{A}_0). \tag{7.6}$$

Since the space of functions on $\mathbf{Z}_+^K$ with finite support separates points of $\mathbf{Z}_+^K$, Λ is one-to-one on $\rho(\mathbf{Z}_+^K)$.

Next, if $\alpha \in \mathbf{Z}_+^K$ and $p \in \Delta_K$ were to satisfy $R_1(\alpha, \cdot) = R((\infty, p), \cdot)$, a similar argument would show that

$$f(\alpha) = \lim_{\rho(\alpha') \to (\infty, p)} f(\alpha'), \qquad f \in \mathscr{D}(\mathscr{A}_0), \tag{7.7}$$

which clearly fails. Finally, let $p, p' \in \Delta_K$ and suppose that $R((\infty, p), \cdot) = R((\infty, p'), \cdot)$. In this case we find that

$$\lim_{\rho(\alpha) \to (\infty, p)} f(\alpha) = \lim_{\rho(\alpha') \to (\infty, p')} f(\alpha'), \qquad f \in \mathscr{D}(\mathscr{A}_0). \tag{7.8}$$

Given $\beta \in \mathbf{Z}_+^K$, define $f(\alpha) = \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_K}{\beta_K} / \binom{|\alpha|}{|\beta|}$ if $\alpha \ge \beta$ and $f(\alpha) = 0$ otherwise. Noting that $f \in \mathscr{D}(\mathscr{A}_0)$, (7.8) becomes

$$\binom{|\beta|}{\beta} p^\beta = \binom{|\beta|}{\beta} (p')^\beta. \tag{7.9}$$

Since monomials separate points of Δ_K, we conclude that Λ is one-to-one on F. $\square$

THEOREM 7.3. *The Ray–Knight compactification of $\mathbf{Z}_+^K$ with respect to the X-process with transition rates given either by* (3.7) *and* (3.8) *or by* (3.13)–(3.15) *is homeomorphic to F defined in* (1.13), *and the corresponding Ray process is, but for a rescaling of time, the Y^p of Theorem* 5.6.

PROOF. Theorems 5.5, 5.6 and 5.7 show that properties (a)–(c) of Lemma 7.2 are satisfied by all X-processes with rates specified as in (4.1) and (4.2). All that remains is to check that twice the rates specified either in (3.7) and (3.8) or in (3.13)–(3.15) give rise to f-functions satisfying (4.2) and (4.3). The necessary argument is given in Lemma 8.2. $\square$

8. Application to the K-allele model (continued). In this section we show that the assumptions made in Sections 4–6 are satisfied by our examples (3.7) and (3.8) and (3.13)–(3.15). This then will finally complete the proof of the transition function expansion, as described in Theorem 3.1.

LEMMA 8.1. *Let (σ_{ij}) be a real symmetric $K \times K$ matrix, and define the function c: $\mathbf{R}_+^K - \{0\} \mapsto [\exp(\min \sigma_{kl}), \exp(\max \sigma_{kl})]$ by* (3.12). (*In the haploid case* (1.19), *this reduces to* (1.21).) *Then*

$$\left| \frac{c(w)}{c(w+\varepsilon^i)} - 1 \right| = O((1+|w|)^{-1}) \tag{8.1}$$

and

$$\left| \frac{c(w+\varepsilon^i)}{c(w)} - \frac{c(w+\varepsilon^i+\varepsilon^j)}{c(w+\varepsilon^j)} \right| = O((1+|w|)^{-2}) \tag{8.2}$$

for $i, j = 1, \dots, K$, uniformly in w.

PROOF. By (3.12), we have

$$c(w) = \mathbf{E}\exp\left\{ \sum_{i,j=1}^{K} \sigma_{ij} \frac{W_i W_j}{|W|^2} \right\} = \exp(\min \sigma_{kl}) \mathbf{E}\{\exp(A(W))\}, \tag{8.3}$$

where $W_1, \dots, W_K$ are independent random variables with W_i being gamma$(w_i, 1)$ distributed, $W := (W_1, \dots, W_K)$ and $A(W) := \sum_{i,j=1}^K \sigma_{ij}^+ (W_i W_j/|W|^2)$ [recall (3.10)]. Of course, $|W| := W_1 + \cdots + W_K$. Consider the function $\hat{c}(w) := c(w)\exp(-\min \sigma_{kl})$. Letting $B_i(W) = \sum_{j=1}^K \sigma_{ij}^+ (W_j/|W|)$, we can write

$$\begin{aligned} \hat{c}(w+\varepsilon^i) = \mathbf{E}\exp\biggl\{ & A(W)\frac{|W|^2}{(|W|+Z)^2} + B_i(W)\frac{2|W|Z}{(|W|+Z)^2} \\ & + \sigma_{ii}^+ \frac{Z^2}{(|W|+Z)^2} \biggr\} \end{aligned} \tag{8.4}$$

and

$$
\begin{aligned}
&\hat{c}(w+\varepsilon^i+\varepsilon^j)\\
&\quad= \mathbf{E}\exp\bigg\{A(W)\frac{|W|^2}{(|W|+Z_1+Z_2)^2}+B_i(W)\frac{2|W|Z_1}{(|W|+Z_1+Z_2)^2}\\
&\qquad\qquad +B_j(W)\frac{2|W|Z_2}{(|W|+Z_1+Z_2)^2}+\sigma^+_{ii}\frac{Z_1^2}{(|W|+Z_1+Z_2)^2}\\
&\qquad\qquad +\sigma^+_{ij}\frac{2Z_1Z_2}{(|W|+Z_1+Z_2)^2}+\sigma^+_{jj}\frac{Z_2^2}{(|W|+Z_1+Z_2)^2}\bigg\}
\end{aligned}
\tag{8.5}
$$

for $i, j = 1, \dots, K$, where Z, Z_1 and Z_2 are independent of each other and of $W_1, \dots, W_K$ and have the gamma(1, 1) distribution. Now we always have $0 \le A(W) \le \overline{\sigma}$ and $0 \le B_i(W) \le \overline{\sigma}$, and hence, defining

$$
\hat{c}_1(w,i) = \mathbf{E}\exp\bigg\{A(W)\Big(1-\frac{2Z}{|W|+Z}\Big)+B_i(W)\frac{2Z}{|W|+Z}\bigg\},
\tag{8.6}
$$

it follows from the inequality $|e^x-1| \le |x|e^{|x|}$ that

$$
\begin{aligned}
&\big|\hat{c}(w+\varepsilon^i)-\hat{c}_1(w,i)\big|\\
&\quad\le \mathbf{E}\bigg\{\exp\bigg\{A(W)\Big(1-\frac{2Z}{|W|+Z}\Big)+B_i(W)\frac{2Z}{|W|+Z}\bigg\}\\
&\qquad\quad \times\bigg|\exp\bigg\{\{A(W)-2B_i(W)+\sigma^+_{ii}\}\frac{Z^2}{(|W|+Z)^2}\bigg\}-1\bigg|\bigg\}\\
&\quad\le e^{2\overline{\sigma}}2\overline{\sigma}\mathbf{E}\bigg\{\frac{Z^2}{(|W|+Z)^2}\bigg\}e^{2\overline{\sigma}} = O((1+|w|)^{-2});
\end{aligned}
\tag{8.7}
$$

in similar fashion, defining

$$
\hat{c}_2(w,i) = \mathbf{E}\bigg\{e^{A(W)}\Big(1-\frac{2Z}{|W|+Z}\{A(W)-B_i(W)\}\Big)\bigg\},
\tag{8.8}
$$

the inequality $|e^x-1-x| \le \frac{1}{2}x^2e^{|x|}$ implies that

$$
\hat{c}_1(w,i)-\hat{c}_2(w,i) = O((1+|w|)^{-2});
\tag{8.9}
$$

of course,

$$
\hat{c}_2(w,i) = \hat{c}(w)-h_i(w),
\tag{8.10}
$$

where

$$h_i(w) := 2\mathbf{E}\left\{e^{A(W)}\frac{Z}{|W|+Z}\{A(W)-B_i(W)\}\right\} \\ = O\big((1+|w|)^{-1}\big). \tag{8.11}$$

Adding (8.7), (8.9) and (8.10) yields

$$\hat{c}(w+\varepsilon^i) = \hat{c}(w) - h_i(w) + O((1+|w|)^{-2}), \tag{8.12}$$

which, with (8.11), completes the proof of (8.1).

For (8.2), analogously to (8.6) and (8.7), we deduce that

$$\begin{aligned}\hat{c}(w+\varepsilon^i+\varepsilon^j) = \mathbf{E}\exp\bigg\{&A(W)\bigg(1-\frac{2Z_1}{|W|+Z_1}-\frac{2Z_2}{|W|+Z_2}\bigg)\\ &+B_i(W)\frac{2Z_1}{|W|+Z_1}+B_j(W)\frac{2Z_2}{|W|+Z_2}\bigg\}\\ &+O\big((1+|w|)^{-2}\big)\end{aligned} \tag{8.13}$$

for $i, j = 1, \dots, K$, and then proceed as before to obtain

$$\hat{c}(w+\varepsilon^i+\varepsilon^j) = \hat{c}(w) - h_i(w) - h_j(w) + O((1+|w|)^{-2}). \tag{8.14}$$

Combining (8.14) with (8.11) and (8.12) then proves (8.2). □

LEMMA 8.2. *In the haploid case* (1.19), *let* $(q(\alpha,\beta))$ *be as in* (3.7) *and* (3.8) *and define* $f_{1i}(\alpha) = 2q(\alpha,\alpha-\varepsilon^i)-\alpha_i|\alpha|$, $f_{2i}^{(1)}(\alpha) = 2q(\alpha,\alpha+\varepsilon^i)$ *and* $f_{3ij}(\alpha)=0$ *for* $i, j = 1, \dots, K$. *In general, let* $(q(\alpha,\beta))$ *be as in* (3.13)–(3.15) *and define* $f_{1i}(\alpha) = 2q(\alpha,\alpha-\varepsilon^i)-\alpha_i|\alpha|$, $f_{2i}^{(2)}(\alpha) = 2q(\alpha,\alpha+\varepsilon^i)$ *and* $f_{3ij}(\alpha) = 2q(\alpha,\alpha+\varepsilon^i+\varepsilon^j)$ *for* $i, j = 1, \dots, K$. *Then* $f_{1i}(0) = f_{2i}^{(1)}(0) = f_{2i}^{(2)}(0) = f_{3ij}(0) = 0$,

$$\big|f_{1i}(\alpha) - f_{1i}(\alpha')\big| \le c_{1i}(1+|\theta|)|\alpha-\alpha'|, \tag{8.15}$$

$$\big|f_{2i}^{(l)}(\alpha) - f_{2i}^{(l)}(\alpha')\big| \le c_{2i}^{(l)}|\alpha-\alpha'|, \tag{8.16}$$

$$\big|f_{3ij}(\alpha) - f_{3ij}(\alpha')\big| \le c_{3ij}|\alpha-\alpha'|, \tag{8.17}$$

for $i, j = 1, \dots, K$, $l = 1, 2$, *and all* $\alpha \in \mathbf{Z}_+^K$, *where the constants* c_{1i}, $c_{2i}^{(l)}$ *and* c_{3ij}, *when regarded as functions of* θ, *are uniformly bounded in* θ. *Furthermore, there exists* $C_1 \ge 0$ *such that* $f_{1i}(\alpha) \ge -C_1\alpha_i$ *for* $i = 1, \dots, K$ *and all* $\alpha \in \mathbf{Z}_+^K$.

PROOF. It suffices for (8.15)–(8.17) to consider $\alpha' = \alpha+\varepsilon^j$ for $j = 1, \dots, K$. For $f_{2i}^{(1)}$ write

$$a_1(\alpha) = |\alpha|/(|\alpha|+|\theta|); \qquad a_2(\alpha) = \alpha_i+\theta_i; \\ a_3(\alpha) = c(\alpha+\varepsilon^i+\theta)/c(\alpha+\theta); \tag{8.18}$$

then it is immediate that, if $\sigma_i^- \neq 0$,

$$
\begin{aligned}
&(2\sigma_i^-)^{-1}\{f_{2i}^{(1)}(\alpha) - f_{2i}^{(1)}(\alpha+\varepsilon^j)\} \\
&\qquad = \big[a_1(\alpha) - a_1(\alpha+\varepsilon^j)\big]a_2(\alpha)a_3(\alpha) \\
&\qquad\quad + a_1(\alpha+\varepsilon^j)\big[a_2(\alpha) - a_2(\alpha+\varepsilon^j)\big]a_3(\alpha) \\
&\qquad\quad + a_1(\alpha+\varepsilon^j)a_2(\alpha+\varepsilon^j)\big[a_3(\alpha) - a_3(\alpha+\varepsilon^j)\big].
\end{aligned} \tag{8.19}
$$

Noting that $|a_3(\alpha)| \le e^{\overline{\sigma}}$, the first term is bounded by

$$
\frac{|\theta|(\alpha_i+\theta_i)}{(|\alpha|+1+|\theta|)(|\alpha|+|\theta|)}\, e^{\overline{\sigma}} \le e^{\overline{\sigma}}; \tag{8.20}
$$

the second is also bounded by $e^{\overline{\sigma}}$, and the last is

$$
\frac{|\alpha|+1}{|\alpha|+1+|\theta|}(\alpha_i+\delta_{ij}+\theta_i)O((|\alpha|+1+|\theta|)^{-2}) \tag{8.21}
$$

by (8.2), which is bounded in α, uniformly in θ. The treatment of $f_{2i}^{(2)}$ is much the same.

In the corresponding argument for

$$
f_{1i}(\alpha) = \alpha_i(|\theta|-1)\frac{c(\alpha-\varepsilon^i+\theta)}{c(\alpha+\theta)} + \alpha_i|\alpha|\left\{\frac{c(\alpha-\varepsilon^i+\theta)}{c(\alpha+\theta)} - 1\right\}, \tag{8.22}
$$

note that $f_{1i}(\alpha) = 0$ if $\alpha_i = 0$, in which case $f_{1i}(\alpha) - f_{1i}(\alpha+\varepsilon^j) = 0$ except when $j = i$; then, from (8.1),

$$
\begin{aligned}
&\big|f_{1i}(\alpha) - f_{1i}(\alpha+\varepsilon^i)\big| \\
&\quad = \left|(|\theta|-1)\frac{c(\alpha+\theta)}{c(\alpha+\varepsilon^i+\theta)} + (|\alpha|+1)\left\{\frac{c(\alpha+\theta)}{c(\alpha+\varepsilon^i+\theta)} - 1\right\}\right| \\
&\quad \le (|\theta|\vee 1)e^{\overline{\sigma}} + (|\alpha|+1)O\big((|\alpha|+1+|\theta|)^{-1}\big).
\end{aligned} \tag{8.23}
$$

If $\alpha_i \ge 1$, consider the two terms in $f_{1i}(\alpha)$ separately. For the first term, take $a_1(\alpha) = \alpha_i$, $a_2(\alpha) = |\theta|-1$ and $a_3(\alpha) = c(\alpha-\varepsilon^i+\theta)/c(\alpha+\theta)$ in (8.19), and argue using (8.2) to obtain a bound $O(|\theta|\vee 1)$; for the second term, take $a_1(\alpha) = \alpha_i$, $a_2(\alpha) = |\alpha|$ and $a_3(\alpha) = c(\alpha-\varepsilon^i+\theta)/c(\alpha+\theta) - 1$ in (8.19), and use (8.1) and (8.2). It also follows from (8.22) that $f_{1i}(\alpha) \ge -C_1\alpha_i$, with

$$
C_1 := \left\{|\theta| - 1 + \sup_{w\in\mathbf{R}_+^K-\{0\}}\ \max_{1\le i\le K} |w|\Big(1 - \frac{c(w)}{c(w+\varepsilon^i)}\Big)\right\} < \infty, \tag{8.24}
$$

by (8.1).

The argument for $|f_{3ij}(\alpha) - f_{3ij}(\alpha')|$ runs along the same lines. Define $a_1(\alpha) = \alpha_i + \delta_{ij} + \theta_i$, $a_2(\alpha) = (|\alpha|+1+|\theta|)^{-1}$, $a_3(\alpha) = c(\alpha+\varepsilon^i+\varepsilon^j+\theta)/c(\alpha+\theta)$, $a_4(\alpha) = |\alpha|/(|\alpha|+|\theta|)$ and $a_5(\alpha) = \alpha_j + \theta_j$; then, for example,

$$
\begin{aligned}
&a_1(\alpha+\varepsilon^k)a_2(\alpha+\varepsilon^k)|a_3(\alpha) - a_3(\alpha+\varepsilon^k)|a_4(\alpha)a_5(\alpha) \\
&\qquad = \frac{\alpha_i + \delta_{ij} + \delta_{ik} + \theta_i}{|\alpha| + 2 + |\theta|} O((|\alpha|+1+|\theta|)^{-2})(\alpha_j + \theta_j)
\end{aligned} \tag{8.25}
$$

and

$$
\begin{aligned}
&a_1(\alpha+\varepsilon^k)|a_2(\alpha) - a_2(\alpha+\varepsilon^k)|a_3(\alpha)a_4(\alpha)a_5(\alpha) \\
&\qquad \le \frac{(\alpha_i + \delta_{ij} + \delta_{ik} + \theta_i)e^{\overline{\sigma}}(\alpha_j + \theta_j)}{(|\alpha|+2+|\theta|)(|\alpha|+1+|\theta|)};
\end{aligned} \tag{8.26}
$$

both terms are bounded in α, uniformly in θ. This proves the lemma. □

Acknowledgement. A.D.B. and S.N.E. thank the Department of Mathematics and Statistics at Monash University for their hospitality during the preparation of this paper.

REFERENCES

DONNELLY, P. and KURTZ, T. G. (1996). A countable representation of the Fleming–Viot measure-valued diffusion. *Ann. Probab.* **24** 698–742.

DONNELLY, P. and KURTZ, T. G. (1999). Genealogical processes for Fleming–Viot models with selection and recombination. *Ann. Appl. Probab.* To appear.

ETHIER, S. N. and GRIFFITHS, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Probab.* **21** 1571–1590.

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

GRIFFITHS, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.* **11** 310–325.

GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theoret. Population Biol.* **17** 37–50.

GRIFFITHS, R. C. and LI, W.-H. (1983). Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. *Theoret. Population Biol.* **32** 19–33.

KIMURA, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proc. Nat. Acad. Sci. U.S.A.* **41** 144–150.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28** 882–901.

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

KRONE, S. M. and NEUHAUSER, C. (1997). Ancestral processes with selection. *Theoret. Population Biol.* **51** 210–237.

LI, Z., SHIGA, T. and YAO, L. (1999). A reversibility problem for Fleming–Viot diffusion processes. *Electronic Comm. Probab.* **4** 65–76.

LINDVALL, T. (1992). *Lectures on the Coupling Method*. Wiley, New York.

OVERBECK, L. and RÖCKNER, M. (1997). Geometric aspects of finite- and infinite-dimensional Fleming-Viot processes. *Random Oper. Stochastic Equations* **5** 35–58.

SATO, K. (1978). Diffusion operators in population genetics and convergence of Markov chains. *Measure Theory Applications to Stochastic Analysis. Lecture Notes in Math.* **695** 127–137. Springer, Berlin.

SHIGA, T. (1981). Diffusion processes in population genetics. *J. Math. Kyoto Univ.* **21** 133–151.
SHIMAKURA, N. (1977). Equations différentielles provenant de la génétique des populations. *Tôhoku Math. J.* **29** 287–318.
TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoret. Population Biol.* **26** 119–164.
WILLIAMS, D. (1979). *Diffusions, Markov Processes, and Martingales* **1**. Wiley, Chichester.
WRIGHT, S. (1949). Adaptation and selection. In *Genetics, Paleontology, and Evolution* (G. L. Jepson, E. Mayr and G. G. Simpson, eds.) 365–389. Princeton Univ. Press.
XIA, A. (1994). Weak convergence of Markov processes with extended generators. *Ann. Probab.* **22** 2183–2202.

A. D. BARBOUR
INSTITUTE FOR MATHEMATICS
UNIVERSITY OF ZÜRICH
WINTERTHURERSTRASSE 190
CH–8057 ZÜRICH
SWITZERLAND
E-MAIL: adb@amath.unizh.ch

S. N. ETHIER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
155 S. 1400 E., JWB 233
SALT LAKE CITY, UTAH 84112
E-MAIL: ethier@math.utah.edu

R. C. GRIFFITHS
DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
E-MAIL: griffiths@stats.ox.ac.uk