CONVERGENCE RESULTS ON MULTITYPE, MULTIVARIATE BRANCHING RANDOM WALKS

J. D. BIGGINS,* *The University of Sheffield*

A. RAHIMZADEH SANI,** *Teacher Training University of Tehran*

Abstract

We consider a multitype branching random walk on d-dimensional Euclidian space. The uniform convergence, as n goes to infinity, of a scaled version of the Laplace transform of the point process given by the nth generation particles of each type is obtained. Similar results in the one-type case, where the transform gives a martingale, were obtained in Biggins (1992) and Barral (2001). This uniform convergence of transforms is then used to obtain limit results for numbers in the underlying point processes. Supporting results, which are of interest in their own right, are obtained on (i) 'Perron–Frobenius theory' for matrices that are smooth functions of a variable $\lambda \in L$ and are nonnegative when $\lambda \in L_- \subset L$, where L is an open set in $\mathbb{C}^d$, and (ii) saddlepoint approximations of multivariate distributions. The saddlepoint approximations developed are strong enough to give a refined large deviation theorem of Chaganty and Sethuraman (1993) as a by-product.

Keywords: Large deviation; local limit theorem; saddlepoint approximation; uniform convergence; multitype branching random walk; Laplace transform; martingale; Perron–Frobenius; nonnegative matrix

2000 Mathematics Subject Classification: Primary 60J80
Secondary 15A48; 60F10

1. Introduction

We consider the multitype branching random walk on the d-dimensional Euclidian space, $\mathbb{R}^d$. The process starts with a single particle located at the origin. This particle produces daughter particles, which are scattered in $\mathbb{R}^d$, to give the first generation; these first-generation particles produce daughter particles to give the second generation; and so on. As usual in branching processes, the nth-generation particles reproduce independently of each other. Each particle in this process is of one of p types; the set of possible types is identified with $\{1, \ldots, p\}$. For each $i \in \{1, \ldots, p\}$ there is a vector of point processes $(Z_{i1}, Z_{i2}, \ldots, Z_{ip})$. Then, when a particle of type i reproduces, the positions of its daughter particles of the various types, relative to the parent's position, are given by a copy of $(Z_{i1}, Z_{i2}, \ldots, Z_{ip})$. The one-type branching random walk has received extensive treatment in the literature. The multitype extension has received less attention, but discussion of it can be found in Mode (1971), Biggins (1976), (1996), Bramson *et al.* (1992), and Kyprianou and Rahimzadeh Sani (2001).

We reserve the notation i, j, and k, drawn from $\{1, \ldots, p\}$, for particle types. Let μ_{ij} be the intensity measure of Z_{ij}. We assume throughout that

$$\text{there exists a } \boldsymbol{\vartheta} \in \mathbb{R}^d \text{ such that } \max_{i,j} \int_{\mathbb{R}^d} e^{-\boldsymbol{\vartheta}^\top \boldsymbol{x}} \mu_{ij}(\mathrm{d}\boldsymbol{x}) < \infty, \tag{1}$$

where $\boldsymbol{\vartheta}^\top \boldsymbol{x} = \sum_i \vartheta_i x_i$ is the usual inner product of vectors. This condition is enough to ensure that convolutions of the μ_{ij} produce well-defined measures.

Technically, the description of the process outlines how to produce the probability measure on the space of trajectories, given the type of the single initial ancestor. There are p such measures, one for each possible initial type. Suppose that the initial ancestor has type i and let Z_{ij}^n be the point process giving the resulting positions of the type-j particles in generation n; then Z_{ij}^1 is distributed like Z_{ij}. The first objective here is to obtain the asymptotic behaviour, as $n \to \infty$, of the Laplace transform of Z_{ij}^n. Note that results for Z_{ij}^n are, in fact, results for the measure on trajectories obtained when the single initial ancestor is of type i. Incorporating the initial type into the notation in this way, rather than through the measure as P_i and E_i, is helpful in calculations.

Before our theorems can be stated, further notation for intensity measures and their transforms is needed. Define $\mu_{ij}^{n\star}$ inductively by

$$\mu_{ij}^{(n+1)\star} = \sum_{k=1}^p \mu_{ik} * \mu_{kj}^{n\star}$$

where '$*$' denotes ordinary convolution of measures. It is easy to confirm, by induction on n, that the point process Z_{ij}^n has the intensity measure $\mu_{ij}^{n\star}$. Furthermore, the counts of the numbers of each type in each generation, given by $(Z_{i1}^n(\mathbb{R}^d), \ldots, Z_{ip}^n(\mathbb{R}^d))$, is a multitype Galton–Watson process, which is discussed in Athreya and Ney (1972), for example. A matrix $\boldsymbol{A}$ of nonnegative entries is called positive regular when, for some positive integer n, all the entries of $\boldsymbol{A}^n$ are strictly positive. The multitype Galton–Watson process is positive regular when the matrix $(P(Z_{ij}(\mathbb{R}^d) > 0))$ is positive regular. Throughout, the embedded Galton–Watson process is assumed to be positive regular. This Galton–Watson process is supercritical when the largest eigenvalue of its mean matrix, $(\mathrm{E}\, Z_{ij}(\mathbb{R}^d))$, exceeds 1. (For a positively regular process, this eigenvalue will be infinite when the mean matrix has any infinite entries.) A supercritical process survives with positive probability. One of the conditions of our theorems will imply that the process is supercritical.

The d-dimensional complex space $\mathbb{C}^d$ is equipped with the maximum metric. Hence, for $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d) \in \mathbb{C}^d$, we have $|\boldsymbol{x}, \boldsymbol{y}| = \max\{|x_i - y_i| \colon i = 1, \ldots, d\}$ where $|x_i - y_i|$ is the usual absolute value in $\mathbb{C}$. Let $B(\boldsymbol{x}, r)$ be the open ball centred at $\boldsymbol{x}$ and of radius r using this metric, and let $\overline{B}(\boldsymbol{x}, r)$ be its closure. (Later, we will also need $S(\boldsymbol{x}, r)$ and $\overline{S}(\boldsymbol{x}, r)$, the open and closed balls in $\mathbb{R}^d$.) In all that follows, we reserve the letters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ for the real and imaginary parts of $\boldsymbol{\lambda} \in \mathbb{C}^d$, so that, with this convention, $\boldsymbol{\lambda} = \boldsymbol{\theta} + \mathrm{i}\boldsymbol{\eta}$. For any $A \subset \mathbb{C}^d$, let A_- be its intersection with $\mathbb{R}^d$; thus

$$A_- = \{\boldsymbol{\lambda} \in A \colon \boldsymbol{\lambda} = \boldsymbol{\theta} + \mathrm{i}\boldsymbol{\eta} \text{ with } \boldsymbol{\eta} = 0\}.$$

Define the Laplace transforms $m_{ij}(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} \in \mathbb{C}^d$, by

$$m_{ij}(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} e^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} \mu_{ij}(\mathrm{d}\boldsymbol{x}) \left(= \mathrm{E} \int_{\mathbb{R}^d} e^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z_{ij}(\mathrm{d}\boldsymbol{x}) \right).$$

Then $|m_{ij}(\boldsymbol{\lambda})| \le m_{ij}(\boldsymbol{\theta})$ and, by Hölder's inequality, $\{\boldsymbol{\theta} \in \mathbb{R}^d : m_{ij}(\boldsymbol{\theta}) < \infty\}$ is a convex set. Let int A be the interior of A and let

$$L = \bigcap_{i,j} \text{int}\{\boldsymbol{\lambda} = \boldsymbol{\theta} + i\boldsymbol{\eta} \in \mathbb{C}^d : m_{ij}(\boldsymbol{\theta}) < \infty\}.$$

Assumption (1) is now strengthened to the following:

$$L \text{ is nonempty.}$$

In fact, L is a nonempty, open, convex subset of $\mathbb{C}^d$ and each $m_{ij}(\boldsymbol{\lambda})$ is analytic in $\boldsymbol{\lambda} \in L$. Let $\boldsymbol{M}(\boldsymbol{\lambda})$ be the matrix given by $\boldsymbol{M}(\boldsymbol{\lambda}) = (m_{ij}(\boldsymbol{\lambda}))$ and let $\boldsymbol{M}^n(\boldsymbol{\lambda})$ be its nth power, with (i, j)th entry $m_{ij}^n(\boldsymbol{\lambda})$. Furthermore, let

$$\mathcal{M}_{ij}^n(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} e^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z_{ij}^n(\mathrm{d}\boldsymbol{x}).$$

Then

$$m_{ij}^n(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} e^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} \mu_{ij}^{n\star}(\mathrm{d}\boldsymbol{x}) = \mathrm{E}\,\mathcal{M}_{ij}^n(\boldsymbol{\lambda}).$$

By analogy with the case in which the entries are nonnegative, the eigenvalue ρ of $\boldsymbol{M}$ is called the *maximum-modulus eigenvalue* if it is a simple root of $\det[z\boldsymbol{I} - \boldsymbol{M}]$ and its modulus is strictly larger than that of all the other roots ($\boldsymbol{I}$ being the identity matrix). Note that, when a maximum-modulus eigenvalue exists, it is automatically unique.

When $\boldsymbol{\vartheta} \in L_-(= L \cap \mathbb{R}^d)$, the entries of $\boldsymbol{M}(\boldsymbol{\vartheta})$ are finite, nonnegative real numbers. The positive regularity of the embedded Galton–Watson process clearly implies that $\boldsymbol{M}(\boldsymbol{\vartheta})$ is then positively regular. This means that the following extension of Perron–Frobenius theory which, is discussed in the final section, applies to $\boldsymbol{M}$. In essence, in the present context, it says that Perron–Frobenius properties extend smoothly to complex arguments with suitably small imaginary parts. Part (iii) of Theorem 1, below, gives an asymptotic estimate, (2), of $\boldsymbol{M}^n(\boldsymbol{\lambda})$ as $n \to \infty$. The key point about (2) is that it gives a uniform bound on the rate of convergence in a suitable neighbourhood. This uniformity has no direct parallel in Perron–Frobenius theory and requires a little care to establish.

Theorem 1. *Suppose that the $p \times p$ matrix $\boldsymbol{M} = (m_{ij})$ of functions defined on the open set $L \subset \mathbb{C}^d$ satisfies the following conditions: for all i and j, $m_{ij}(\boldsymbol{\lambda})$ are analytic functions in $\boldsymbol{\lambda} \in L$, and, for all $\tilde{\boldsymbol{\lambda}} \in \tilde{L} \subset L$, $\boldsymbol{M}(\tilde{\boldsymbol{\lambda}})$ is positive regular. Then there is an open set $\Omega \subset L$ containing $\tilde{L}$ such that, for $\boldsymbol{\lambda} \in \Omega$, the following statements hold.*

(i) *$\boldsymbol{M}(\boldsymbol{\lambda})$ has a maximum-modulus eigenvalue, $\rho(\boldsymbol{\lambda})$, that is analytic in $\boldsymbol{\lambda}$.*

(ii) *The left- and right-eigenvectors associated with $\rho(\boldsymbol{\lambda})$, namely $\boldsymbol{u}(\boldsymbol{\lambda})$ and $\boldsymbol{v}(\boldsymbol{\lambda})$, normalised so that $\sum_{i=1}^p u_i(\boldsymbol{\lambda}) = 1$ and $\sum_{i=1}^p u_i(\boldsymbol{\lambda})v_i(\boldsymbol{\lambda}) = 1$, are analytic in $\boldsymbol{\lambda}$ and, for all i, $u_i(\boldsymbol{\lambda}) \ne 0$ and $v_i(\boldsymbol{\lambda}) \ne 0$.*

(iii) *For any $\boldsymbol{\vartheta} \in \tilde{L}$, there is a neighbourhood, $B \in \Omega$, containing $B(\boldsymbol{\vartheta}, \delta)$ for some $\delta > 0$, and constants $K < \infty$ and $\gamma \in (0, 1)$ such that, for all n, i, and j,*

$$\sup_{\boldsymbol{\lambda} \in B} |\rho(\boldsymbol{\lambda})^{-n}(M(\boldsymbol{\lambda})^n)_{ij} - v_i(\boldsymbol{\lambda})u_j(\boldsymbol{\lambda})| \le K\gamma^n. \tag{2}$$

All the notation introduced in Theorem 1 is now applied to $\boldsymbol{M}$, the matrix of Laplace transforms, with $\tilde{L} = L_-$. Thus, $\boldsymbol{M}$ has maximum-modulus eigenvalue ρ and eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}$ defined on the set Ω, with the properties described in Theorem 1. For all $\boldsymbol{\lambda} \in \Omega$, $n = 1, 2, \ldots$, and i and j, define

$$\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) = \frac{v_j(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})} \rho(\boldsymbol{\lambda})^{-n} \mathcal{M}_{ij}^n(\boldsymbol{\lambda}) = \frac{v_j(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})} \rho(\boldsymbol{\lambda})^{-n} \int_{\mathbb{R}^d} \mathrm{e}^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z_{ij}^n(\mathrm{d}\boldsymbol{x}), \qquad (3)$$

which is a scaled version of the Laplace transform of the point process Z_{ij}^n. Establishing the uniform convergence of such transforms in the one-type case is an integral part of the approach to results on the distribution of the points described in Biggins (1992). For the multitype case, results on the distribution of points were obtained in Bramson *et al.* (1992), where particles were confined to the integer lattice of $\mathbb{R}$. More precise results, of the kind given in Biggins (1992), should hold and should not be limited to the lattice case. One such result is given in Theorem 7, below.

Let $\mathcal{F}_n$ be the σ-algebra that contains all information on the multitype branching random walk up to generation n. In the one-type case, (3) defines a martingale with respect to $\mathcal{F}_n$, which makes aspects of the study simpler. In the multitype case, it is the sequence

$$\mathcal{W}_i^n(\boldsymbol{\lambda}) = \sum_j \mathcal{W}_{ij}^n(\boldsymbol{\lambda}) = \sum_j \frac{v_j(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})} \rho(\boldsymbol{\lambda})^{-n} \int_{\mathbb{R}^d} \mathrm{e}^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z_{ij}^n(\mathrm{d}\boldsymbol{x}) \qquad (4)$$

that is a martingale; this is well known and proved here in Lemma 1. Our approach to the convergence of $\{\mathcal{W}_{ij}^n(\boldsymbol{\lambda})\}$ involves first considering the convergence of the martingale $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$. Note that, given any sample path of the process, each $\mathcal{W}_{ij}^n(\boldsymbol{\lambda})$ and $\mathcal{W}_i^n(\boldsymbol{\lambda})$ is analytic in $\boldsymbol{\lambda} \in \Omega$.

We now introduce certain sets that will be used to define the region where convergence occurs. For $\alpha \in (1, 2]$, let

$$\Omega_\alpha^2 = \left\{ \boldsymbol{\lambda} \in \Omega \colon \alpha \boldsymbol{\lambda} \in \Omega, \ \frac{\rho(\alpha \boldsymbol{\theta})}{|\rho(\boldsymbol{\lambda})|^\alpha} < 1 \right\}, \qquad (5)$$

$$\Omega_\alpha^3 = \mathrm{int}\left\{ \boldsymbol{\lambda} = \boldsymbol{\theta} + \mathrm{i}\boldsymbol{\eta} \in \Omega \colon \ \max_i \mathrm{E}[\mathcal{W}_i^1(\boldsymbol{\theta})^\alpha] < \infty \right\}, \qquad (6)$$

$$\Lambda_\alpha = \Omega_\alpha^2 \cap \Omega_\alpha^3 \quad \text{and} \quad \Lambda = \bigcup_{1 < \alpha \leq 2} \Lambda_\alpha.$$

These are all open sets in $\mathbb{C}^d$. Error estimates we derive involve $(\rho(\alpha \boldsymbol{\theta})/|\rho(\boldsymbol{\lambda})|^\alpha)^n$. Hence, having $\boldsymbol{\lambda} \in \Omega_\alpha^2$ ensures that such bounds decay quickly with n. Obviously, having $\boldsymbol{\lambda} \in \Omega_\alpha^3$ imposes a moment condition. The approach requires $\boldsymbol{\lambda}$ to satisfy both these conditions for the same $\alpha \in (1, 2]$, which leads to the definition of Λ_α and then Λ.

The process is supercritical, that is, $\rho(\boldsymbol{0}) > 1$, whenever there exists a $\boldsymbol{\theta} \in \Omega_\alpha^2$ for some $\alpha \in (1, 2]$. To check this, first note that $\log \rho(\boldsymbol{\theta})$ is a convex function; see Kingman (1961), Miller (1961), or Seneta (1973, Theorem 3.7). Hence,

$$\frac{(\alpha - 1)}{\alpha} \log \rho(\boldsymbol{0}) + \frac{1}{\alpha} \log \rho(\alpha \boldsymbol{\theta}) \geq \log \rho(\boldsymbol{\theta}),$$

and $\log(\rho(\alpha \boldsymbol{\theta})/\rho(\boldsymbol{\theta})^\alpha) < 0$ when $\boldsymbol{\theta} \in \Omega_\alpha^2$. Then $(\alpha - 1)\log \rho(\boldsymbol{0}) > 0$ and, so, $\rho(\boldsymbol{0}) > 1$.

Theorem 2. *Let $\alpha \in (1, 2]$ and $\boldsymbol{\lambda} = \boldsymbol{\theta} + \mathrm{i}\boldsymbol{\eta} \in \Lambda_\alpha$. Then $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ converges almost surely and in αth mean, as $n \to \infty$, for each i.*

This gives the following result on mean convergence as an immediate consequence.

Corollary 1. *Suppose that* $\lambda \in \Lambda$. *Then* $\{\mathcal{W}_i^n(\lambda)\}$ *converges almost surely and in mean, as* $n \to \infty$, *for each* i.

The mean convergence of $\{\mathcal{W}_i^n(\lambda)\}$ for real λ was also discussed in Bramson *et al.* (1992), Kyprianou and Rahimzadeh Sani (2001), and, rather briefly, in Biggins and Kyprianou (2004).

Theorem 3. *Assume that* $\Lambda_- \neq \varnothing$. *Then there is an open set* Γ, *with* $\Lambda_- \subset \Gamma \subset \Lambda$, *such that, for each* i, *the martingale* $\{\mathcal{W}_i^n(\lambda)\}$ *converges uniformly in any compact subset of* Γ, *almost surely and in mean, as* $n \to \infty$. *Furthermore, the limit* $\mathcal{W}_i(\lambda)$ *is analytic in* $\lambda \in \Gamma$.

By building on these convergence results for the martingale, we will obtain analogous results for $\{\mathcal{W}_{ij}^n(\lambda)\}$.

Theorem 4. *Let* $\alpha \in (1, 2]$ *and assume that* $(\Lambda_\alpha)_- \neq \varnothing$. *Then there is an open set* Γ_1, *with* $(\Lambda_\alpha)_- \subset \Gamma_1 \subset \Lambda_\alpha$, *such that, for all* $\lambda \in \Gamma_1$, $\{\mathcal{W}_{ij}^n(\lambda)\}$ *converges to* $u_j(\lambda)v_j(\lambda)\mathcal{W}_i(\lambda)$, *almost surely and in* αth *mean, as* $n \to \infty$, *where* $\mathcal{W}_i(\lambda)$ *is the limit of the martingale* $\{\mathcal{W}_i^n(\lambda)\}$ *as* $n \to \infty$.

Theorem 5. *Assume that* $\Lambda_- \neq \varnothing$. *Then there is an open set* Γ_2, *with* $\Lambda_- \subset \Gamma_2 \subset \Lambda$, *such that, for all* i, j, $\{\mathcal{W}_{ij}^n(\lambda)\}$ *converges uniformly in any compact subset of* Γ_2, *almost surely and in mean, as* $n \to \infty$, *to the random variable* $u_j(\lambda)v_j(\lambda)\mathcal{W}_i(\lambda)$, *where* $\mathcal{W}_i(\lambda)$ *is the limit of the martingale* $\{\mathcal{W}_i^n(\lambda)\}$ *as* $n \to \infty$.

The sequences $\{\mathcal{W}_i^n(\lambda)\}$ and $\{\mathcal{W}_{ij}^n(\lambda)\}$ converge on the set Γ_2 introduced in Theorem 5. However, to move from information on transforms to information on the associated measures, a result on the behaviour of $\mathcal{M}_{ij}^n(\lambda)$ is needed for $\theta \in L_-$, but in the case where λ need not be in Γ_2.

The branching random walk is *strongly nonlattice* when it is positively regular and, for some (k, l) and some $\theta \in \{\vartheta : m_{kl}(\vartheta) < \infty\}$,

$$\left| \frac{m_{kl}(\theta + i\eta)}{m_{kl}(\theta)} \right| = 1 \quad \text{only when } \eta = 0. \tag{7}$$

This follows the usage of 'strongly nonlattice' in Bhattacharya (1977, Equation (1.64)), rather than that in Stone (1965).

Theorem 6. *For a strongly nonlattice branching random walk, for any set* $\mathcal{K} \subset \Lambda_-$ *that is compact in* $\mathbb{R}^d$ *and any* $\epsilon \in (0, 1)$, *there is an* $\varepsilon < 1$ *such that*

$$\varepsilon^{-n} \sup_{i,j} \sup_{\theta \in \mathcal{K}} \sup_{\epsilon \leq |\eta| \leq \epsilon^{-1}} \left| \frac{\mathcal{M}_{ij}^n(\theta + i\eta)}{\rho^n(\theta)} \right| \to 0 \quad \text{as } n \to \infty, \tag{8}$$

almost surely.

Between them, Theorems 5 and 6 provide enough information on the behaviour of the transforms $\mathcal{M}_{ij}^n(\lambda)$ to develop good estimates of the associated measures. In the one-type case, in Biggins (1992), this step relies heavily on the corresponding results for sums of independent, identically distributed random variables, obtained in Stone (1967). Unfortunately, this approach does not extend directly, since the existing results for Markov additive processes, which are here the analogue of independent, identically distributed random variables, are not suitable. Instead,

the treatment has to go back to the general methods for saddlepoint approximations. The results needed, which give an approximation to a measure based only on a few of its characteristics, are contained in Theorems 9 and 10 in Section 5. That section is independent of the rest of the paper and is of more general interest. We illustrate there how Theorems 9 and 10 connect with refined large deviation results of Chaganty and Sethuraman (1993) and Stone (1967).

It is worth noting that, in this study, the measures of interest, Z_{ij}^n, are atomic. Hence, if (7) were replaced by the stronger, but more tractable, Cramér's condition (Bhattacharya (1977, Equation (1.36))), that condition could not transfer to the Z_{ij}^n.

The assumption that $\log \rho$ is strictly convex will be imposed. This amounts to saying that the branching random walk is truly d-dimensional, but we do not demonstrate this. For $\boldsymbol{\theta} \in L_-$, let $(-1)^i \boldsymbol{v}_i(\boldsymbol{\theta})$ be the array of ith derivatives of $\log \rho(\boldsymbol{\theta})$. Then $\boldsymbol{v}_2(\boldsymbol{\theta})$ is a positive-definite matrix when $\log \rho$ is strictly convex. Let $\mathfrak{D}$ be the space of continuously differentiable real-valued functions on $\mathbb{R}^d$ and, if $f \in \mathfrak{D}$, let its vector of derivatives be ∇f. We need functions that decay suitably at infinity, so we introduce

$$\mathfrak{G}(G) = \left\{ f \in \mathfrak{D} \colon \int_0^\infty \sup\{|f(\boldsymbol{x})|, |\nabla f(\boldsymbol{x})| \colon |\boldsymbol{x}| \geq (r-1)\}(r+1)^{d+1}\, \mathrm{d}r \leq G \right\}.$$

Furthermore, for $C \subset \mathbb{R}^d$, we use the notation $C + \boldsymbol{a} = \{\boldsymbol{c} + \boldsymbol{a} \colon \boldsymbol{c} \in C\}$.

Theorem 7. *Assume that* $\log \rho$ *is strictly convex and that the process is strongly nonlattice. Let* $\mathcal{K}$ *be a compact subset of* Λ_-. *Let* h *be such that, for some finite* G *and all* $\boldsymbol{\theta} \in \mathcal{K}$, *the function* $\mathrm{e}^{\boldsymbol{\theta}^\top \boldsymbol{x}} h(\boldsymbol{x})$ *is in* $\mathfrak{G}(G)$. *Then, with*

$$\xi(\boldsymbol{\theta}) = -\boldsymbol{\theta}^\top \boldsymbol{v}_1(\boldsymbol{\theta}) - \log \rho(\boldsymbol{\theta})$$

and C *any convex set in* $\mathbb{R}^d$,

$$n^{d/2} \mathrm{e}^{n\xi(\boldsymbol{\theta})} \int_C h(\boldsymbol{x}) Z_{ij}^n(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1(\boldsymbol{\theta})) \to \frac{v_i(\boldsymbol{\theta}) u_j(\boldsymbol{\theta}) W_i(\boldsymbol{\theta})}{\sqrt{(2\pi)^d \det[\boldsymbol{v}_2(\boldsymbol{\theta})]}} \int_C \mathrm{e}^{\boldsymbol{\theta}^\top \boldsymbol{x}} h(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

uniformly in $\boldsymbol{\theta} \in \mathcal{K}$ *and* C, *almost surely on* $\mathscr{S}$, *where* $\mathscr{S}$ *is the survival set of the underlying Galton–Watson process.*

The condition that $\mathrm{e}^{\boldsymbol{\theta}^\top \boldsymbol{x}} h(\boldsymbol{x})$ be in $\mathfrak{G}(G)$ becomes more restrictive as $\mathcal{K}$ becomes larger. In particular, it forces $h(\boldsymbol{x})$ to decay rapidly with $|\boldsymbol{x}|$ when the origin is in the interior of $\mathcal{K}$.

Corollary 2. *Let* $b < \infty$. *Under the conditions of Theorem 7,*

$$n^{d/2} \mathrm{e}^{n\xi(\boldsymbol{\theta})} Z_{ij}^n(C + n\boldsymbol{v}_1(\boldsymbol{\theta})) \to \frac{v_i(\boldsymbol{\theta}) u_j(\boldsymbol{\theta}) W_i(\boldsymbol{\theta})}{\sqrt{(2\pi)^d \det[\boldsymbol{v}_2(\boldsymbol{\theta})]}} \int_C \mathrm{e}^{\boldsymbol{\theta}^\top \boldsymbol{x}}\, \mathrm{d}\boldsymbol{x}$$

uniformly in convex $C \subset \{\boldsymbol{x} \colon |\boldsymbol{x}| \leq b\}$ *and* $\boldsymbol{\theta} \in \mathcal{K}$, *almost surely on* $\mathscr{S}$.

It is worth pointing out that if, for some $\boldsymbol{y}$, $\mathcal{K}$ lies inside the half-space $\{\boldsymbol{\theta} \colon \boldsymbol{\theta}^\top \boldsymbol{y} > 0\}$, then the result of this corollary can also hold for some sets C that are not bounded.

The relationships between ρ, $\boldsymbol{v}_1$, and ξ are the usual ones associated with large deviation and saddlepoint calculations; see Jensen (1995, Section 2.2), for example. The assumption that ρ is strictly convex means that compact subsets of Λ_- translate, under $\boldsymbol{y} = \boldsymbol{v}_1(\boldsymbol{\theta})$, into compact subsets of $\mathrm{int}\{\boldsymbol{v}_1(\boldsymbol{\theta}) \colon \boldsymbol{\theta} \in L_-\}$.

Theorem 7 has the following analogue for the measures $\mu_{ij}^{n\star}$, with the proof requiring only obvious changes.

Theorem 8. *Under the conditions of Theorem 7,*

$$n^{d/2}e^{n\xi(\theta)}\int_C h(x)\mu_{ij}^{n\star}(\mathrm{d}x + n v_1(\theta)) \to \frac{v_i(\theta)u_j(\theta)}{\sqrt{(2\pi)^d \det[v_2(\theta)]}}\int_C e^{\theta^\top x}h(x)\,\mathrm{d}x$$

uniformly in $\theta \in \mathcal{K}$ and C.

2. Proofs of Theorems 2 and 3

For any $\lambda = \theta + i\eta \in \Omega$, define the functions

$$\bar{v}(\lambda) = \max_{i,j}\left|\frac{v_j(\lambda)}{v_i(\lambda)}\right|, \qquad \underline{v}(\lambda) = \min_{i,j}\left|\frac{v_j(\lambda)}{v_i(\lambda)}\right|, \qquad \nu(\lambda) = \frac{\bar{v}(\lambda)}{\underline{v}(\theta)},$$

and

$$\phi(\lambda) = \frac{\rho(\theta)}{|\rho(\lambda)|}.$$

All of these are strictly positive, continuous functions of $\lambda \in \Omega$, where Ω is as in Theorem 1. Also, since $|m_{ij}(\lambda)| \leq m_{ij}(\theta)$, we know that $|\rho(\lambda)| \leq \rho(\theta)$ (Lancaster and Tismenetsky (1985), Theorem 15.2.1)) and, so, $\phi(\lambda) \geq 1$.

For each $\alpha \in (1,2]$, define Ω_α by $\Omega_\alpha = \{\lambda \in \Omega : \alpha\lambda \in \Omega\}$. Then Ω_α is an open subset of Ω. For $\alpha \in (1,2]$ and $\lambda \in \Omega_\alpha$, let

$$\nu_1(\lambda) = \frac{\bar{v}(\lambda)^\alpha}{\underline{v}(\alpha\theta)} \quad \text{and} \quad \kappa(\lambda) = \frac{\rho(\alpha\theta)}{|\rho(\lambda)|^\alpha}.$$

Then ν_1 and κ are strictly positive, continuous functions of $\lambda \in \Omega_\alpha$; they depend on α, but this has been suppressed in the notation. Note that Ω_α^2, defined in (5), can now be written as $\Omega_\alpha^2 = \{\lambda \in \Omega_\alpha : \kappa(\lambda) < 1\}$ and is an open subset of Ω_α. Define

$$\beta(\theta) = \max_i \mathrm{E}[\mathcal{W}_i^1(\theta)^\alpha].$$

Then Ω_α^3, defined in (6), can be written as $\Omega_\alpha^3 = \mathrm{int}\{\lambda = \theta + i\eta \in \Omega : \beta(\theta) < \infty\}$ and β is a real-valued, continuous function on Ω_α^3.

Let $\alpha \geq 1$. Define the αth absolute central moment of a random (complex-valued) variable X, $\sigma^\alpha(X)$, by

$$\sigma^\alpha(X) = \mathrm{E}[|X - \mathrm{E}\,X|^\alpha],$$

and the αth absolute central moment conditional on the σ-algebra $\mathcal{G}$ by

$$\sigma^\alpha(X \mid \mathcal{G}) = \mathrm{E}[|X - \mathrm{E}[X \mid \mathcal{G}]|^\alpha \mid \mathcal{G}].$$

Let $\{z_{ik;s}^l : s\}$ be the positions of the particles making up Z_{ik}^l, where by $z_{ik;s}^l$ we denote the position of the sth particle of type k in generation l when the initial ancestor is of type i. Now, by looking at the particles in the nth generation as the $(n-l)$th-generation children of the particles in generation l, we can introduce $Z_{kj}^n(\cdot|l,s)$ as the point process giving the positions of the type-j particles in generation n descended from the sth particle of type k in generation l, relative to the

latter particle's position. Given $\mathcal{F}_l$, $\{Z_{kj}^n(\cdot|l, s) : s\}$ are independent copies of the point process Z_{kj}^{n-l}. Thus, conditional on $\mathcal{F}_l$, the random variables

$$\mathcal{W}_k^{n-l}(\boldsymbol{\lambda}|l, s) = \sum_j \frac{v_j(\boldsymbol{\lambda})}{v_k(\boldsymbol{\lambda})} \frac{1}{\rho(\boldsymbol{\lambda})^{n-l}} \int_{\mathbb{R}^d} e^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z_{kj}^n(d\boldsymbol{x}|l, s) \tag{9}$$

are, as s varies, independent and identical copies of $\mathcal{W}_k^{n-l}(\boldsymbol{\lambda})$. Furthermore,

$$\mathcal{W}_i^n(\boldsymbol{\lambda}) = \sum_k \frac{v_k(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})} \frac{1}{\rho(\boldsymbol{\lambda})^l} \sum_s \exp(-\boldsymbol{\lambda}^\top z_{ik;s}^l) \mathcal{W}_k^{n-l}(\boldsymbol{\lambda}|l, s). \tag{10}$$

Lemma 1. $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ *is a martingale with respect to* $\{\mathcal{F}_n\}$.

Proof. From Theorem 1, $E\,\mathcal{W}_k^1(\boldsymbol{\lambda}) = 1$. Therefore, decomposition (10) with $l = n-1$ gives the result.

The following result, which is an extension of von Bahr and Esseen (1965, Equation (4)) to complex-valued random variables, was proved in Biggins (1992). Barral (2001) provided a very pretty approach to uniform convergence in the one-type case by applying a lemma like this to Banach-space-valued random variables.

Lemma 2. *If* $\{X_n\}$ *are independent, complex-valued random variables with* $E\,X_n = 0$ *or, more generally, martingale differences, then for* $\alpha \in (1, 2]$ *there exists a constant* $C > 0$, *depending on* α *but independent of* n *and the sequence* $\{X_n\}$, *such that*

$$E\left[\left|\sum_{j=1}^n X_j\right|^\alpha\right] \le C \sum_{j=1}^n E[|X_j|^\alpha].$$

The next lemma, which is the multitype version of Biggins (1992, Lemma 2), gives some bounds related to the martingale $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$. In bounding formulae like these, the argument $\boldsymbol{\lambda}$ will often be omitted; thus, for example, $\phi \equiv \phi(\boldsymbol{\lambda})$.

Lemma 3. *Let* $\alpha \in (1, 2]$, $\boldsymbol{\lambda} = \boldsymbol{\theta} + i\boldsymbol{\eta} \in \Lambda_\alpha$, *and* $\varphi = v_1 \phi^\alpha v^\alpha \beta$. *Then there are constants* c_1, c_2, *and* c_3, *depending on* α *but not on* $\boldsymbol{\lambda}$, i, *or* n, *such that the following statements hold.*

(i) $(E\,|\mathcal{W}_i^{n+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|)^\alpha \le E[|\mathcal{W}_i^{n+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha] \le c_1 \varphi \kappa^n$.

(ii) $\sigma^\alpha(\mathcal{W}_i^n(\boldsymbol{\lambda})) \le c_2 \varphi (1 - \kappa)^{-1}$.

(iii) $\sum_{k=n}^\infty E\,|\mathcal{W}_i^{k+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^k(\boldsymbol{\lambda})| \le c_3 \varphi^{1/\alpha} \kappa^{n/\alpha} (1 - \kappa^{1/\alpha})^{-1}$.

(iv) *For* $\boldsymbol{\vartheta} \in (\Lambda_\alpha)_-$, *there exists a* $\delta > 0$ *such that*

$$\sup\{E\,|\mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha : \boldsymbol{\lambda} \in B(\boldsymbol{\vartheta}, \delta), i, n\} < \infty.$$

Proof. The first inequality in part (i) is just $E[|X|^\alpha] \ge (E\,|X|)^\alpha$. Since $E[|\mathcal{W}_i^1(\boldsymbol{\lambda})|^\alpha] \ge 1$, we obtain

$$E[|\mathcal{W}_i^1(\boldsymbol{\lambda}) - 1|^\alpha] \le 4\,E[|\mathcal{W}_i^1(\boldsymbol{\lambda})|^\alpha] \le 4\phi^\alpha v^\alpha \beta.$$

Using (10), Lemma 2, and the fact that, given $\mathcal{F}_n$, $\mathcal{W}_k^1(\boldsymbol{\lambda}|n, s)$ are independent copies of $\mathcal{W}_k^1(\boldsymbol{\lambda})$, we have

$$
\begin{aligned}
\sigma^\alpha(\mathcal{W}_i^{n+1}(\boldsymbol{\lambda}) \mid \mathcal{F}_n) &= \mathrm{E}[|\mathcal{W}_i^{n+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha \mid \mathcal{F}_n] \\
&\leq C \sum_k \left| \frac{v_k(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})\rho(\boldsymbol{\lambda})^n} \right|^\alpha \sum_s |\exp(-\boldsymbol{\lambda}^\top z_{ik;s}^n)|^\alpha \, \mathrm{E}[|\mathcal{W}_k^1(\boldsymbol{\lambda}) - 1|^\alpha] \\
&\leq c_1 v_1 \kappa^n \mathcal{W}_i^n(\alpha\boldsymbol{\theta})(\phi^\alpha v^\alpha \beta) \\
&= c_1 \varphi \kappa^n \mathcal{W}_i^n(\alpha\boldsymbol{\theta}).
\end{aligned}
$$

Taking expectations now gives part (i).

Note that, since $\boldsymbol{\lambda} \in \Lambda_\alpha$, we have $\kappa < 1$. Part (ii) now follows directly from summing over n in part (i) and using Lemma 2. To prove part (iii), sum the inequalities in part (i).

The continuity of the functions v, v_1, ϕ, and κ on Λ_α and part (ii) mean that there exists a constant $C > 0$, depending only on α and $\delta > 0$, such that, for any $\boldsymbol{\lambda} \in \overline{B}(\boldsymbol{\vartheta}, \delta)$, we have $\sigma^\alpha(\mathcal{W}_i^n(\boldsymbol{\lambda})) \leq C$. Since $\mathrm{E}[|\mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha] \leq 2 + 2\sigma^\alpha(\mathcal{W}_i^n(\boldsymbol{\lambda}))$, this implies part (iv).

Proof of Theorem 2. From part (iii) of Lemma 3, $\mathrm{E} \sum_{n=0}^\infty |\mathcal{W}_i^{n+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|$ is finite and, so, $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ is a Cauchy sequence almost surely. Thus, as $n \to \infty$, $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ converges to $\mathcal{W}_i(\boldsymbol{\lambda})$ almost surely. Applying Fatou's lemma, Lemma 2, and Lemma 3(i) gives

$$
\begin{aligned}
\mathrm{E}[|\mathcal{W}_i(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha] &\leq \liminf_{N \to \infty} \mathrm{E}[|\mathcal{W}_i^{n+N}(\boldsymbol{\lambda}) - \mathcal{W}_i^n(\boldsymbol{\lambda})|^\alpha] \\
&\leq C \sum_{j=0}^\infty \mathrm{E}[|\mathcal{W}_i^{n+j+1}(\boldsymbol{\lambda}) - \mathcal{W}_i^{n+j}(\boldsymbol{\lambda})|^\alpha] \\
&\leq C c_1 \varphi \frac{\kappa^n}{1 - \kappa}.
\end{aligned}
$$

Since $\kappa < 1$, $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ converges in αth mean.

The distinguished boundary (see Hörmander (1973)) of $B(\boldsymbol{x}, r)$, which is a subset of the topological boundary, is defined by

$$
D(\boldsymbol{x}, r) = \{\boldsymbol{y} \in \mathbb{C}^d : |x_s - y_s| = r, \ s = 1, \ldots, d\}.
$$

We assume $D(\boldsymbol{x}, r)$ to be parametrized such that

$$
D(\boldsymbol{x}, r) = \{\boldsymbol{z}(\boldsymbol{t}) = (z_1(\boldsymbol{t}), \ldots, z_d(\boldsymbol{t})) : z_s(\boldsymbol{t}) = x_s + re^{it_s}, \ s = 1, \ldots, d, \ \boldsymbol{t} \in I\},
$$

where $I = [0, 2\pi]^d$ is a d-dimensional closed cube in $\mathbb{R}^d$. The next lemma, which is Biggins (1992, Lemma 3), is the key to obtaining bounds that hold uniformly.

Lemma 4. *If f is analytic on the open ball $B = B(\boldsymbol{x}, 2r')$ and $r < r'$, then*

$$
\sup\{|f(\boldsymbol{\lambda})| : \boldsymbol{\lambda} \in B(\boldsymbol{x}, r)\} \leq \pi^{-d} \int_I |f(\boldsymbol{z}(\boldsymbol{t}))| \, \mathrm{d}\boldsymbol{t},
$$

where $\boldsymbol{z}(\boldsymbol{t}) \in D(\boldsymbol{x}, 2r)$.

Proof of Theorem 3. The proof closely follows that of the one-type result of Biggins (1992). Let $x \in \Lambda_-$ be fixed. There exist an $\alpha \in (1, 2]$ and a $B = B(x, 3r) \subset \Lambda_\alpha$, which means that $B \subset \Omega_\alpha^2$ and $B \subset \Omega_\alpha^3$. For any two positive integers n and $N \geq n$, we apply Lemma 4 to $\mathcal{W}_i^{N+1}(\lambda) - \mathcal{W}_i^n(\lambda)$ to obtain

$$\pi^d \sup_{\lambda \in B(x,r)} |\mathcal{W}_i^{N+1}(\lambda) - \mathcal{W}_i^n(\lambda)| \leq \int_I \sum_{m=n}^N |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt,$$

$$\leq \sum_{m=n}^\infty \int_I |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt,$$

where $z(t) \in D(x, 2r) \subset B$. Applying Fubini's theorem and Lemma 3(iii), we obtain

$$\mathrm{E} \sum_{m=n}^\infty \int_I |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt \leq \int_I C\varphi^{1/\alpha} \frac{\kappa^{n/\alpha}}{1 - \kappa^{1/\alpha}} \, dt,$$

where the argument $z(t)$ has been suppressed on the right-hand side. Recall that $\kappa < 1$ throughout Ω_α^2 and $\beta < \infty$ throughout Ω_α^3. Therefore, using the continuity of the various functions, there exist constants $\delta < 1$ and $K < \infty$ such that the integrand on the right-hand side can be bounded by $K\delta^n$ throughout B and, hence, throughout $D(x, 2r)$. Thus,

$$\mathrm{E} \sum_{m=n}^\infty \int_I |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt \leq K\delta^n \int_I dt < \infty$$

and, so,

$$\sum_{m=0}^\infty \int_I |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt < \infty \quad \text{almost surely.}$$

Therefore, $\{\mathcal{W}_i^n(\lambda)\}$ is a Cauchy sequence under the supremum norm $\| \cdot \|$ on $B(x, r)$, which implies uniform convergence on $B(x, r)$, almost surely.

For the remainder of the proof, the supremum norm is defined over $B(x, r)$. From the almost-sure uniform convergence of the martingale, for fixed n, as $N \to \infty$,

$$\|\mathcal{W}_i^N - \mathcal{W}_i^n\| \to \|\mathcal{W}_i - \mathcal{W}_i^n\|$$

almost surely. Let n be fixed; then, by taking the expectation of both sides and applying first Fatou's lemma and then Lemma 4, we obtain

$$\mathrm{E} \|\mathcal{W}_i - \mathcal{W}_i^n\| \leq \liminf_{N \to \infty} \mathrm{E} \|\mathcal{W}_i^N - \mathcal{W}_i^n\|$$

$$\leq \pi^{-d} \mathrm{E} \int_I \sum_{m=n}^\infty |\mathcal{W}_i^{m+1}(z(t)) - \mathcal{W}_i^m(z(t))| \, dt$$

$$\leq \pi^{-d} K\delta^n \int_I dt < \infty$$

$$\to 0 \quad \text{as } n \to \infty.$$

This proves uniform convergence on $B(x, r)$, in mean.

Finally, define

$$\Gamma = \bigcup_{\boldsymbol{x} \in \Lambda_-} B(\boldsymbol{x}, r).$$

Then, by a compactness argument, the uniform convergence of $\{\mathcal{W}_i^n(\boldsymbol{\lambda})\}$ in the open balls in Γ implies the uniform convergence in any compact subset of Γ.

3. Proofs of Theorems 4 and 5

Let $1 \le l \le n$ and let i and j be fixed. We start with a decomposition that is similar to (10),

$$\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) = \sum_k \sum_s \frac{v_k(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})} \frac{1}{\rho(\boldsymbol{\lambda})^l} \exp(-\boldsymbol{\lambda}^\top z_{ik;s}^l) \mathcal{W}_{kj}^{n-l}(\boldsymbol{\lambda}|l, s), \tag{11}$$

where, given $\mathcal{F}_l$, $\mathcal{W}_{kj}^{n-l}(\boldsymbol{\lambda}|l, s)$ are independent copies of $\mathcal{W}_{kj}^{n-l}(\boldsymbol{\lambda})$ as s varies. Hence,

$$\mathrm{E}[\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l] = \frac{v_j(\boldsymbol{\lambda})}{v_i(\boldsymbol{\lambda})\rho(\boldsymbol{\lambda})^n} \sum_k \sum_s \exp(-\boldsymbol{\lambda}^\top z_{ik;s}^l) m_{kj}^{n-l}(\boldsymbol{\lambda}). \tag{12}$$

Therefore, given any sample path of the process, $\mathrm{E}[\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l]$ is an analytic function of $\boldsymbol{\lambda} \in \Omega$.

For all $\boldsymbol{\lambda} \in \Omega$, define

$$g_{ij}^n(\boldsymbol{\lambda}|l) = \mathcal{W}_{ij}^n(\boldsymbol{\lambda}) - \mathrm{E}[\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l]$$

and

$$h_{ij}^n(\boldsymbol{\lambda}|l) = \mathrm{E}[\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l] - u_j(\boldsymbol{\lambda})v_j(\boldsymbol{\lambda})\mathcal{W}_i^l(\boldsymbol{\lambda}).$$

The functions $g_{ij}^n(\boldsymbol{\lambda}|\cdot)$ and $h_{ij}^n(\boldsymbol{\lambda}|\cdot)$ are analytic functions of $\boldsymbol{\lambda} \in \Omega$. Furthermore,

$$|\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) - u_j(\boldsymbol{\lambda})v_j(\boldsymbol{\lambda})\mathcal{W}_i(\boldsymbol{\lambda})| \le |g_{ij}^n(\boldsymbol{\lambda}|l)| + |h_{ij}^n(\boldsymbol{\lambda}|l)| + |u_j(\boldsymbol{\lambda})v_j(\boldsymbol{\lambda})||\mathcal{W}_i^l(\boldsymbol{\lambda}) - \mathcal{W}_i(\boldsymbol{\lambda})|.$$

Hence, $|\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) - u_j(\boldsymbol{\lambda})v_j(\boldsymbol{\lambda})\mathcal{W}_i(\boldsymbol{\lambda})|^\alpha$ is less than or equal to

$$3(|g_{ij}^n(\boldsymbol{\lambda}|l)|^\alpha + |h_{ij}^n(\boldsymbol{\lambda}|l)|^\alpha + |u_j(\boldsymbol{\lambda})v_j(\boldsymbol{\lambda})|^\alpha |\mathcal{W}_i^l(\boldsymbol{\lambda}) - \mathcal{W}_i(\boldsymbol{\lambda})|^\alpha). \tag{13}$$

The idea now is to let $l = \lfloor \frac{1}{2}n \rfloor$, where $\lfloor x \rfloor$ is the greatest integer not exceeding x, and let n tend to infinity. This motivates the lemmas we now give before returning to the main proof. It is in these results that we need the uniformity proved in Theorem 1(iii).

Lemma 5. *Suppose that $\boldsymbol{\vartheta} \in (\Lambda_\alpha)_-$. Then there are positive constants c_1, c_2, and c_3 depending on α, a constant $\gamma < 1$, and a neighbourhood B of $\boldsymbol{\vartheta}$ such that, for all $\boldsymbol{\lambda} \in B$, $n = 1, 2, \ldots$, and all i and j, we have*

(i) $|h_{ij}^n(\boldsymbol{\lambda}|l)| \le c_1 \gamma^{n-l} v \phi^l \mathcal{W}_i^l(\boldsymbol{\theta})$,

(ii) $\mathrm{E} \sum_{l=0}^n |h_{ij}^n(\boldsymbol{\lambda}|l)|^\alpha \le c_2 \phi^{n\alpha}$, *and*

(iii) $\sigma^\alpha(\mathcal{W}_{ij}^n(\boldsymbol{\lambda})) \le c_3 \phi^{n\alpha}$.

Proof. Substituting, using (4) and (12), we have

$$h_{ij}^n(\lambda|l) = \frac{v_j(\lambda)}{v_i(\lambda)\rho(\lambda)^l} \sum_k \sum_s \exp(-\lambda^\top z_{ik;s}^l)[\rho(\lambda)^{-(n-l)}m_{kj}^{n-l}(\lambda) - u_j(\lambda)v_k(\lambda)].$$

Therefore, by Theorem 1(iii), there exist constants $c_1 > 0$ and $\gamma \in (0, 1)$ and a neighbourhood B of ϑ such that, for all $\lambda \in B$ and all n,

$$|h_{ij}^n(\lambda|l)| \le c_1 v \phi^l \gamma^{n-l} \sum_k \sum_s \left[\frac{v_k(\theta)}{v_i(\theta)} \rho(\theta)^{-l} \exp(-\theta^\top z_{ik;s}^l) \right]$$

$$= c_1 v \phi^l \gamma^{n-l} W_i^l(\theta), \tag{14}$$

as required. This and Lemma 3(iv) combine, after making B smaller if necessary, to show that, for some constant c_1', $\mathrm{E}\,|h_{ij}^n(\lambda|l)|^\alpha \le c_1' v^\alpha (\gamma/\phi)^{(n-l)\alpha}\phi^{n\alpha}$. Since $\gamma < 1$ and $\phi \ge 1$, we have $(\gamma/\phi) < 1$ for all $\lambda \in B$. Thus,

$$\mathrm{E} \sum_{l=0}^n |h_{ij}^n(\lambda|l)|^\alpha \le c_1' v^\alpha \frac{1}{1 - (\gamma/\phi)^\alpha} \phi^{n\alpha}.$$

The continuity of the functions v and ϕ now implies part (ii). Finally, using Lemma 2,

$$\sigma^\alpha(W_{ij}^n(\lambda)) \le C \sum_{l=1}^n \mathrm{E}[|\,\mathrm{E}[W_{ij}^n(\lambda) \mid \mathcal{F}_l] - \mathrm{E}[W_{ij}^n(\lambda) \mid \mathcal{F}_{l-1}]|^\alpha]$$

$$\le 3^\alpha C \left(|u_j(\lambda)v_j(\lambda)|^\alpha \sum_{l=1}^n \mathrm{E}[|W_i^l(\lambda) - W_i^{l-1}(\lambda)|^\alpha] + 2\,\mathrm{E} \sum_{l=0}^n |h_{ij}^n(\lambda|l)|^\alpha \right).$$

By suitable bounding of the continuous functions involved in Lemma 3(i), for all $\lambda \in B(\vartheta, \delta)$ with δ sufficiently small, we have

$$\sum_{l=1}^\infty \mathrm{E}[|W_i^l(\lambda) - W_i^{l-1}(\lambda)|^\alpha] \le K_1 < \infty$$

and, so, using this and part (ii), we have

$$\sigma^\alpha(W_{ij}^n(\lambda)) \le 3^\alpha C(|u_j(\lambda)v_j(\lambda)|^\alpha K_1 + 2c_2\phi^{n\alpha}).$$

Now the boundedness of $u_j(\lambda)v_j(\lambda)$ on B and the fact that $\phi \ge 1$ combine to prove part (iii).

Lemma 6. *Suppose that $\alpha \in (1, 2]$ and $\vartheta \in (\Lambda_\alpha)_-$. Then there exists a $B = B(\vartheta, \delta) \subset \Lambda_\alpha$ such that, for $l = \lfloor \frac{1}{2}n \rfloor$,*

   (i) *for all $\lambda \in B$, $h_{ij}^n(\lambda|l) \to 0$ as $n \to \infty$, almost surely and in αth mean; and*

   (ii) *$\sup\{|h_{ij}^n(\lambda|l)|\colon \lambda \in B\} \to 0$ as $n \to \infty$, almost surely and in mean.*

Proof. Let $B = B(\vartheta, \delta)$ be a neighbourhood of ϑ throughout which (14) holds. Take δ to be smaller if necessary, so that $\phi\gamma < 1$ throughout B. Now application of Theorems 2 and 3 to $W_i^l(\theta)$ together with simple bounding in (14) gives all the claimed results.

Lemma 7. *Suppose that $\alpha \in (1, 2]$ and $\boldsymbol{\vartheta} \in (\Lambda_\alpha)_-$. Then there exists a neighbourhood of $\boldsymbol{\vartheta}$, say $B = B(\boldsymbol{\vartheta}, r) \subset \Lambda_\alpha$, such that, for $l = \lfloor \frac{1}{2} n \rfloor$,*

(i) *for all $\boldsymbol{\lambda} \in B$, $g_{ij}^n(\boldsymbol{\lambda}|l) \to 0$ as $n \to \infty$, almost surely and in αth mean; and*

(ii) $\sup\{|g_{ij}^n(\boldsymbol{\lambda}|l)|: \boldsymbol{\lambda} \in B\} \to 0$ *as $n \to \infty$, almost surely and in mean.*

Proof. First note that, using (11) and Lemma 2,

$$\sigma^\alpha(\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l) \leq C\bar{v}(\boldsymbol{\lambda})^\alpha |\rho(\boldsymbol{\lambda})|^{-l\alpha} \sum_k \sum_s \exp(-\alpha\boldsymbol{\theta}^\top z_{ik;s}^l)\sigma^\alpha(\mathcal{W}_{kj}^{n-l}(\boldsymbol{\lambda})). \tag{15}$$

Let $B = B(\boldsymbol{\vartheta}, \delta)$ be inside the neighbourhood of $\boldsymbol{\vartheta}$ introduced in Lemma 5. Then, for some positive constant c and for all $\boldsymbol{\lambda} \in B$, $n = 1, 2, \ldots$, and all i and j, applying Lemma 5(iii) to the bound (15) gives

$$\sigma^\alpha(\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l) \leq cv_1 \mathcal{W}_i^l(\alpha\boldsymbol{\theta})\phi^{\alpha(n-l)}\kappa^l. \tag{16}$$

Now, since $\mathrm{E}\, \mathcal{W}_i^l(\alpha\boldsymbol{\theta}) = 1$, we have

$$\mathrm{E}[|g_{ij}^n(\boldsymbol{\lambda}|l)|^\alpha] = \mathrm{E}\, \sigma^\alpha(\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l) \leq cv_1\phi^{\alpha(n-l)}\kappa^l.$$

Choose δ to be smaller if necessary, so that $\phi^\alpha\kappa < 1$ for all $\boldsymbol{\lambda} \in B$. Then the right-hand side of the above inequality converges to 0 geometrically quickly, as $n \to \infty$, when $l = \lfloor \frac{1}{2} n \rfloor$. This implies that $g_{ij}^n(\boldsymbol{\lambda}|l) \to 0$ in αth mean and almost surely.

Jensen's inequality and (16) give

$$\mathrm{E}\, |g_{ij}^n(\boldsymbol{\lambda}|l)| \leq (\mathrm{E}\, \sigma^\alpha(\mathcal{W}_{ij}^n(\boldsymbol{\lambda}) \mid \mathcal{F}_l))^{1/\alpha} \leq c^{1/\alpha}v_1^{1/\alpha}\phi^{n-l}\kappa^{l/\alpha}.$$

For $l = \lfloor \frac{1}{2} n \rfloor$, $\phi^{n-l}\kappa^{l/\alpha}$ is asymptotically equivalent to $(\phi\kappa^{1/\alpha})^l$, and $\phi\kappa^{1/\alpha} \leq \gamma_1 < 1$ in B. Hence, there exists a $C' < \infty$ such that, for all $\boldsymbol{\lambda} \in B$, $n = 1, 2, \ldots, l = \lfloor \frac{1}{2} n \rfloor$, and all i and j, we have $\mathrm{E}\, |g_{ij}^n(\boldsymbol{\lambda}|l)| \leq C'\gamma_1^l$. Let $2r < \delta$ and apply Lemma 4 to the analytic functions $g_{ij}^n(\boldsymbol{\lambda}|l)$ to obtain

$$\sup_{\boldsymbol{\lambda} \in B(\boldsymbol{\vartheta}, r)} |g_{ij}^n(\boldsymbol{\lambda}|l)| \leq \pi^{-d} \int_I |g_{ij}^n(\boldsymbol{z}(\boldsymbol{t})|l)| \, \mathrm{d}\boldsymbol{t},$$

where $\boldsymbol{z}(\boldsymbol{t}) \in D(\boldsymbol{\vartheta}, 2r)$. The expectation of the right-hand side here goes to 0 geometrically quickly. Hence, $g_{ij}^n(\boldsymbol{\lambda}|l)$ converges to 0 uniformly in $\boldsymbol{\lambda} \in B(\boldsymbol{\vartheta}, r)$, almost surely and in mean.

Proof of Theorems 4 and 5. In (13), let $l = \lfloor \frac{1}{2} n \rfloor$ and then let $n \to \infty$. In a suitable neighbourhood of $\boldsymbol{\vartheta} \in (\Lambda_\alpha)_-$, the almost-sure and αth-mean convergence of $\mathcal{W}_i^n(\boldsymbol{\lambda}) - \mathcal{W}_i(\boldsymbol{\lambda})$, $g_{ij}^n(\boldsymbol{\lambda}|l)$, and $h_{ij}^n(\boldsymbol{\lambda}|l)$ are contained in Theorem 2, Lemma 7(i), and Lemma 6(i), respectively. Now take the union of these neighbourhoods to be Γ_1. This proves Theorem 4.

Similarly, the almost-sure and mean uniform convergence of $\mathcal{W}_i^n(\boldsymbol{\lambda}) - \mathcal{W}_i(\boldsymbol{\lambda})$, $g_{ij}^n(\boldsymbol{\lambda}|l)$, and $h_{ij}^n(\boldsymbol{\lambda}|l)$ are contained in Theorem 3, Lemma 7(ii) and Lemma 6(ii), respectively. A union of suitable neighbourhoods now provides Γ_2, proving Theorem 5.

4. Proof of Theorem 6

Lemma 8. *If the branching random walk is strongly nonlattice then, for any compact set $\mathcal{K} \subset \{\boldsymbol{\theta} \in \mathbb{R}^d: \sup_{i,j} m_{ij}(\boldsymbol{\theta}) < \infty\}$ and all $a \in (0, 1)$, there exists an $\varepsilon < 1$ such that*

$$\varepsilon^{-n} \sup_{i,j} \sup_{\boldsymbol{\theta} \in \mathcal{K}} \sup_{a \leq |\boldsymbol{\eta}| \leq a^{-1}} \left| \frac{m_{ij}^n(\boldsymbol{\theta} + i\boldsymbol{\eta})}{\rho^n(\boldsymbol{\theta})} \right| \to 0 \quad \text{as } n \to \infty. \tag{17}$$

Proof. Let (k, l) be such that (7) holds. As a function of $\boldsymbol{\eta}$, $m_{kl}(\boldsymbol{\theta} + i\boldsymbol{\eta})/m_{kl}(\boldsymbol{\theta})$ is the characteristic function of the probability measure $m_{kl}(\boldsymbol{\theta})^{-1}e^{-\boldsymbol{\theta}^\top x}\mu_{kl}(dx)$ and (7) implies that this measure is strongly nonlattice. Then, using this and continuity, we have

$$\tilde{\varepsilon} = \sup_{\boldsymbol{\theta} \in \mathcal{K}} \sup_{a \leq |\boldsymbol{\eta}| \leq a^{-1}} \left| \frac{m_{kl}(\boldsymbol{\theta} + i\boldsymbol{\eta})}{m_{kl}(\boldsymbol{\theta})} \right| < 1.$$

Let $\boldsymbol{M}^\dagger(\boldsymbol{\theta})$ be the matrix with (k, l)th entry $\tilde{\varepsilon}m_{kl}(\boldsymbol{\theta})$ and with other entries the same as those in $\boldsymbol{M}(\boldsymbol{\theta})$, and let $(M^\dagger(\boldsymbol{\theta})^n)_{ij}$ be the (i, j)th entry of its nth power. Let $\rho^\dagger(\boldsymbol{\theta})$ be the eigenvalue of $\boldsymbol{M}^\dagger(\boldsymbol{\theta})$ of maximum modulus, with right-eigenvector $\boldsymbol{v}^\dagger(\boldsymbol{\theta})$. Then $\rho^\dagger(\boldsymbol{\theta}) < \rho(\boldsymbol{\theta})$ (see Seneta (1973, Theorem 1.1)). Furthermore, since Theorem 1 applies to $\boldsymbol{M}^\dagger(\boldsymbol{\theta})$, there exists an $\varepsilon < 1$ such that $\rho^\dagger(\boldsymbol{\theta}) < \varepsilon\rho(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathcal{K}$, and

$$\sup_{\boldsymbol{\theta} \in \mathcal{K}} \max_{i,j} \frac{v_i^\dagger(\boldsymbol{\theta})}{v_j^\dagger(\boldsymbol{\theta})} = c < \infty.$$

Now,

$$\sup_{a \leq |\boldsymbol{\eta}| \leq a^{-1}} |(M(\boldsymbol{\theta} + i\boldsymbol{\eta})^n)_{ij}| \leq (M^\dagger(\boldsymbol{\theta})^n)_{ij} \leq c(\varepsilon\rho(\boldsymbol{\theta}))^n,$$

proving (17).

Proof of Theorem 6. It is sufficient to prove (8) with the compact set $\mathcal{K}$ being some closed ball in $\mathbb{R}^d$ centred on $\boldsymbol{\vartheta} \in \Lambda_-$, since a simple covering argument extends the result to general $\mathcal{K}$. We begin the proof by introducing various bounds leading to the appropriate ball to consider. Recall that $S(\boldsymbol{x}, r)$ is a ball in $\mathbb{R}^d$ centred on $\boldsymbol{x}$ with radius r, and $\overline{S}(\boldsymbol{x}, r)$ is its closure.

Choose $\boldsymbol{\vartheta} \in \Lambda_-$; then $\boldsymbol{\vartheta} \in \Lambda_\alpha$ for some α, which is now fixed. Now take δ_1 such that $S(\boldsymbol{\vartheta}, 2\delta_1) \subset \Lambda_\alpha$ and $S_1 = \overline{S}(\boldsymbol{\vartheta}, \delta_1)$. Let ε_1 be an ε such that (17) holds when $\mathcal{K} = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in S_1\}$ and let $a = \frac{1}{2}\epsilon$ (where ϵ is as defined in the statement of the theorem). Recall that

$$\frac{\rho(\alpha\boldsymbol{\theta})^{1/\alpha}}{\rho(\boldsymbol{\theta})} < 1, \qquad \boldsymbol{\theta} \in \Lambda_\alpha.$$

Let

$$\varepsilon_2 = \max\left\{\varepsilon_1, \sup_{\boldsymbol{\theta} \in S_1} \frac{\rho(\alpha\boldsymbol{\theta})^{1/\alpha}}{\rho(\boldsymbol{\theta})}\right\}.$$

Now let $S_2 = \overline{S}(\boldsymbol{\vartheta}, 2\delta_2)$ with $2\delta_2 < \delta_1$, and let

$$\overline{\rho} = \sup_{\boldsymbol{\theta} \in S_2} \rho(\boldsymbol{\theta}) \quad \text{and} \quad \underline{\rho} = \inf_{\boldsymbol{\theta} \in S_2} \rho(\boldsymbol{\theta}).$$

Take δ_2 small enough that

$$\varepsilon_3 := \varepsilon_2 \frac{\sup_{\boldsymbol{\theta} \in S_2} \rho(\boldsymbol{\theta})}{\inf_{\boldsymbol{\theta} \in S_2} \rho(\boldsymbol{\theta})} = \varepsilon_2 \frac{\overline{\rho}}{\underline{\rho}} < 1. \tag{18}$$

Note that, for $\boldsymbol{\theta} \in S_2$,

$$\rho(\alpha\boldsymbol{\theta})^{1/\alpha} \leq \varepsilon_2\rho(\boldsymbol{\theta}) \leq \varepsilon_2\overline{\rho}. \tag{19}$$

These preliminaries provide a suitably small neighbourhood of ϑ, S_2, in which various inequalities hold. Now let $S = \overline{S}(\vartheta, \delta_2) \subset S_2 = \overline{S}(\vartheta, 2\delta_2)$ and define the two regions

$$R = \{\theta + i\eta \colon \theta \in S, \ \epsilon \leq |\eta| \leq \epsilon^{-1}\} \quad \text{and} \quad R_2 = \{\theta + i\eta \colon \theta \in S_2, \ \tfrac{1}{2}\epsilon \leq |\eta| \leq (\tfrac{1}{2}\epsilon)^{-1}\}.$$

We will prove (8) for $\theta \in S$, that is, for $\lambda \in R$. Clearly,

$$\sup_{\lambda \in R} |\mathcal{M}_{ij}^n(\lambda)| \leq \sup_{\lambda \in R} |\mathcal{M}_{ij}^n(\lambda) - m_{ij}^n(\lambda)| + \sup_{\lambda \in R} |m_{ij}^n(\lambda)|. \tag{20}$$

Furthermore,

$$\sup_{\lambda \in R} |m_{ij}^n(\lambda)| \leq \sup_{\lambda \in R_2} |m_{ij}^n(\lambda)| = \sup_{\lambda \in R_2} \left| \frac{m_{ij}^n(\lambda)}{\rho(\theta)^n} \rho(\theta)^n \right| \leq (\varepsilon_1 \overline{\rho})^n \leq (\varepsilon_2 \overline{\rho})^n. \tag{21}$$

Using Lemma 2, we have

$$\mathrm{E}[|\mathcal{M}_{ij}^n(\lambda) - m_{ij}^n(\lambda)|^\alpha] \leq C \sum_{l=0}^{n-1} \mathrm{E}[|\,\mathrm{E}[\mathcal{M}_{ij}^n(\lambda) \mid \mathcal{F}_{l+1}] - \mathrm{E}[\mathcal{M}_{ij}^n(\lambda) \mid \mathcal{F}_l]|^\alpha], \tag{22}$$

and splitting at the lth generation shows that

$$\mathrm{E}[\mathcal{M}_{ij}^n(\lambda) \mid \mathcal{F}_l] = \sum_k \left(\sum_s \exp(-\lambda^\top z_{ik;s}^l) \right) m_{kj}^{n-l}(\lambda)$$
$$= \sum_k \left(\int_{\mathbb{R}^d} \exp(-\lambda^\top z) Z_{ik}^l(\mathrm{d}z) \right) m_{kj}^{n-l}(\lambda). \tag{23}$$

Following (9), let

$$\mathcal{M}_{kj}^{n-l}(\lambda|l, s) = \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda^\top x} Z_{kj}^n(\mathrm{d}x|l, s);$$

then

$$\int_{\mathbb{R}^d} \exp(-\lambda^\top z) Z_{ik}^{l+1}(\mathrm{d}z) = \sum_h \sum_s \exp(-\lambda^\top z_{ih;s}^l) \mathcal{M}_{hk}^1(\lambda|l, s). \tag{24}$$

Let $\Delta_l = \mathrm{E}[\mathcal{M}_{ij}^n(\lambda) \mid \mathcal{F}_{l+1}] - \mathrm{E}[\mathcal{M}_{ij}^n(\lambda) \mid \mathcal{F}_l]$. Then, using (23) and (24), we have

$$\Delta_l = \sum_k \left(\sum_h \sum_s \exp(-\lambda^\top z_{ih;s}^l)(\mathcal{M}_{hk}^1(\lambda|l, s) - m_{hk}(\lambda)) \right) m_{kj}^{n-l-1}(\lambda)$$

and, given $\mathcal{F}_l$, $\{\mathcal{M}_{hk}^1(\lambda|l, s) - m_{hk}(\lambda) \colon s\}$ are independent variables with zero mean. We need a bound on the moments of these variables. Let

$$c_1 = 4 \left(\max_{i,j} \sup_{\theta \in S_2} \mathrm{E}[|\mathcal{M}_{ij}^1(\theta)|^\alpha] \right)^{1/\alpha},$$

which is finite because $S_2 \subset \Lambda_\alpha$. Then, simple bounding gives

$$\sup\{\mathrm{E}[|\mathcal{M}_{ij}^1(\lambda) - m_{ij}(\lambda)|^\alpha] \colon \lambda = \theta + i\eta, \ \theta \in S_2, i, j\} \leq (c_1)^\alpha.$$

Let

$$c_2 = \max_{i,j} \sup_{\boldsymbol{\theta} \in S_2} \frac{v_j(\alpha\boldsymbol{\theta})}{v_i(\alpha\boldsymbol{\theta})},$$

which is finite, by Theorem 1(ii), because $S_2 \subset \Lambda_\alpha$. Then, using the fact that

$$m_{ij}^n(\alpha\boldsymbol{\theta}) v_j(\alpha\boldsymbol{\theta}) \leq \rho(\alpha\boldsymbol{\theta})^n v_i(\alpha\boldsymbol{\theta})$$

and (19), for $\boldsymbol{\theta} \in S_2$ we have

$$m_{ij}^n(\alpha\boldsymbol{\theta}) \leq c_2(\varepsilon_2 \overline{\rho})^{n\alpha}.$$

Now, applying Lemma 2 again, (21), and the two bounds just given, for any $\boldsymbol{\lambda} \in R_2$ we have

$$\begin{aligned}
\mathrm{E}[|\Delta_{l-1}|^\alpha] &\leq C \sum_k \left(\sum_h m_{ih}^{l-1}(\alpha\boldsymbol{\theta}) \, \mathrm{E}[|\mathcal{M}_{hk}^1(\boldsymbol{\lambda}) - m_{hk}(\boldsymbol{\lambda})|^\alpha] \right) |m_{kj}^{n-l}(\boldsymbol{\lambda})|^\alpha \\
&\leq C p^2 c_2 (\varepsilon_2 \overline{\rho})^{(l-1)\alpha} (c_1)^\alpha (\varepsilon_2 \overline{\rho})^{(n-l)\alpha} \\
&\leq c_3 (\varepsilon_2 \overline{\rho})^{n\alpha},
\end{aligned}$$

where c_3 is independent of $\boldsymbol{\theta} \in S_2$ and l. Hence, (22) gives

$$\mathrm{E}[|\mathcal{M}_{ij}^n(\boldsymbol{\lambda}) - m_{ij}^n(\boldsymbol{\lambda})|^\alpha] \leq n c_3 (\varepsilon_2 \overline{\rho})^{n\alpha}$$

for any $\boldsymbol{\lambda} \in R_2$.

The region R can be covered with a finite number of open balls such that, when their radii are doubled, they still lie inside R_2. Hence, using (22), the bound just obtained, and Lemma 4, for a suitable constant c_4 we have

$$\mathrm{E} \sup_{\boldsymbol{\lambda} \in R} |\mathcal{M}_{ij}^n(\boldsymbol{\lambda}) - m_{ij}^n(\boldsymbol{\lambda})| \leq c_4 (\varepsilon_2 \overline{\rho})^n n^{1/\alpha}.$$

Combining this with (20) and (21) gives

$$\mathrm{E} \sup_{\boldsymbol{\lambda} \in R} |\mathcal{M}_{ij}^n(\boldsymbol{\lambda})| \leq (c_4 n^{1/\alpha} + c_1)(\varepsilon_2 \overline{\rho})^n$$

and, so, using (18), we have

$$\mathrm{E} \sup_{\boldsymbol{\lambda} \in R} \left| \frac{\mathcal{M}_{ij}^n(\boldsymbol{\lambda})}{\rho(\boldsymbol{\theta})^n} \right| \leq \frac{1}{\underline{\rho}^n} \mathrm{E} \sup_{\boldsymbol{\lambda} \in R} |\mathcal{M}_{ij}^n(\boldsymbol{\lambda})| \leq (c_4 n^{1/\alpha} + c_1) \varepsilon_3^n.$$

Hence, (8) holds for any $\varepsilon \in (\varepsilon_3, 1)$.

5. Approximation of measures

This section contains the preparatory work on saddlepoint approximations necessary for the proof of Theorem 7. The idea is to develop explicit estimates that apply to a particular measure through only a few of its attributes. Most treatments of these matters, made with a view to application to variables that are, or look like, sums of independent, identically distributed random variables, bring n into the picture sooner than we do. The treatment draws on ideas from Stone (1967), von Bahr (1967), Bhattacharya (1972), (1977), Chaganty and Sethuraman (1993), and Jensen (1995).

Recall that $S(\boldsymbol{x}, \epsilon) = \{\boldsymbol{y} \in \mathbb{R}^d \colon |\boldsymbol{x} - \boldsymbol{y}| < \epsilon\}$. Let Γ be a probability measure on $\mathbb{R}^d$, with $\gamma := \Gamma(S(\boldsymbol{0}, 1)) > \frac{1}{2}$, whose characteristic function vanishes outside $\{\boldsymbol{t} \colon |\boldsymbol{t}| \geq \zeta\}$ for some finite ζ. Let Γ_ϵ be the measure given by $\Gamma_\epsilon(A) = \Gamma(\{\boldsymbol{x} \colon \epsilon \boldsymbol{x} \in A\})$, so that its characteristic function vanishes outside $\{\boldsymbol{t} \colon |\boldsymbol{t}| \geq \epsilon^{-1}\zeta\}$. Let ν and μ be probability measures on $\mathbb{R}^d$, let b_ϵ be the supremum of the modulus of the density of $(\nu - \mu) * \Gamma_\epsilon$, and let q be the supremum of the density of μ. For any f, let

$$\omega_f(\boldsymbol{x}, \epsilon) = \sup\{|f(\boldsymbol{y}) - f(\boldsymbol{z})| \colon \boldsymbol{y}, \boldsymbol{z} \in S(\boldsymbol{x}, \epsilon)\}$$

and

$$\overline{f}^\epsilon(\boldsymbol{x}) = \sup\{|f(\boldsymbol{y})| \colon \boldsymbol{y} \in S(\boldsymbol{x}, \epsilon)\}.$$

The next result is contained in, and easily derived from, Lemma 2.2 of Bhattacharya (1972) and the remark following it.

Lemma 9. *For any real-valued, bounded, Borel-measurable function f, we have*

$$\sup_{\boldsymbol{u} \in \mathbb{R}^d} \left| \int f(\boldsymbol{x})(\nu - \mu)(\mathrm{d}\boldsymbol{x} + \boldsymbol{u}) \right| \leq \frac{1}{2\gamma - 1} \left[b_\epsilon \int \overline{f}^\epsilon(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + q \int \omega_f(\boldsymbol{x}, 2\epsilon) \, \mathrm{d}\boldsymbol{x} \right].$$

The idea now is to make this bound simpler by imposing suitable smoothness conditions on f. The set $\mathfrak{G}'(G)$ is made up of those $f \in \mathfrak{D}$ for which there is a bounded decreasing function g with $\int_0^\infty g(r)(r + 1)^{d-1} \, \mathrm{d}r = G$ and

$$\sup\{|f(\boldsymbol{x})|, |\nabla f(\boldsymbol{x})| \colon |\boldsymbol{x}| \geq (r - 1)\} \leq g(r).$$

Note that $\mathfrak{G}(G)$, introduced to state Theorem 7, is defined in the same way, but with

$$\int_0^\infty g(r)(r + 1)^{d+1} \, \mathrm{d}r = G;$$

therefore, $\mathfrak{G}(G) \subset \mathfrak{G}'(G)$. Also, let $\mathfrak{C}$ denote the convex sets in $\mathbb{R}^d$.

Lemma 10. *For a constant Δ, depending only on the dimension, and $\epsilon < 1$, we have*

$$\sup\left\{ \left| \int_C f(\boldsymbol{x})(\nu - \mu)(\mathrm{d}\boldsymbol{x} + \boldsymbol{u}) \right| \colon \boldsymbol{u} \in \mathbb{R}^d, \, C \in \mathfrak{C}, \, f \in \mathfrak{G}'(G) \right\} \leq G \Delta (b_\epsilon + \epsilon q).$$

Proof. For a set $A \subset \mathbb{R}^d$, let $A^\epsilon = \{\boldsymbol{y} \colon |\boldsymbol{y} - \boldsymbol{a}| < \epsilon, \, \boldsymbol{a} \in A\}$, and let ∂A be the boundary of A. Let $\mathfrak{S}$ be the surface area of the unit ball. Then, for $C \in \mathfrak{C}$, from von Bahr (1967, Equation (15)), we have

$$\int 1_{\partial C^\epsilon}(\boldsymbol{x}) g(|\boldsymbol{x}|) \, \mathrm{d}\boldsymbol{x} \leq 2\mathfrak{S} G \epsilon.$$

Take $\epsilon < 1$ and let $h(\boldsymbol{x}) = f(\boldsymbol{x}) 1_C(\boldsymbol{x})$, with $f \in \mathfrak{G}'(G)$. If $\boldsymbol{x} \in \partial C^\epsilon$ then

$$\omega_h(\boldsymbol{x}, \epsilon) \leq 2 \sup\{|f(\boldsymbol{y})| \colon \boldsymbol{y} \in S(\boldsymbol{x}, \epsilon)\} \leq 2g(|\boldsymbol{x}|).$$

If $\boldsymbol{x} \in S(\boldsymbol{x}, \epsilon) \subset C$ then, using Taylor's theorem, $\omega_h(\boldsymbol{x}, \epsilon) = \omega_f(\boldsymbol{x}, \epsilon) \leq 2\epsilon g(|\boldsymbol{x}|)$. Hence,

$$\omega_h(\boldsymbol{x}, \epsilon) \leq 2g(|\boldsymbol{x}|)(1_{\partial C^\epsilon}(\boldsymbol{x}) + \epsilon 1_C(\boldsymbol{x}))$$

and, so,

$$\int \omega_h(\boldsymbol{x}, \epsilon)\, \mathrm{d}\boldsymbol{x} \le 2 \int (1_{\partial C^\epsilon}(\boldsymbol{x}) + \epsilon 1_C(\boldsymbol{x})) g(|\boldsymbol{x}|)\, \mathrm{d}\boldsymbol{x} \le 6\epsilon \mathfrak{S} G.$$

Also,

$$\int \overline{h}^\epsilon(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \le \int \overline{f}^\epsilon(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \le \int g(|\boldsymbol{x}|)\, \mathrm{d}\boldsymbol{x} = \mathfrak{S} G.$$

Substituting these into the bound in Lemma 9 gives the result.

Recall that, for $A \subset \mathbb{C}^d$, A_- is its intersection with $\mathbb{R}^d$. Let $\mathcal{K}$ be compact in $\mathbb{R}^d$, let $\tau < 1$, and let

$$\mathcal{K}_\tau = \{\boldsymbol{\theta} + \boldsymbol{\lambda} \in \mathbb{C}^d : \boldsymbol{\theta} \in \mathcal{K},\ |\boldsymbol{\lambda}| \le \tau\}.$$

Now let B be an open set containing $\mathcal{K}_\tau$, with B_- convex. For Z a measure on $\mathbb{R}^d$, let

$$\widehat{Z}(\boldsymbol{\lambda}) = \int \mathrm{e}^{-\boldsymbol{\lambda}^\top \boldsymbol{x}} Z(\mathrm{d}\boldsymbol{x}).$$

Suppose that $\log \widehat{Z}$ is analytic on B and strictly convex on B_-. Let $-\boldsymbol{m}(\boldsymbol{\theta})$ be the vector of its first derivatives and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ the matrix of its second derivatives on B_-. Then $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is positive definite, because $\log \widehat{Z}$ is strictly convex. Let $\mathfrak{u}$ be a (finite) bound on the modulus of all derivatives of $\log \widehat{Z}$ over $\mathcal{K}_\tau$, up to and including order three. Let $\mathfrak{c}$ be a lower bound on the smallest eigenvalue of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ as $\boldsymbol{\theta}$ varies through $\mathcal{K}$, and let $\mathfrak{r} = \mathfrak{u}/\mathfrak{c}$. Since $\log \widehat{Z}$ is strictly convex and analytic on B_-, we can take $\mathfrak{c} > 0$. Also, let

$$\varepsilon(a) = a^{-d} \sup\left\{\left|\frac{\widehat{Z}(\boldsymbol{\theta} + \mathrm{i}\boldsymbol{t})}{\widehat{Z}(\boldsymbol{\theta})}\right| : \boldsymbol{\theta} \in \mathcal{K},\ a \le |\boldsymbol{t}| < \frac{1}{a}\right\}.$$

Fix $\boldsymbol{\vartheta} \in \mathcal{K}$. Let ν be the probability measure given by

$$\nu(A) = \frac{\int 1_A(\boldsymbol{x}) \mathrm{e}^{-\boldsymbol{\vartheta}^\top \boldsymbol{x}} Z(\mathrm{d}\boldsymbol{x})}{\int \mathrm{e}^{-\boldsymbol{\vartheta}^\top \boldsymbol{y}} Z(\mathrm{d}\boldsymbol{y})},$$

with mean $\boldsymbol{m}\ (= \boldsymbol{m}(\boldsymbol{\vartheta}))$ and covariance matrix $\boldsymbol{\Sigma}\ (= \boldsymbol{\Sigma}(\boldsymbol{\vartheta}))$. Let $\widehat{\nu}(\boldsymbol{\lambda})$ be the Laplace transform of ν, i.e. $\widehat{\nu}(\boldsymbol{\lambda}) = \widehat{Z}(\boldsymbol{\vartheta} + \boldsymbol{\lambda})/\widehat{Z}(\boldsymbol{\vartheta})$. Let $\mathfrak{N}$ be the normal distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$. The idea is to use $\mathfrak{N}$ to approximate ν and, hence, Z, through Lemma 10, in a way that is suitably uniform. Note that ν and $\mathfrak{N}$ both depend on $\boldsymbol{\vartheta}$; we use '$\nu_{\boldsymbol{\vartheta}}$' and '$\mathfrak{N}_{\boldsymbol{\vartheta}}$' in the next definition to emphasise this. Let

$$A(Z) = \sup\left\{\left|\int_C f(\boldsymbol{x})(\nu_{\boldsymbol{\vartheta}} - \mathfrak{N}_{\boldsymbol{\vartheta}})(\mathrm{d}\boldsymbol{x} + \boldsymbol{u})\right| : \boldsymbol{u} \in \mathbb{R}^d,\ f \in \mathfrak{G}'(G),\ C \in \mathfrak{C},\ \boldsymbol{\vartheta} \in \mathcal{K}\right\}.$$

Obviously, by bounding $A(Z)$ we allow f to be shifted arbitrarily, f to vary within $\mathfrak{G}'(G)$, and integration over an arbitrary convex set.

Theorem 9. *For $\delta \le \min\{\tau, \mathrm{e}^{-d}/4\mathfrak{r}\}$, $\epsilon \le \min\{1, \zeta/\delta, \zeta\delta\}$, and a constant Δ' depending only on the dimension, we have*

$$A(Z) \le G\Delta'\left[\frac{\mathfrak{r}}{\mathfrak{c}^{(d+1)/2}} + \varepsilon(\epsilon/\zeta) + \frac{(\mathrm{e}^{-\delta^2 \mathfrak{c}/4} + \epsilon)}{\mathfrak{c}^{d/2}}\right].$$

Proof. Let q be a bound on the density of $\mathfrak{N}$ and let b_ϵ be a bound on the modulus of the density of $(\nu - \mathfrak{N}) * \Gamma_\epsilon$. Note that $q \leq (2\pi\mathfrak{c})^{-d/2}$, which is independent of $\boldsymbol{\vartheta}$.

For $|\boldsymbol{\lambda}| \leq \tau$, let

$$\psi(\boldsymbol{\lambda}) = \log \widehat{\nu}(\boldsymbol{\lambda}) - \boldsymbol{m}^\top \boldsymbol{\lambda} + \tfrac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}$$
$$= \log \widehat{Z}(\boldsymbol{\vartheta} + \boldsymbol{\lambda}) - \log \widehat{Z}(\boldsymbol{\vartheta}) - \boldsymbol{m}^\top \boldsymbol{\lambda} + \tfrac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}.$$

Then, by arrangement, ψ is analytic on $|\boldsymbol{\lambda}| \leq \tau$, with $\psi(\boldsymbol{0}) = 0$ and all its first and second derivatives vanishing at $\boldsymbol{0}$.

Using Taylor's theorem and the analyticity of $\log \widehat{Z}$ on B, for $|\boldsymbol{\lambda}| \leq \tau$ we have

$$|\psi(\boldsymbol{\lambda})| \leq \frac{d^3}{3!} \mathfrak{u} |\boldsymbol{\lambda}|^3 \leq \mathrm{e}^d \mathfrak{u} |\boldsymbol{\lambda}|^3$$

and, so, for $|\boldsymbol{\lambda}| \leq \delta$,

$$|\psi(\boldsymbol{\lambda})| \leq \mathrm{e}^d \mathfrak{u} |\boldsymbol{\lambda}|^3 \leq \left(4\mathrm{e}^d \frac{\mathfrak{u}}{\mathfrak{c}} \delta \right) \frac{\mathfrak{c}}{4} |\boldsymbol{\lambda}|^2 \leq \frac{\mathfrak{c}}{4} |\boldsymbol{\lambda}|^2.$$

Using these two inequalities gives

$$|\exp(\psi(\boldsymbol{\lambda})) - 1| \leq \mathrm{e}^d \mathfrak{u} |\boldsymbol{\lambda}|^3 \exp\left(\frac{\mathfrak{c}}{4} |\boldsymbol{\lambda}|^2 \right) \quad \text{for } |\boldsymbol{\lambda}| \leq \delta.$$

The key point is that the right-hand side here does not depend on $\boldsymbol{\vartheta}$. Using this bound, we have

$$\int_{|\boldsymbol{t}| < \delta} |\widehat{\nu}(\mathrm{i}\boldsymbol{t}) - \mathrm{e}^{\mathrm{i}\boldsymbol{m}^\top \boldsymbol{t}} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2}| \, \mathrm{d}\boldsymbol{t} \leq \int_{|\boldsymbol{t}| < \delta} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2} |\mathrm{e}^{\psi(\mathrm{i}\boldsymbol{t})} - 1| \, \mathrm{d}\boldsymbol{t}$$

$$\leq \mathrm{e}^d \mathfrak{u} \int_{|\boldsymbol{t}| < \delta} |\boldsymbol{t}|^3 \mathrm{e}^{\mathfrak{c}|\boldsymbol{t}|^2/4} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2} \, \mathrm{d}\boldsymbol{t}$$

$$\leq \mathrm{e}^d \mathfrak{u} \int_{|\boldsymbol{t}| < \delta} |\boldsymbol{t}|^3 \mathrm{e}^{-|\boldsymbol{t}|^2 \mathfrak{c}/4} \, \mathrm{d}\boldsymbol{t}$$

$$\leq \mathrm{e}^d \mathfrak{u} \frac{1}{\mathfrak{c}^{(d+3)/2}} \int_0^\infty |z|^3 \mathrm{e}^{-|z|^2/4} \, \mathrm{d}z$$

$$\leq \frac{\Delta_1 \mathfrak{r}}{\mathfrak{c}^{(d+1)/2}},$$

where Δ_1 depends only on the dimension d. Thus,

$$(2\pi)^d b_\epsilon \leq \int_{|\boldsymbol{t}| < \zeta/\epsilon} |\widehat{\nu}(\mathrm{i}\boldsymbol{t}) - \mathrm{e}^{\mathrm{i}\boldsymbol{m}^\top \boldsymbol{t}} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2}| \, \mathrm{d}\boldsymbol{t}$$

$$\leq \int_{|\boldsymbol{t}| < \delta} |\widehat{\nu}(\mathrm{i}\boldsymbol{t}) - \mathrm{e}^{\mathrm{i}\boldsymbol{m}^\top \boldsymbol{t}} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2}| \, \mathrm{d}\boldsymbol{t} + \int_{\delta \leq |\boldsymbol{t}| < \zeta/\epsilon} |\widehat{\nu}(\mathrm{i}\boldsymbol{t})| \, \mathrm{d}\boldsymbol{t} + \int_{\delta \leq |\boldsymbol{t}|} \mathrm{e}^{-\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}/2} \, \mathrm{d}\boldsymbol{t}$$

$$\leq \frac{\Delta_1 \mathfrak{r}}{\mathfrak{c}^{(d+1)/2}} + \varepsilon\left(\frac{\epsilon}{\zeta} \right) \frac{\mathfrak{S}}{d} + \mathrm{e}^{-\delta^2 \mathfrak{c}/4} \frac{(4\pi)^{d/2}}{\mathfrak{c}^{d/2}}.$$

Substituting these estimates for b_ϵ and q into Lemma 10 gives the result.

For the main result, we also need to approximate the normal distribution $\mathfrak{N}$. To formulate the theorem, let

$$B(\boldsymbol{v}) = \sup\left\{\left|\left|\int_C f(\boldsymbol{x})\mathfrak{N}(\mathrm{d}\boldsymbol{x} + \boldsymbol{m} - \boldsymbol{v}) - \frac{\int_C f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{\sqrt{(2\pi)^d \det[\boldsymbol{\Sigma}]}}\right|\right| : f \in \mathfrak{G}(G),\ C \in \mathfrak{C},\ \boldsymbol{\vartheta} \in \mathcal{K}\right\}.$$

Recall that $\mathfrak{N}$, $\boldsymbol{m}$, and $\boldsymbol{\Sigma}$ each depends on $\boldsymbol{\vartheta}$.

Theorem 10. *For a constant Δ'', depending only on the dimension, we have*

$$B(\boldsymbol{v}) \leq \Delta'' \frac{1 + |\boldsymbol{v}|^2}{\mathfrak{c}^{(d+2)/2}} G.$$

Proof. Temporarily, let $h(\boldsymbol{x}) = f(\boldsymbol{x})\mathbf{1}_C(\boldsymbol{x})$ and, for a fixed $\boldsymbol{\vartheta} \in \mathcal{K}$, let

$$B_* = \left|\sqrt{(2\pi)^d \det[\boldsymbol{\Sigma}]}\int h(\boldsymbol{x})\mathfrak{N}(\mathrm{d}\boldsymbol{x} + \boldsymbol{m} - \boldsymbol{v}) - \int h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right|.$$

Then

$$\begin{aligned}
B_* &= \left|\int h(\boldsymbol{x} + \boldsymbol{v})(\exp(-\tfrac{1}{2}\boldsymbol{x}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{x}) - 1)\,\mathrm{d}\boldsymbol{x}\right| \\
&\leq \frac{1}{2}\int |f(\boldsymbol{x} + \boldsymbol{v})|(\boldsymbol{x}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&\leq \frac{1}{2\mathfrak{c}}\int |f(\boldsymbol{x})||\boldsymbol{x} - \boldsymbol{v}|^2\,\mathrm{d}\boldsymbol{x} \\
&\leq \frac{1}{2\mathfrak{c}}\int g(|\boldsymbol{x}|)|\boldsymbol{x}|^2\,\mathrm{d}\boldsymbol{x} + \frac{|\boldsymbol{v}|^2}{2\mathfrak{c}}\int g(|\boldsymbol{x}|)\,\mathrm{d}\boldsymbol{x} \\
&\leq \frac{1 + |\boldsymbol{v}|^2}{2\mathfrak{c}}\mathfrak{S}G
\end{aligned}$$

and, so,

$$\frac{B_*}{\sqrt{(2\pi)^d \det[\boldsymbol{\Sigma}]}} \leq \frac{B_*}{\sqrt{(2\pi)^d \mathfrak{c}^d}} \leq \frac{\mathfrak{S}}{2\sqrt{(2\pi)^d}}\frac{1 + |\boldsymbol{v}|^2}{\mathfrak{c}^{(d+2)/2}}G,$$

as required. $\qquad\square$

It is worth, very briefly, relating these results to those in Chaganty and Sethuraman (1993). Temporarily following the notation used there, let T_n be a univariate random variable with moment-generating function $\exp(n\psi_n(z))$. Suppose that ψ_n is analytic, bounded in n and z on Ω, and there has a second derivative bounded below by α. Let $\{\tau_n\}$ be a positive, bounded sequence inside Ω_- and suppose that, for any $a > 0$,

$$\varepsilon_n(a) = a\sup\left\{\left|\frac{\exp(n\psi_n(\tau_n + \mathrm{i}t))}{\exp(n\psi_n(\tau_n))}\right| : a \leq |t| < \frac{1}{a}\right\} = o(n^{-1/2}).$$

These are the conditions in Chaganty and Sethuraman (1993, Theorem 3.3). Under these conditions we can take $\mathfrak{c} = n\alpha$ and $\mathfrak{u} = nU$, and $\mathfrak{r}$ is bounded.

Take f to be a function of $x \in \mathbb{R}$ with a bounded derivative and with $f(x) = 0$ for $x \in (-\infty, -1]$ and $f(x) = \mathrm{e}^{-\tau x}$ for $x \in [0, \infty)$. For $\tau \in (0, B]$, it is routine calculus to show that there exists a constant C such that f is in both $\mathfrak{G}'(C/\tau)$ and $\mathfrak{G}(C/\tau^3)$. By applying

Theorems 9 and 10 with this f, the convex set C being $(0, \infty)$ and $\mathcal{K}$ being the point set $\{\tau_n\}$ when estimating the nth distribution, we obtain

$$\frac{\mathrm{P}((T_n - n\psi_n'(\tau_n)) \in (0, \infty))}{\exp(n\psi_n(\tau_n) - n\tau_n\psi'(\tau_n))} - \frac{1}{\tau_n\sqrt{2\pi n\psi_n''(\tau_n)}} = o\left(\frac{1}{\tau_n\sqrt{n}}\right),$$

which contains the conclusion of Chaganty and Sethuraman (1993, Theorem 3.3). Clearly, a multivariate version of this result would also follow directly from the discussion here, as would a result uniform in τ in compact subsets of $(0, \infty) \cap \Omega_-$. Also, when sums of independent, identically distributed variables are considered, when $\psi_n(z)$ does not depend on n, the result of Stone (1967, Theorem 3) is easily derived (for the nonlattice case).

6. Proof of Theorem 7 and Corollary 2

Proof of Theorem 7. Note first that $\mathrm{P}(\mathcal{W}_i(\boldsymbol{\theta}) = 0)$ is a fixed point of the multivariate generating function of the underlying Galton–Watson process. Hence, when the martingale converges in mean, so that $\mathrm{E}\,\mathcal{W}_i(\boldsymbol{\theta}) = 1$, these probabilities must be less than 1 and, so, must equal the extinction probabilities from that starting type. Thus, $\mathcal{W}_i(\boldsymbol{\theta}) > 0$ agrees with the survival set almost surely for $\boldsymbol{\theta} \in \Lambda_-$. The continuity of $\mathcal{W}_i(\boldsymbol{\theta})$ now means that the null set can be chosen independently of $\boldsymbol{\theta}$. For the rest of the calculation, we deal with sample paths in $\mathcal{S}$. Then there exists an N such that $\mathcal{W}_{ij}^n(\boldsymbol{\theta}) > 0$ for all $n \geq N$.

We take $n \geq N$. Let $\mathcal{K}$ be a compact subset of Λ_- and fix j. Let the function w^n be defined by

$$w^n(\boldsymbol{\lambda}) = \log\left(\frac{\mathcal{W}_{ij}^n(\boldsymbol{\lambda})v_i(\boldsymbol{\lambda})}{v_j(\boldsymbol{\lambda})}\right).$$

Using Theorems 1(ii), 3, and 5, there exists an open B (with $\Lambda_- \subset B \subset \Lambda$) such that, for each n, w^n is analytic in $\boldsymbol{\lambda} \in B$, $\{w^n\}$ converges as $n \to \infty$ to an analytic function on B, and, for some τ,

$$\mathcal{K}_\tau = \{\boldsymbol{\theta} + \boldsymbol{\lambda} \in \mathbb{C}^d : \boldsymbol{\theta} \in \mathcal{K}, \, |\boldsymbol{\lambda}| \leq \tau\} \subset B.$$

Analyticity of w^n and its limit on B mean that all its derivatives are uniformly bounded on $\mathcal{K}_\tau$.

To make the connection with the previous section, for fixed j and n let the measure Z be Z_{ij}^n and, for fixed $\boldsymbol{\theta}$, let

$$\nu(\mathrm{d}\boldsymbol{x}) = \frac{\mathrm{e}^{-\boldsymbol{\theta}^\top \boldsymbol{x}}Z(\mathrm{d}\boldsymbol{x})}{\int \mathrm{e}^{-\boldsymbol{\theta}^\top \boldsymbol{y}}Z(\mathrm{d}\boldsymbol{y})}.$$

Note that, with $f(\boldsymbol{x}) = \mathrm{e}^{\boldsymbol{\theta}^\top \boldsymbol{x}}h(\boldsymbol{x})$ and $\boldsymbol{v}_1 \equiv \boldsymbol{v}_1(\boldsymbol{\theta})$,

$$\mathrm{e}^{n\xi(\boldsymbol{\theta})}\int_C h(\boldsymbol{x})Z_{ij}^n(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1(\boldsymbol{\theta})) = \frac{\mathrm{e}^{-n\boldsymbol{\theta}^\top \boldsymbol{v}_1(\boldsymbol{\theta})}}{\rho(\boldsymbol{\theta})^n}\int_C f(\boldsymbol{x})\mathrm{e}^{-\boldsymbol{\theta}^\top \boldsymbol{x}}Z_{ij}^n(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1(\boldsymbol{\theta}))$$

$$= \frac{\int \mathrm{e}^{-\boldsymbol{\theta}^\top \boldsymbol{x}}Z_{ij}^n(\mathrm{d}\boldsymbol{x})}{\rho(\boldsymbol{\theta})^n}\int_C f(\boldsymbol{x})\nu(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1)$$

$$= \frac{v_i(\boldsymbol{\theta})\mathcal{W}_{ij}^n(\boldsymbol{\theta})}{v_j(\boldsymbol{\theta})}\int_C f(\boldsymbol{x})\nu(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1).$$

Now, by Theorem 5, $v_i(\boldsymbol{\theta})\mathcal{W}_{ij}^n(\boldsymbol{\theta})/v_j(\boldsymbol{\theta})$ converges to $v_i(\boldsymbol{\theta})u_j(\boldsymbol{\theta})\mathcal{W}_i(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta} \in \mathcal{K}$. By assumption, $f \in \mathfrak{G}(G)$, and, so, the results of the previous section can be applied to

consider the convergence of $\int_C f(\boldsymbol{x}) \nu(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1)$, once we show that Z satisfies the appropriate conditions.

For $\boldsymbol{\lambda} \in B$, substitution shows that

$$\log(\widehat{Z}(\boldsymbol{\lambda})) - n\log(\rho(\boldsymbol{\lambda})) = w^n(\boldsymbol{\lambda}).$$

It follows from the uniform boundedness of the derivatives of w^n that $\{\boldsymbol{m} - n\boldsymbol{v}_1 : n\}$ and $\{\boldsymbol{\Sigma} - n\boldsymbol{v}_2 : n\}$ are bounded uniformly in $\boldsymbol{\theta} \in \mathcal{K}$. Furthermore, there are constants $l > 0$ and $L < \infty$ such that, for all sufficiently large n, we have $\mathfrak{c}/n \geq l$ and $\mathfrak{u}/n \leq L$ and, thus, $\mathfrak{r} \leq L/l$.

For any $\epsilon > 0$,

$$\varepsilon(\epsilon) = \epsilon^{-d} \sup\left\{ \left| \frac{\widehat{Z}(\boldsymbol{\theta} + i\boldsymbol{\eta})}{\widehat{Z}(\boldsymbol{\theta})} \right| : \boldsymbol{\theta} \in \mathcal{K}, \, \epsilon \leq |\boldsymbol{\eta}| < \frac{1}{\epsilon} \right\}$$

$$= \epsilon^{-d} \sup\left\{ \frac{v_j(\boldsymbol{\theta})}{v_i(\boldsymbol{\theta})\mathcal{W}_{ij}^n(\boldsymbol{\theta})} \left| \frac{\mathcal{M}_{ij}^n(\boldsymbol{\theta} + i\boldsymbol{\eta})}{\rho(\boldsymbol{\theta})^n} \right| : \boldsymbol{\theta} \in \mathcal{K}, \, \epsilon \leq |\boldsymbol{\eta}| < \frac{1}{\epsilon} \right\}$$

$$\to 0$$

geometrically quickly as $n \to \infty$, using Theorems 5 and 6. Hence, using Theorem 9, $n^{d/2} A(Z_{ij}^n) \to 0$. Furthermore, since $\{\boldsymbol{m} - n\boldsymbol{v}_1 : n\}$ is uniformly bounded, Theorem 10 gives $n^{(d+2)/2} B(\boldsymbol{m} - n\boldsymbol{v_1}) \to 0$. Hence,

$$n^{d/2} \left| \int_C f(\boldsymbol{x}) \nu(\mathrm{d}\boldsymbol{x} + n\boldsymbol{v}_1) - \frac{1}{\sqrt{(2\pi)^d \det[\boldsymbol{v}_2]}} \int_C f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| \to 0$$

uniformly in $\boldsymbol{\theta} \in \mathcal{K}$ and $C \in \mathfrak{C}$.

Proof of Corollary 2. In Theorem 7, take $f \in \mathfrak{D}$ to be 1 on $|\boldsymbol{x}| \leq b$ and 0 on $|\boldsymbol{x}| > b + 1$, with bounded derivatives. The corollary then follows.

7. Extended Perron–Frobenius

A matrix $\boldsymbol{M} = \{m_{ij}\}_{p \times p}$, has nth power $\boldsymbol{M}^n$ with entries denoted by m_{ij}^n. The eigenvalues of $\boldsymbol{M}$ are the zeros of the characteristic polynomial $q(z) = \det[z\boldsymbol{I} - \boldsymbol{M}]$, which is of degree p. Denote the roots of $q(z)$ by $\rho_1, \ldots, \rho_p$, with the roots listed in order so that

$$|\rho_1| \geq |\rho_2| \geq \cdots \geq |\rho_p|.$$

The eigenvalue ρ_1 of $\boldsymbol{M}$ is the maximum-modulus eigenvalue if $|\rho_1| > |\rho_2|$, and is then denoted by ρ. Let the entries $m_{ij}(\boldsymbol{\lambda})$ of the matrix $\boldsymbol{M}(\boldsymbol{\lambda}) = \{m_{ij}(\boldsymbol{\lambda})\}_{p \times p}$ be functions of $\boldsymbol{\lambda} \in L$. Clearly, the eigenvalues and their multiplicities all depend on $\boldsymbol{\lambda}$.

Proof of Theorem 1. Parts (i) and (ii) of Theorem 1 arise from routine applications of the implicit function theorem; details can be found in Biggins and Rahimzadeh Sani (2004). Similar results have been obtained before, particularly for matrices of Laplace transforms; see, for example, Miller (1961, Theorem 1(a)), Lancaster and Tismenetsky (1985, Theorem 11.5.1), Ney and Numelin (1987, Theorem 4.1), and Kontoyiannis and Meyn (2003, Proposition 4.8(iii)). Now, by part (i), there is an open set containing $\tilde{L}$ on which $\rho(\boldsymbol{\lambda}) = |\rho_1(\boldsymbol{\lambda})| > |\rho_2(\boldsymbol{\lambda})|$, where $\rho = \rho_1$ (the eigenvalues here depend on $\boldsymbol{\lambda}$ but this is left implicit in the notation). For the rest of the proof, $\boldsymbol{\lambda}$ will be confined to this open set.

The resolvent of $M(\lambda)$ is defined by

$$R(z) = \{r_{ij}(z)\}_{p \times p} = (I - zM(\lambda))^{-1},$$

which, for all i and j, has the expansion $r_{ij}(z) = \sum_{n=0}^{\infty} z^n m_{ij}^n(\lambda)$ when $z < \|M(\lambda)\|^{-1}$, where $\| \cdot \|$ is a matrix norm (see Lancaster and Tismenetsky (1985, Theorem 11.1.1)). For a fixed $\tilde{\lambda} \in \tilde{L}$, let $B = B(\tilde{\lambda}, \tilde{\delta})$ and M_1 be such that, for all $\lambda \in B$, $\|M(\lambda)\| < M_1$. Now we take $|z| < 1/M_1$ and $\lambda \in B$, and suppress λ in the notation.

Let

$$h(z) = \det[I - zM] = (1 - z\rho) \prod_{k=2}^{p} (1 - z\rho_k).$$

Inverting $I - zM$ and rewriting using partial fractions, we obtain

$$r_{ij}(z) = \frac{d_{ij}(z)}{(1 - z\rho) \prod_{k=2}^{p}(1 - z\rho_k)} = \frac{a_{ij}}{1 - z\rho} + \frac{\sum_{k=0}^{p-2} b_{k,ij} z^k}{\prod_{k=2}^{p}(1 - z\rho_k)}, \tag{25}$$

where $d_{ij}(z)$ are polynomials in z of degree at most $(p-1)$, given by the appropriate entry in the adjoint matrix of $I - zM$, and a_{ij} and $b_{k,ij}$ are defined in such a way that (25) holds. For fixed λ, $|b_{k,ij}/\rho^k|$ are bounded over i, j, and k (with $k \le p - 2$) by C, say. Then, expansion of (25) gives

$$r_{ij}(z) = \sum_{n=0}^{\infty} \left(a_{ij}\rho^n + \sum_{h=0}^{p-2} b_{h,ij} \sum_{k_2 + \cdots + k_p = n-h} \rho_2^{k_2} \cdots \rho_p^{k_p} \right) z^n \tag{26}$$

and, so,

$$m_{ij}^n - a_{ij}\rho^n = \sum_{h=0}^{p-2} b_{h,ij} \sum_{k_2 + \cdots + k_p = n-h} \rho_2^{k_2} \cdots \rho_p^{k_p}.$$

Hence, for $\alpha \in [|\rho_2|/|\rho|, 1)$, we have

$$|\rho^{-n} m_{ij}^n - a_{ij}| \le C(p-1)(n+1)^{p-1} \alpha^{n-p}. \tag{27}$$

Let A be the matrix with entries $\{a_{ij}\}$. Then $\rho^{-n}M^n \to A$ and, so, $MA = AM = \rho A$. Parts (i) and (ii) of the theorem now give $A = cvu^\top$ for some scalar c; however,

$$v = \rho^{-n}M^n v \to Av = cv,$$

using $u^\top v = 1$, and, so, $c = 1$.

To translate (26) with $a_{ij} = v_i u_j$ into an asymptotic estimate of $M(\lambda)^n$, the boundedness of $b_{k,ij}$ as λ varies is now needed. Let $c_1(B)$ denote the continuously differentiable functions on B. The function $h(z)$ is a polynomial in z with coefficients in $c_1(B)$ and $1/\rho$ is a simple root of this polynomial. Hence, $h_1(z) = h(z)/(1 - \rho z)$ is a polynomial in z with coefficients in $c_1(B)$; the same is true of $d_{ij}(z)$. From (25) and $A = vu^\top$, we have

$$d_{ij}(z) = v_i u_j h_1(z) + \sum_{k=0}^{p-2} b_{k,ij} z^k (1 - z\rho).$$

Equating powers of z here shows that the $b_{k,ij}$ are in $c_1(B)$. Hence, the supremum of $|b_{k,ij}(\boldsymbol{\lambda})/\rho(\boldsymbol{\lambda})^k|$ over i, j, k, and $\boldsymbol{\lambda}$ (with $k \leq p - 2$) in the closed ball of radius $\frac{1}{2}\tilde{\delta}$ centred at $\tilde{\boldsymbol{\lambda}}$ will be finite; denote this finite supremum by C.

Let 3ϵ be less than $|\rho(\tilde{\boldsymbol{\lambda}})| - |\rho_2(\tilde{\boldsymbol{\lambda}})|$ and small enough that balls of radius 3ϵ centred on the distinct eigenvalues of $\boldsymbol{M}(\tilde{\boldsymbol{\lambda}})$ are disjoint. Then any point within a distance ϵ of one of $\rho_2(\tilde{\boldsymbol{\lambda}}), \ldots, \rho_p(\tilde{\boldsymbol{\lambda}})$ is smaller in magnitude than every point in the ϵ-ball centred on $\rho(\tilde{\boldsymbol{\lambda}})$. Let $\delta < \frac{1}{2}\tilde{\delta}$ be small enough to ensure that the maximum distance between the eigenvalues at $\tilde{\boldsymbol{\lambda}}$ and those at $\boldsymbol{\lambda} \in B(\tilde{\boldsymbol{\lambda}}, \delta)$, after a suitable pairing, is less than ϵ. Then, for $\boldsymbol{\lambda} \in B(\tilde{\boldsymbol{\lambda}}, \delta)$ and $j = 2, 3, \ldots, p$,

$$\left| \frac{\rho_j(\boldsymbol{\lambda})}{\rho(\boldsymbol{\lambda})} \right| \leq \frac{|\rho_2(\tilde{\boldsymbol{\lambda}})| + \epsilon}{\rho(\tilde{\boldsymbol{\lambda}}) - \epsilon} \leq \frac{\rho(\tilde{\boldsymbol{\lambda}}) - 2\epsilon}{\rho(\tilde{\boldsymbol{\lambda}}) - \epsilon} = \alpha < 1.$$

With these definitions, the estimate (27) holds uniformly on $B(\tilde{\boldsymbol{\lambda}}, \delta)$, and this implies (2).

Although we do not need it here, it is worth noting the following result, which applies to suitable matrices of Fourier transforms. Its proof requires only obvious modifications of the discussion already given.

Theorem 11. *Theorem 1 remains true when 'analytic' is replaced by 'q-times differentiable* $(q \geq 1)$*' and '*$\mathbb{C}^d$*' is replaced by '*$\mathbb{R}^d$*'.*

Acknowledgements

References

ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, Berlin.

BARRAL, J. (2001). Generalized vector multiplicative cascades. *Adv. Appl. Prob.* **33,** 874–895.

BHATTACHARYA, R. N. (1972). Recent results on refinements of the central limit theorem. In *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* (Berkeley, CA, 1970/71), Vol. 2, University of California Press, Berkeley, CA, pp. 453–484.

BHATTACHARYA, R. N. (1977). Refinements of the multidimensional central limit theorem and applications. *Ann. Prob.* **5,** 1–27.

BIGGINS, J. D. (1976). The first- and last-birth problems for a multitype age-dependent branching process. *Adv. Appl. Prob.* **8,** 446–459.

BIGGINS, J. D. (1992). Uniform convergence of martingales in the branching random walk. *Ann. Prob.* **20,** 137–151.

BIGGINS, J. D. (1996). How fast does a general branching random walk spread? In *Classical and Modern Branching Processes* (IMA Vols Math. Appl. **84**), eds K. B. Athreya and P. Jagers. Springer, New York, pp. 19–40.

BIGGINS, J. D. AND KYPRIANOU, A. E. (2004). Measure change in multitype branching. *Adv. Appl. Prob.* **36,** 544–581.

BIGGINS, J. D. AND RAHIMZADEH SANI, A. (2004). Extended Perron–Frobenius results. Preprint 539/04, Department of Probability and Statistics, University of Sheffield. Available at: http://www.shef.ac.uk/~st1jdb/epfr.html.

BRAMSON, M., NEY, P. AND TAO, J. (1992). The population composition of a multitype branching random walk. *Ann. Appl. Prob.* **2,** 575–596.

CHAGANTY, N. R. AND SETHURAMAN, J. (1993). Strong large deviation and local limit theorems. *Ann. Prob.* **21,** 1671–1690.

HÖRMANDER, L. (1973). *An Introduction to Complex Analysis in Several Variables*. North-Holland, Amsterdam.

JENSEN J. L. (1995). *Saddlepoint Approximations*. Oxford University Press.

KINGMAN, J. F. C. (1961). A convexity property of positive matrices. *Quart. J. Math. Oxford* **12,** 283–284.

KONTOYIANNIS, Y. AND MEYN, S. P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Prob.* **13,** 304–362.

KYPRIANOU, A. E. AND RAHIMZADEH SANI A. (2001). Martingale convergence and the functional equation in the multi-type branching random walk. *Bernoulli* **7,** 593–604.

LANCASTER, P. AND TISMENETSKY, M. (1985). *The Theory of Matrices*, 2nd edn. Academic Press, New York.

MILLER, H. D. (1961). A convexity property in the theory of random variables defined on a finite Markov chain. *Ann. Math. Statist*. **32,** 1260–1270.

MODE, C. J. (1971). *Multitype Branching Processes. Theory and Applications.* American Elsevier, New York.

NEY, P. AND NUMELIN, E. (1987). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Prob.* **15,** 561–592.

SENETA, E. (1973). *Non-Negative Matrices.* Allen and Unwin, London.

STONE, C. (1965). A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.* **36,** 546–551.

STONE, C. (1967). On local and ratio limit theorems. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* (Berkeley, CA, 1965/66), Vol. 2, University of California Press, Berkeley, CA, pp. 217–224.

VON BAHR, B. (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7,** 71–88.

VON BAHR, B. AND ESSEEN, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$. *Ann. Math. Statist.* **36,** 299–303.