University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Peköz, EA;Röllin, A;Ross, N

Title:
Exponential and laplace approximation for occupation statistics of branching random walk

Date:
2020-01-01

Citation:
Peköz, E. A., Röllin, A. & Ross, N. (2020). Exponential and laplace approximation for occupation statistics of branching random walk. Electronic Journal of Probability, 25 (none), https://doi.org/10.1214/20-EJP461.

Persistent Link:
https://hdl.handle.net/11343/290321

License:
CC BY

Electron. J. Probab. **25** (2020), article no. 55, 1–22.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP461

Exponential and Laplace approximation for occupation statistics of branching random walk

Erol A. Peköz [*] Adrian Röllin[†] Nathan Ross[‡]

Abstract

We study occupancy counts for the critical nearest-neighbor branching random walk on the d-dimensional lattice, conditioned on non-extinction. For $d \geqslant 3$, Lalley and Zheng [4] showed that the properly scaled joint distribution of the number of sites occupied by j generation-n particles, $j = 1, 2, \ldots$, converges in distribution as n goes to infinity, to a deterministic multiple of a single exponential random variable. The limiting exponential variable can be understood as the classical Yaglom limit of the total population size of generation n. Here we study the second order fluctuations around this limit, first, by providing a rate of convergence in the Wasserstein metric that holds for all $d \geqslant 3$, and second, by showing that for $d \geqslant 7$, the weak limit of the scaled joint differences between the number of occupancy-j sites and appropriate multiples of the total population size converge in the Wasserstein metric to a multivariate symmetric Laplace distribution. We also provide a rate of convergence for this latter result.

Keywords: branching random walk; distributional approximation; exponential distribution; multivariate symmetric Laplace distribution.
AMS MSC 2010: 60F05; 60J80.
Submitted to EJP on September 4, 2019, final version accepted on April 20, 2020.
Supersedes arXiv:1909.01617.

1 Introduction

Branching random walk (BRW) is a fundamental mathematical model of a population evolving in time and space, which has been intensely studied for more than 50 years due to its connection to population genetics and superprocesses; see, e.g., [1, Chapter 9] and references. Among this literature, the most relevant to our study is [4, Theorem 5],

[*]Boston University. http://people.bu.edu/pekoz/
E-mail: pekoz@bu.edu
[†]National University of Singapore. https://blog.nus.edu.sg/roellin/
E-mail: adrian.roellin@nus.edu.sg
[‡]University of Melbourne. https://www.nathanrossprob.com
E-mail: nathan.ross@unimelb.edu.au

which states that the exponential distribution arises asymptotically for certain occupation statistics of a critical BRW conditioned on non-extinction. Their result is closely related to the classical theorem of Yaglom [13], which says that the distribution of the size of a critical Galton-Watson process, properly scaled and conditioned on non-extinction, converges to the exponential distribution. Yaglom's theorem has a large related literature of embellishments and extensions, e.g., [5] and [2] give elegant probabilistic proofs, and [7] gives a rate of convergence using Stein's method.

We now define the nearest neighbor critical BRW on the d-dimensional integer lattice. At each time step $n = 1, 2, \ldots$, every particle generates an independent number of offspring having distribution X with $\mathbb{E}[X] = 1$, $\mathrm{Var}(X) = \sigma^2 < \infty$, and each offspring moves to a site randomly chosen from the $2d + 1$ sites having distance less than or equal to 1 from the site of its parent. We say that a site has *multiplicity j in the nth generation* if there are exactly j particles from the nth generation at that site. Starting the process from a single particle at the origin, let Z_n be the number of particles in nth generation, and let $M_n(j)$ be the total number of multiplicity j sites in the nth generation. Lalley and Zheng [4, Theorem 5] showed that, when $d \geqslant 3$, there are constants $\kappa_1, \kappa_2 \ldots$ with $\sum_{j \geqslant 1} j\kappa_j = 1$ such that, as $n \to \infty$,

$$\mathscr{L}\left(\frac{Z_n}{n}, \frac{M_n(1)}{n}, \frac{M_n(2)}{n}, \ldots \middle| Z_n > 0\right) \to \mathscr{L}\big((1, \kappa_1, \kappa_2, \ldots)Z\big), \tag{1.1}$$

where $Z \sim \mathrm{Exp}(\sigma^2/2)$ is exponential rate $\sigma^2/2$, and the convergence is with respect to the product topology (which is the same as convergence of finite dimensional distributions). We study the second order fluctuations in this limit, working with the finite dimensional distributions of (1.1) in the L_1-Wasserstein metric. More precisely, let $\|\cdot\|_1$ be the L_1-norm on $\mathbb{R}^r$, let

$$\mathcal{H}_r = \{h : \mathbb{R}^r \to \mathbb{R} : |h(x) - h(y)| \leqslant \|x - y\|_1 \text{ for every } x, y \in \mathbb{R}^r\}$$

be the set of L_1-Lipschitz continuous functions with constant 1, and define

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathscr{L}(V)) := \sup_{h \in \mathcal{H}_r} |\,\mathbb{E}[h(W) - h(V)]|. \tag{1.2}$$

Our first main result is as follows. Below and throughout the paper, we use c to represent constants that do not depend on n, but possibly $\mathscr{L}(X)$, the dimension d, and the length of the vector r, and can differ from line to line. We also disregard the pathological case where $\mathrm{Var}(X) = 0$.

Theorem 1.1. *With the definitions above, and any $r \geqslant 1$, $d \geqslant 3$, and offspring variable X satisfying $\mathbb{E}[X] = 1, \mathbb{E}[X^3] < \infty, \mathrm{Var}(X) = \sigma^2$, there are positive constants $\kappa_1, \ldots, \kappa_r$, so that for $Z \sim \mathrm{Exp}(\sigma^2/2)$,*

$$d_{\mathrm{W}}\left(\mathscr{L}\left(\frac{Z_n}{n}, \frac{M_n(1)}{n}, \ldots, \frac{M_n(r)}{n}\middle| Z_n > 0\right), \mathscr{L}\Big((1, \kappa_1, \ldots, \kappa_r)Z\Big)\right) \leqslant cn^{-\frac{d-2}{2(d+1)}},$$

for some constant c independent of n.

Our next result refines Theorem 1.1 when $d \geqslant 7$ and the offspring distribution has finite higher moments. To state the result, we define the r-dimensional symmetric Laplace distribution with covariance matrix Σ, denoted $\mathrm{SL}_r(\Sigma)$, as follows: If $E \sim \mathrm{Exp}(1)$ is independent of $\boldsymbol{Z}$, which is a centered multivariate normal vector with covariance matrix Σ, then $\mathscr{L}(\sqrt{E}\boldsymbol{Z}) = \mathrm{SL}_r(\Sigma)$. More important for our purposes is that the symmetric Laplace distribution arises as the scaled limit of a geometric random sum of i.i.d. centered random variables; see Section 2.1 for more details.

Theorem 1.2. *Recall the definitions above, and let $r \geqslant 1$, $d \geqslant 7$, and offspring variable X satisfy $\mathbb{E}[X] = 1$ and $\mathbb{E}\big[X^{5+\lfloor \frac{18}{d-6} \rfloor}\big] < \infty$. For the positive constants $\kappa_1, \dots, \kappa_r$ from Theorem 1.1, there exists a non-negative definite matrix $\widetilde{\Sigma}$, and a constant c independent of n, such that*

$$d_{\mathrm{W}}\left(\mathscr{L}\left(\frac{M_n(1)-\kappa_1 Z_n}{\sqrt{n}}, \dots, \frac{M_n(r)-\kappa_r Z_n}{\sqrt{n}} \Big| Z_n > 0\right), \mathrm{SL}_r\big(\widetilde{\Sigma}\big)\right) \leqslant c n^{-\frac{2d-9}{6(2d+1)}}.$$

Before discussing the ideas behind the proofs of these two theorems, we make a few remarks. The limiting covariance matrix $\widetilde{\Sigma}$ is a constant multiple ($\mathrm{Var}(X)/2$) of the limit of the (unconditional) covariance matrix of $(M_n(j) - Z_n \,\mathbb{E}[M_n(j)])_{j=1}^r$, given by Lemma 2.9 below. We are only able to show the limit exists, and cannot exclude the possibility that $\widetilde{\Sigma}$ is degenerate or even zero. Where Theorem 1.2 applies, it implies a rate of convergence of $n^{-1/2}$ in Theorem 1.1. However, we present the results in this way, as it is conceptually natural, and simplifies the presentation of the proofs. An interesting open question from our study is what is the minimal dimension for which the convergence in Theorem 1.2 occurs? The assumption that $d \geqslant 7$ stems from Lemma 2.9, giving the behavior of the covariance matrix. It may be possible to sharpen some estimates, e.g., (2.21), but there are others that may be sharp and still require $d \geqslant 7$, for example, the upper bound of (2.23), which must be $\mathrm{o}(1)$ (as $n \to \infty$) for our arguments to go through.

We now turn to a discussion of the proofs, where a key result is the following rate of convergence for Yaglom's theorem from [7].

Theorem 1.3. *([7, Theorem 3.3]) With the definitions above, and any offspring variable X satisfying $\mathbb{E}[X] = 1, \mathbb{E}[X^3] < \infty, \mathrm{Var}(X) = \sigma^2$, we have*

$$d_{\mathrm{W}}\big(\mathscr{L}\left(Z_n/n \,|\, Z_n > 0\right), \mathrm{Exp}(\sigma^2/2)\big) \leqslant \frac{c \log n}{n}.$$

With this result in hand, the basic intuition behind Theorems 1.1 and 1.2 is that the number of multiplicity j sites in generation n is approximately a sum of a random number of conditionally independent random variables. The number of summands is Z_m, for a well-chosen $m < n$, and a given summand represents the contribution from only descendants of a single generation m particle. If m is large, then Theorem 1.3 implies that Z_m will be roughly exponential with large mean; hence approximately geometrically distributed; and, if $d \geqslant 3$, the summands approximate the true variable because the random walk is transient, and most low occupancy sites consist of particles descended from exactly one individual in generation m; see Lemma 2.5. Thus the vector of Theorem 1.1 is approximately a geometric sum with small parameter, which, by Rényi's Theorem for geometric sums, is close to its mean times an exponential. For Theorem 1.2, the idea is similar, but the summands need to be centered for the Laplace limit to arise. A first thought is to subtract the mean of the summands, but in fact we must subtract the mean times a variable with mean one that is highly correlated to the summand to get the correct scaling; see Lemma 2.9. In order to obtain rates of convergence for the Laplace distribution, we prove a general approximation result for random sums, Theorem 2.8 below.

The approach of [4] used to obtain (1.1) uses a similar idea, but the conditioning is different to ours, and does not seem amenable to obtaining the error bounds necessary for Theorems 1.1 and 1.2. Here we use couplings via an explicit construction of $(M_n(j)|Z_n > 0)$, which is an elaboration of [5], along with Theorem 1.3, to evaluate the bounds necessary to obtain Theorems 1.1 and 1.2. We also use the explicit representation in a novel way to compare two different conditionings appearing in our argument; see Lemma 2.3. The use of the Wasserstein metric is essential in our argument, even if Kolmogorov bounds are the eventual goal (via standard smoothing arguments).

The organization of the paper is as follows. In the next section we provide constructions and lemmas used to prove Theorems 1.1 and 1.2. In Section 2.1 we state and prove our general Laplace approximation result, and then apply it to prove Theorem 1.2. Section 3 gives some auxiliary multivariate normal approximation results that are adapted to our setting, and used in the proofs.

2 Constructions, moment bounds and proofs

To prove Theorems 1.1 and 1.2, we first need to relate $\mathscr{L}(Z_m|Z_n>0)$ to $\mathscr{L}(Z_m|Z_m>0)$ (in Lemma 2.3 below). We use the size-biased tree construction from [5].

Size-biased tree construction. Assume that the tree is labeled and ordered, so if w and v are vertices in the tree from the same generation and w is to the left of v, then the offspring of w is to the left of the offspring of v, too. Start in Generation 0 with one vertex v_0 and let it have a number of offspring distributed according to a size-biased version X^s of X, so that

$$\mathbb{P}(X^s=k)=k\,\mathbb{P}(X=k)/\,\mathbb{E}[X].$$

Pick one of the offspring of v_0 uniformly at random and call it v_1. To each of the siblings of v_1, attach an independent Galton-Watson branching process with offspring distribution X. For v_1 proceed as for v_0, that is, give it a size-biased number of offspring, pick one at uniformly at random, call it v_2, attach independent Galton-Watson branching process to the siblings of v_2 and so on. For $1\leqslant j\leqslant n$, denote by $L_{n,j}$ and $R_{n,j}$ the number of particles in generation n of this tree that are descendants of the siblings of v_j to the left and right (excluding v_j). This gives rise to the size-biased tree.

From this construction we define another tree T_n as follows. For $1\leqslant j\leqslant n$, let $R'_{n,j}$ be independent random variables with

$$\mathscr{L}(R'_{n,j})=\mathscr{L}(R_{n,j}|L_{n,j}=0).$$

Start with a single "marked" particle in generation 0, represented as the root vertex of T_n, and give this particle $R'_{n,1}$ offspring. Then choose the leftmost offspring of the marked particle as the generation 1 marked particle, and give it $R'_{n,2}$ offspring. To continue, the generation j marked particle is the leftmost offspring of the marked particle in Generation $j-1$, and has $R'_{n,j+1}$ offspring. In addition, every non-marked particle has descendants according to an independent Galton-Watson tree with offspring distribution $\mathscr{L}(X)$. Let T_n be the tree generated in this way to generation n. The key fact from [5, Theorem C(i)], is that the distribution of T_n is the same as the *entire tree* created from an ordinary Galton-Watson process with offspring distribution X conditional on non-extinction up to Generation n. Moreover, this tree and its marked particles can be closely coupled to the size-biased tree and the v_j "spine" offspring.

Now, let $A_{k,j}=\{L_{k,j}=0\}$ and let $X_{m,j},X'_{m,j}$ be random variables that are independent of each other and the size-biased tree constructed above, such that

$$\mathscr{L}(X_{m,j})=\mathscr{L}(R_{m,j}+L_{m,j}|A_{m,j}),\qquad \mathscr{L}(X'_{m,j})=\mathscr{L}(R_{m,j}+L_{m,j}|A_{n,j}),$$

and let

$$\begin{aligned}Y_{m,j}&=R_{m,j}+L_{m,j}+(X_{m,j}-R_{m,j}-L_{m,j})\mathrm{I}[A^c_{m,j}],\\ Y'_{m,j}&=R_{m,j}+L_{m,j}+(X'_{m,j}-R_{m,j}-L_{m,j})\mathrm{I}[A^c_{n,j}].\end{aligned}\tag{2.1}$$

Define also $Y_m=1+\sum_{j=1}^m Y_{m,j}$ and $Y'_m=1+\sum_{j=1}^m Y'_{m,j}$. We have the following result.

Lemma 2.1. *Using the notation and definitions of the size-biased tree construction, for $j=1,\dots,m$ and $k=m,n$, we have*

$$(i)\ \ \mathscr{L}(Y_m)=\mathscr{L}(Z_m|Z_m>0);\ \ \mathscr{L}(Y'_m)=\mathscr{L}(Z_m|Z_n>0);$$

(ii) *there is a constant* K *such that if* $n-m>K$, *then*

$$\mathbb{E}\big[(L_{m,j}+R_{m,j})\mathrm{I}[A_{n,j}\setminus A_{m,j}]\big]\leqslant c\bigg[\frac{(n-m)^2}{(n-j+1)^2}+\frac{\log(m-j+2)}{m-j+1}\bigg];$$

(iii) $\mathbb{E}[(R_{m,j}+L_{m,j})|A_{k,j}]\leqslant c$;

(iv) $\mathbb{E}\big[X_{m,j}\mathrm{I}[A^c_{m,j}]\big]\leqslant\dfrac{c}{m-j+1}$, *and* $\mathbb{E}\big[X'_{m,j}\mathrm{I}[A^c_{n,j}]\big]\leqslant\dfrac{c}{n-j+1}$.

Before proving the lemma, we state a key result for controlling the conditioning on non-extinction is the following second order version of "Kolmogorov's estimate" found in [11, Display between (5) and (6)].

Lemma 2.2 (Kolmogorov's estimate)**.** *If the offspring variable* X *satisfies* $\mathbb{E}[X]=1$, $\mathbb{E}[X^3]<\infty$, *and* $\mathrm{Var}(X)=\sigma^2$, *then as* $n\to\infty$,

$$\mathbb{P}(Z_n>0)=\frac{2}{n\sigma^2}+\mathrm{O}\bigg(\frac{\log^2(n)}{n^2}\bigg).$$

In particular, $n\,\mathbb{P}(Z_n>0)\to 2/\sigma^2$.

Proof of Lemma 2.1. Statement (i) follows from the key fact stated above from [5, Proof of Theorem C(i)]. For (ii), let X_j denote the number of siblings of v_j, distributed as $\mathscr{L}(X^s-1)$, $I_j\sim\mathrm{U}\{1,\dots,X_j+1\}$ be the position of v_j in its siblings (labeled left to right), and for $i=1,\dots,I_j-1$, let $L^{(i)}_{m,j}$ be the number of offspring in generation m descended from the ith sibling (labeled from the left) of v_j and $A^{(i)}_{k,j}$ be the event $\{L^{(i)}_{k,j}=0\}$; $k=m,n$. Then

$$\begin{aligned}
&\mathbb{E}\big[L_{m,j}\mathrm{I}[A_{n,j}\setminus A_{m,j}]\big]\\
&\quad\leqslant\mathbb{E}\Big[\sum_{i,\ell=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(\ell)}_{n,j}\setminus A^{(\ell)}_{m,j}]\Big]\\
&\quad=\mathbb{E}\Big[\sum_{i=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\Big]+\mathbb{E}\Big[\sum_{\substack{i,\ell=1\\ \ell\neq i}}^{I_j-1}\mathbb{E}\big[L^{(i)}_{m,j}|I_j,X_j\big]\,\mathbb{P}(A^{(\ell)}_{n,j}\setminus A^{(\ell)}_{m,j}|I_j,X_j)\Big]\\
&\quad\leqslant\mathbb{E}\Big[\sum_{i=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\Big]+\mathbb{P}(Z_{m-j}>0)\,\mathbb{E}\Big[\sum_{\substack{i,\ell=1\\ \ell\neq i}}^{I_j-1}\mathbb{E}\big[L^{(i)}_{m,j}|I_j,X_j\big]\Big]\\
&\quad\leqslant\mathbb{E}\Big[\sum_{i=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\Big]+\mathbb{P}(Z_{m-j}>0)\,\mathbb{E}\big[X_j\,\mathbb{E}[L_{m,j}|X_j]\big]\\
&\quad\leqslant\mathbb{E}\Big[\sum_{i=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\Big]+\mathbb{P}(Z_{m-j}>0)\,\mathbb{E}[X_j^2]\\
&\quad\leqslant\mathbb{E}\Big[\sum_{i=1}^{I_j-1}L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\Big]+\frac{c}{m-j+1},
\end{aligned}$$

where the first inequality is a union bound, the equality is by independence of lineages, the second inequality is because $\mathscr{L}(L^{(i)}_{m,j})=\mathscr{L}(Z_{m-j})$, and the last is by Lemma 2.2. To bound further, we have that, conditional on X_j,I_j, using the independence of lineages,

$$\begin{aligned}
&\mathbb{E}\big[L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}\setminus A^{(i)}_{m,j}]\big]\\
&\quad=\mathbb{E}\Big[L^{(i)}_{m,j}\mathrm{I}[A^{(i)}_{n,j}]\big|L^{(i)}_{m,j}>0\Big]\,\mathbb{P}(L^{(i)}_{m,j}>0)
\end{aligned}$$

$$= \mathbb{E}\Big[\frac{L^{(i)}_{m,j}}{m-j+1}\Big(\mathbb{P}(Z_{n-m}=0)^{m-j+1}\Big)^{\frac{L^{(i)}_{m,j}}{m-j+1}}\big|L^{(i)}_{m,j}>0\Big]\,\mathbb{P}(Z_{m-j}>0)(m-j+1)$$
$$\leqslant c\,\mathbb{E}\Big[\frac{L^{(i)}_{m,j}}{m-j+1}\Big(\mathbb{P}(Z_{n-m}=0)^{m-j+1}\Big)^{\frac{L^{(i)}_{m,j}}{m-j+1}}\big|L^{(i)}_{m,j}>0\Big],$$

where we have used Lemma 2.2. The function $g(x) = xa^x$ is 1-Lipschitz on $(0,\infty)$ for any $a<1$, so we can apply Theorem 1.3 and use the fact that $\mathscr{L}(L^{(i)}_{m,j}) = \mathscr{L}(Z_{m-j})$ to find that for $W\sim \mathrm{Exp}(2/\sigma^2)$,

$$\begin{aligned}
&\mathbb{E}\Big[\frac{L^{(i)}_{m,j}}{m-j+1}\Big(\mathbb{P}(Z_{n-m}=0)^{m-j+1}\Big)^{\frac{L^{(i)}_{m,j}}{m-j+1}}\big|L^{(i)}_{m,j}>0\Big]\\
&\quad\leqslant \mathbb{E}\Big[W\Big(\mathbb{P}(Z_{n-m}=0)^{m-j+1}\Big)^{W}\Big]+c\frac{\log(m-j+1)}{m-j+1}\\
&\quad= c\bigg[\Big(2\sigma^{-2}-(m-j+1)\log\big(\mathbb{P}(Z_{n-m}=0)\big)\Big)^{-2}+\frac{\log(m-j+1)}{m-j+1}\bigg]\\
&\quad\leqslant c\bigg[\Big(2\sigma^{-2}+(m-j+1)\,\mathbb{P}(Z_{n-m}>0)\Big)^{-2}+\frac{\log(m-j+1)}{m-j+1}\bigg]\\
&\quad\leqslant c\bigg[\Big(1+\frac{m-j+1}{n-m}-c'\frac{(m-j+1)(\log(n-m)^2}{(n-m)^2}\Big)^{-2}+\frac{\log(m-j+1)}{m-j+1}\bigg],
\end{aligned}$$

where we have used Lemma 2.2, and we now choose K so that $c'\log(n-m)^2/(n-m)<1/2$ whenever $n-m>K$. Using this bound and combining the last three displays, we have

$$\begin{aligned}
&\mathbb{E}\big[L_{m,j}\mathrm{I}[A_{n,j}\setminus A_{m,j}]\big]\\
&\quad\leqslant c\bigg\{\mathbb{E}[X_j]\Big[\Big(1+\frac{m-j+1}{n-m}-c'\frac{(m-j+1)\log(n-m)^2}{(n-m)^2}\Big)^{-2}\\
&\qquad\qquad+\frac{\log(m-j+1)}{m-j+1}\Big]+\frac{1}{m-j+1}\bigg\}\\
&\quad\leqslant c\bigg\{\bigg(1+\Big(\frac{1}{2}\Big)\Big(\frac{m-j+1}{n-m}\Big)\bigg)^{-2}+\frac{\log(m-j+1)}{m-j+1}+\frac{1}{m-j+1}\bigg\}\\
&\quad\leqslant c\bigg\{\Big(1+\frac{m-j+1}{n-m}\Big)^{-2}+\frac{\log(m-j+2)}{m-j+1}\bigg\}.
\end{aligned}\tag{2.2}$$

Now noting that given I_j and X_j, $R_{m,j}$ is independent of $A_{m,j}$ and $A_{n,j}$, we easily find from Lemma 2.2 that

$$\mathbb{E}\big[R_{m,j}\mathrm{I}[A_{n,j}\setminus A_{m,j}]\big]\leqslant \mathbb{E}\Big[\mathbb{E}\big[R_{m,j}|X_j,I_j\big]\,\mathbb{P}\big(A_{n,j}\setminus A_{m,j}|X_j,I_j\big)\Big].$$

A union bound implies

$$\mathbb{P}\big(A_{n,j}\setminus A_{m,j}|X_j,I_j\big)\leqslant \mathbb{P}\big(A^c_{m,j}|X_j,I_j\big)\leqslant X_j\,\mathbb{P}(Z_{m-j}>0),$$

and clearly $\mathbb{E}[R_{m,j}|X_j,I_j]\leqslant X_j$, so altogether, using Lemma 2.2,

$$\mathbb{E}\big[R_{m,j}\mathrm{I}[A_{n,j}\setminus A_{m,j}]\big]\leqslant \mathbb{E}[X_j^2]\,\mathbb{P}(Z_{m-j}>0)\leqslant\frac{c}{m-j+1}.$$

Combining this with (2.2) shows (ii).

For (iii), we show that for $k=m,n$,

$$\mathbb{E}\big[(R_{m,j}+L_{m,j})\mathrm{I}[A_{k,j}]\big]\leqslant c\,\mathbb{E}[X_j]\,\mathbb{P}(A_{k,j}),\tag{2.3}$$

which easily implies the result. To show (2.3), we use the following correlation inequality: if f is non-decreasing and g is non-increasing, then $\mathrm{Cov}(f(X), g(X)) \leqslant 0$. Note that

$$\begin{aligned} 0 = \mathbb{E}\big[L_{m,j}\mathrm{I}[A_{m,j}]|X_j, I_j\big] &\leqslant \mathbb{E}\big[L_{m,j}|X_j, I_j\big]\,\mathbb{E}\big[\mathrm{I}[A_{m,j}]|X_j, I_j\big], \\ \mathbb{E}\big[R_{m,j}\mathrm{I}[A_{m,j}]|X_j, I_j\big] &\leqslant \mathbb{E}\big[R_{m,j}|X_j, I_j\big]\,\mathbb{E}\big[\mathrm{I}[A_{m,j}]|X_j, I_j\big], \end{aligned}$$

where the first line is obvious and the second is because of conditional independence. Since $\mathbb{E}[R_{m,j} + L_{m,j}|X_j, I_j] = X_j$, we then have

$$\mathbb{E}\big[(R_{m,j} + L_{m,j})\mathrm{I}[A_{m,j}]|X_j, I_j\big] \leqslant X_j\,\mathbb{P}\big(A_{m,j}|X_j, I_j\big) = X_j\,\mathbb{E}\big[\mathbb{P}(Z_{m-j} = 0)^{I_j - 1}|X_j\big].$$

But we can couple I_j to X_j in such a way that it is non-decreasing with X_j, and thus the correlation inequality implies

$$\mathbb{E}\Big[X_j\,\mathbb{E}\big[\mathbb{P}(Z_{m-j} = 0)^{I_j - 1}|X_j\big]\Big] \leqslant \mathbb{E}[X_j]\,\mathbb{E}\big[\mathbb{P}(Z_{m-j} = 0)^{I_j - 1}\big] = \mathbb{E}[X_j]\,\mathbb{P}(A_{m,j}).$$

Combining the last two displays implies (2.3) for $k = m$. For $k = n$, using similar ideas,

$$\mathbb{E}\big[L_{m,j}\mathrm{I}[A_{n,j}]|L_{m,j}\big] = L_{m,j}\,\mathbb{P}(Z_{n-m} = 0)^{L_{m,j}},$$

and the second factor decreases with $L_{m,j}$ and so

$$\mathbb{E}\big[L_{m,j}\mathrm{I}[A_{n,j}]\big] \leqslant \mathbb{E}[L_{m,j}]\,\mathbb{P}(A_{n,j}) \leqslant \mathbb{E}[X_j]\,\mathbb{P}(A_{n,j}).$$

Finally,

$$\mathbb{E}\big[R_{m,j}\mathrm{I}[A_{n,j}]|X_j, I_j\big] = (X_j + 1 - I_j)\,\mathbb{P}(Z_{n-j} = 0)^{I_j - 1} \leqslant X_j\,\mathbb{P}(Z_{n-j} = 0)^{I_j - 1}.$$

But again we can couple (X_j, I_j) such that I_j is non-decreasing in X_j, and thus

$$\mathbb{E}\big[R_{m,j}\mathrm{I}[A_{n,j}]\big] \leqslant \mathbb{E}[X_j]\,\mathbb{E}\big[P(Z_{n-j} = 0)^{I_j - 1}\big] = \mathbb{E}[X_j]\,\mathbb{P}(A_{n,j}).$$

Finally, to show (iv), we have

$$\begin{aligned} \mathbb{E}\big[X_{m,j}\mathrm{I}[A^c_{m,j}]\big] &= \mathbb{E}[X_{m,j}]\,\mathbb{P}(A^c_{m,j}) = \mathbb{E}[R_{m,j} + L_{m,j}|A_{m,j}]\,\mathbb{P}(A^c_{m,j}), \\ \mathbb{E}\big[X'_{m,j}\mathrm{I}[A^c_{n,j}]\big] &= \mathbb{E}[X'_{m,j}]\,\mathbb{P}(A^c_{n,j}) = \mathbb{E}[R_{m,j} + L_{m,j}|A_{n,j}]\,\mathbb{P}(A^c_{n,j}), \end{aligned}$$

and then (iv) easily follows from (iii) and

$$\mathbb{P}(A^c_{k,j}) = \mathbb{E}\big[\mathbb{P}\big(A^c_{k,j}|X_j\big)\big] \leqslant \mathbb{E}[X_j]\,\mathbb{P}(Z_{k-j} > 0) \leqslant \frac{c}{k-j+1}. \qquad \square$$

Lemma 2.3. *If Z_k denotes the size of the kth generation of a Galton-Watson tree with offspring distribution X satisfying $\mathbb{E}[X] = 1$ and $\mathbb{E}[X^3] < \infty$, and if $m \leqslant n$ with $n - m \to \infty$, then*

$$d_{\mathrm{W}}(\mathscr{L}(Z_m|Z_m > 0), \mathscr{L}(Z_m|Z_n > 0)) \leqslant c\big[(n-m) + \log^2(m)\big].$$

Proof of Lemma 2.3. The coupling definition of Wasserstein distance and Lemma 2.1, parts $(i), (ii), (iv)$, yield an upper bound of

$$\begin{aligned} \mathbb{E}|Y'_m - Y_m| &\leqslant \sum_{j=1}^{m} \mathbb{E}\big[X_{m,j}\mathrm{I}[A^c_{m,j}] + X'_{m,j}\mathrm{I}[A^c_{n,j}]\big] \\ &\qquad + \sum_{j=1}^{m} \mathbb{E}\big[(R_{m,j} + L_{m,j})\mathrm{I}[A^c_{m,j} \cap A_{n,j}]\big] \\ &\leqslant c\sum_{j=1}^{m}\left(\frac{1}{m-j+1} + \frac{(n-m)^2}{(n-j+1)^2} + \frac{\log(m-j+2)}{m-j+1}\right) \\ &\leqslant c\big[(n-m) + \log^2(m)\big]. \end{aligned}$$

$\square$

To continue, we need a lemma giving some moment information for variables in the BRW.

Lemma 2.4. *Assume the definitions and constructions above, and let $Y_{n;m}$ denote the number of particles in generation n of the BRW that occupy a site with another generation-n particle that has a different ancestor at generation m. We have the following:*

1. *The limit $\kappa_j := \lim_{n\to\infty} \mathbb{E}[M_n(j)]$ exists, and*
$$|\,\mathbb{E}[M_n(j)] - \kappa_j| \leqslant cn^{1-d/2};$$
2. $\mathrm{Var}(M_n(j)) \leqslant 1 + n\sigma^2$;
3. $\mathbb{E}[Y_{n;m}|Z_n > 0] \leqslant cn(1 + m\sigma^2)(n-m)^{-d/2}$.

Proof of Lemma 2.4. Part 1 follows from the proof of [4, Proposition 21], where they obtain
$$|\,\mathbb{E}[M_{n+1}(j)] - \mathbb{E}[M_n(j)]| \leqslant cn^{-d/2},$$
which means $\mathbb{E}[M_{n+1}(j)], n = 1, 2, \dots$ has a limit and the bound follows since
$$\sum_{i=n}^{\infty} i^{-d/2} \leqslant cn^{1-d/2}.$$
Part 2 follows from $M_n(j) \leqslant Z_n, \mathbb{E}[Z_n] = 1$ and $\mathrm{Var}(Z_n) = n\sigma^2$ to get $\mathrm{Var}(M_n(j)) \leqslant \mathbb{E}[Z_n^2] \leqslant 1 + n\sigma^2$. To prove Part 3, we have
$$\mathbb{E}[Y_{n;m}|Z_n > 0] = \frac{\mathbb{E}\big[Y_{n;m}\mathrm{I}[Z_n > 0]\big]}{\mathbb{P}(Z_n > 0)} \leqslant \frac{\mathbb{E}[Y_{n;m}]}{\mathbb{P}(Z_n > 0)} \leqslant cn(n-m)^{-d/2}\,\mathbb{E}[Z_m^2],$$
where in the last inequality we have used Lemma 2.2 and [4, Corollary 20]. The lemma follows after noting that $\mathrm{Var}(Z_m) = m\sigma^2$ and $\mathbb{E}[Z_m] = 1$. □

We give a construction of the critical BRW, building from the size-biased tree construction.

BRW construction. To construct $M_n(j)$ conditional on $Z_n > 0$, first generate T_n from the size-bias tree section above. Since this tree is distributed as the Galton-Watson tree given $Z_n > 0$, we construct the conditional BRW by attaching a random direction to each offspring, chosen uniformly and independently from the $2d + 1$ available directions for the nearest-neighbor random walk. It is obvious that this "modified" BRW process has the same distribution as the original conditioned on non-extinction to generation n. For the modified process, let $\hat{Z}_k$ denote the size of generation k and $\hat{M}_n(j)$ be the number of multiplicity j sites in generation n, then, in particular, we have
$$\mathscr{L}\big(Z_m, Z_n, (M_n(j))_{j=1}^r | Z_n > 0\big) = \mathscr{L}\big(\hat{Z}_m, \hat{Z}_n, (\hat{M}_n(j))_{j=1}^r\big).$$

Modified BRW construction. A key to our approach is the following lemma that shows the cost of replacing $\hat{M}_n(i)$ by a sum of a random sum of conditionally independent variables. Given $\hat{Z}_m$, for $i = 2, \dots, \hat{Z}_m$, let $Z_{n,m}^i$ be the number of generation n offspring of the ith particle in generation m of the modified BRW construction; here the labelling is left to right (so particle 1 is always the marked particle), and note these are distributed as the sizes of the $(n-m)$th generations of i.i.d. Galton-Watson trees with offspring distribution $\mathscr{L}(X)$. Let also $M_{n,m}^i(j)$ be the number of sites having exactly j generation-n descendants from the generation m particle labeled i in the critical BRW construction above, where the counts ignore particles descended from other generation m particles at those sites. Also let $(Z_{n,m}^1, M_{n,m}^1(j)$ be an independent copy of $(Z_{n-m}, M_{n-m}(j))$. Note that given $\hat{Z}_m$, $M_{n,m}^1(j), \dots, M_{n,m}^{\hat{Z}_m}(j)$ are i.i.d.

Lemma 2.5. *For the variables described above and $m < n$,*

$$\mathbb{E}\Big|\hat{M}_n(j) - \sum_{i=1}^{\hat{Z}_m} M^i_{n,m}(j)\Big| \leqslant c\big(nm(n-m)^{-d/2} + (n-m)\big).$$

Proof. The differences between the two variables are (i) multiplicity j sites with more than 1 ancestor from generation m, (ii) multiplicity $k > j$ sites with exactly j particles descended from some single generation m particle, (iii) the number of multiplicity j sites with only descendants of the first particle of generation m, and (iv) $M^1_{n,m}(j)$. But (i) and (ii) together are bounded by $Y_{n;m}$ (from Lemma 2.4) and (iii) is bounded by $\sum_{i=m+1}^n R'_{n,i}$. Thus, using using Items 1 and 3 of Lemma 2.4 and *(iii)* of Lemma 2.1, we have

$$\begin{aligned}\mathbb{E}\Big|\hat{M}_n(j) - \sum_{i=1}^{\hat{Z}_m} M^i_{n,m}(j)\Big| &\leqslant \mathbb{E}[Y_{n;m}|Z_n > 0] + \sum_{i=m+1}^n \mathbb{E}[R'_{n,i}] + \mathbb{E}[M_{n-m}(j)] \\ &\leqslant c\big(n(1+m\sigma^2)(n-m)^{-d/2} + (n-m)\big).\end{aligned}$$

□

Before proving Theorem 1.1, we state and prove a simple lemma.

Lemma 2.6. *For any nonnegative random variable Y on the same space as $(Z_j)_{0\leqslant j\leqslant n}$ and $m < n$, we have*

$$\mathbb{E}[Y|Z_n > 0] \leqslant \frac{cn}{m}\,\mathbb{E}[Y|Z_m > 0]. \tag{2.4}$$

Proof. Using Kolmogorov's approximation,

$$\begin{aligned}&\mathbb{E}[Y|Z_n > 0] \\ &\quad= \frac{\mathbb{E}[Y1_{Z_n>0}]}{\mathbb{P}(Z_n>0)} \leqslant \frac{\mathbb{E}[Y1_{Z_m>0}]}{\mathbb{P}(Z_n>0)} = \mathbb{E}[Y|Z_m>0]\frac{\mathbb{P}(Z_m>0)}{\mathbb{P}(Z_n>0)} \leqslant \frac{cn}{m}\,\mathbb{E}[Y|Z_m>0].\end{aligned}$$

□

Proof of Theorem 1.1. Fix $n \geqslant 4$ and

$$1 \leqslant m := \big\lfloor n - n^{3/(d+1)}\big\rfloor = \big\lfloor n\big(1 - n^{-(d-2)/(d+1)}\big)\big\rfloor < n.$$

The triangle inequality implies

$$\begin{aligned}&d_{\mathrm{W}}\left(\mathscr{L}\left(\frac{Z_n}{n}, \frac{M_n(1)}{n}, \dots, \frac{M_n(r)}{n}\Big|Z_n>0\right), \mathscr{L}\big((1,\kappa_1,\dots,\kappa_r)Z\big)\right) \\ &\quad\leqslant d_{\mathrm{W}}\left(\mathscr{L}\left(\frac{Z_n}{n}, \frac{M_n(1)}{n}, \dots, \frac{M_n(r)}{n}\Big|Z_n>0\right), \mathscr{L}\left((1,\kappa_1,\dots,\kappa_r)\frac{Z_m}{m}\Big|Z_n>0\right)\right) \qquad (2.5) \\ &\qquad+ d_{\mathrm{W}}\left(\mathscr{L}\left((1,\kappa_1,\dots,\kappa_r)\frac{Z_m}{m}\Big|Z_n>0\right), \mathscr{L}\big((1,\kappa_1,\dots,\kappa_r)Z\big)\right). \qquad (2.6)\end{aligned}$$

Using Lemma 2.3 and Theorem 1.3, we find (2.6) is upper bounded by $rc(n-m)/m = \mathrm{O}(n^{\frac{2-d}{d+1}})$.

From the coupling definition of the Wasserstein metric, (2.5) is upper bounded by

$$\mathbb{E}\left|\frac{\hat{Z}_n}{n} - \frac{\hat{Z}_m}{m}\right| + \sum_{j=1}^r \mathbb{E}\left|\frac{\hat{M}_n(j)}{n} - \frac{\kappa_j\hat{Z}_m}{m}\right| \leqslant b_1 + b_2 + \sum_{j=1}^r (e_{1,j} + e_{2,j}),$$

where the hat-couplings are those in the BRW description above and

$$b_1 = \frac{1}{n}\,\mathbb{E}\big|\hat{Z}_n - \hat{Z}_m\big|, \qquad\qquad b_2 = \frac{n-m}{nm}\,\mathbb{E}[\hat{Z}_m],$$

$$e_{1,j} = \frac{1}{n}\,\mathbb{E}\left|\hat{M}_n(j) - \sum_{i=1}^{\hat{Z}_m} M^i_{n,m}(j)\right|, \qquad e_{2,j} = \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{\hat{Z}_m} M^i_{n,m}(j) - \frac{\kappa_j \hat{Z}_m}{m}\right|.$$

For b_1, we have that

$$\left|\hat{Z}_n - \hat{Z}_m\right| \leqslant \sum_{i=m+1}^{n} R'_{n,i} + \left|\sum_{i=2}^{\hat{Z}_m}\big(Z^i_{n,m} - 1\big)\right|,$$

where recall the $Z^i_{n,m}$ are the number of generation n offspring of the ith particle in generation m of the modified BRW construction, which are distributed as the sizes of the $(n-m)$th generations of i.i.d. Galton-Watson trees with offspring distribution $\mathscr{L}(X)$. Thus, using (iii) of Lemma 2.1, conditioning on $\hat{Z}_m$, and using the fact that $\mathbb{E}[Z^i_{n,m}] = 1$, $\mathrm{Var}(Z^i_{n,m}) = (n-m)\sigma^2$, Cauchy-Schwarz, and then Jensen's inequality, we have

$$b_1 \leqslant \frac{c}{n}\Big(n - m + \mathbb{E}\Big[\sqrt{(n-m)\sigma^2\hat{Z}_m}\Big]\Big) \leqslant \frac{c}{n}\Big(n - m + \sqrt{(n-m)\sigma^2}\sqrt{\mathbb{E}[\hat{Z}_m]}\Big) \leqslant cn^{\frac{2-d}{2(d+1)}}.$$

Now, using (2.4) which says $\mathbb{E}[\hat{Z}_m] = \mathbb{E}[Z_m | Z_n > 0] \leqslant cn$, we find $b_2 = \mathrm{O}(n^{\frac{2-d}{d+1}})$. And Lemma 2.5 implies that $e_{1,j} = \mathrm{O}(n^{\frac{2-d}{2(d+1)}})$.

For $e_{2,j}$, recall that conditional on $\hat{Z}_m$, $M^1_{n,m}(j), M^2_{n,m}(j), \ldots$ are i.i.d. random variables all having the same distribution as $M_{n-m}(j)$. Thus, letting Z_m be distributed as the size of the mth generation in a Galton-Watson tree with offspring distribution X that is independent of the $M^i_{n,m}(j)$, we find

$$\begin{aligned}
\frac{m}{cn}e_{2,j} &\leqslant \mathbb{E}\bigg\{\mathbb{E}\Big[\Big|n^{-1}\sum_{i=1}^{Z_m} M^i_{n,m}(j) - \kappa_j Z_m/m\Big|\Big|Z_m, Z_m > 0\Big]\Big|Z_m > 0\bigg\} \\
&\leqslant \mathbb{E}\bigg\{\mathbb{E}\Big[\big(n^{-1}\sum_{i=1}^{Z_m} M^i_{n,m}(j) - \kappa_j Z_m/m\big)^2\Big|Z_m, Z_m > 0\Big]^{1/2}\Big|Z_m > 0\bigg\} \\
&= \mathbb{E}\bigg\{\Big[n^{-2}Z_m\,\mathrm{Var}(M_{n-m}(j)) + \big(n^{-1}Z_m\,\mathbb{E}[M_{n-m}(j)] - \kappa_j Z_m/m\big)^2\Big]^{1/2}\Big|Z_m > 0\bigg\} \\
&\leqslant \mathbb{E}\Big[n^{-1}\sqrt{Z_m}\,\mathrm{Var}(M_{n-m}(j))^{1/2} + n^{-1}Z_m\big|\mathbb{E}[M_{n-m}(j)] - n\kappa_j/m\big|\big|Z_m > 0\Big] \\
&\leqslant n^{-1}\,\mathbb{E}\big[\sqrt{Z_m}|Z_n > 0\big]\,\mathrm{Var}(M_{n-m}(j))^{1/2} \\
&\quad + n^{-1}\,\mathbb{E}[Z_m|Z_n > 0]\big(|\,\mathbb{E}[M_{n-m}(j)] - \kappa_j| + \kappa_j(n-m)/m\big) \\
&\leqslant cn^{-\frac{d-2}{2(d+1)}},
\end{aligned}$$

where the last line follows by $\mathbb{E}[\sqrt{Z_m}|Z_n > 0] \leqslant \sqrt{\mathbb{E}[Z_m|Z_n > 0]} \leqslant cn^{1/2}$–using (2.4)–and Parts 1 and 2 of Lemma 2.4, which imply

$$\mathrm{Var}(M_{n-m}(j))^{1/2} \leqslant c\sqrt{n-m} \leqslant cn^{1.5/(d+1)} \text{ and } |\,\mathbb{E}[M_{n-m}(j)] - \kappa_j| \leqslant c(n-m)^{1-d/2}. \quad \square$$

2.1 Laplace distribution approximation

The centered multivariate symmetric Laplace distribution is a cousin to the Gaussian distribution that arises in a number of contexts and applications; see [3] for a book length treatment of this distribution. The r-dimensional distribution is denoted $\mathrm{SL}_r(\Sigma)$, where the parameter Σ is an $r \times r$ positive definite matrix. In general, its law is the same as that of $\sqrt{E}\boldsymbol{Z}$, where $E \sim \mathrm{Exp}(1)$ and $\boldsymbol{Z}$ is a centered multivariate normal vector with covariance matrix Σ. The covariance matrix of $\mathrm{SL}_r(\Sigma)$ is Σ, which can thus be thought of as a scaling parameter. The characteristic function is evidently

$$\mathbf{u} \mapsto \frac{1}{1 + \frac{1}{2}\mathbf{u}^\top\Sigma\mathbf{u}},$$

and from this it's easy to see a number of equivalent characterizations in the case $r = 1$: if E_1, E_2 are independent and $\mathrm{Exp}(1)$, then $E_1 - E_2 \sim \mathrm{SL}_1(2)$; if $B \sim \mathrm{Ber}(1/2)$ independent of E_1, E_2, then $BE_1 - (1-B)E_2 \sim \mathrm{SL}_1(2)$. The 1-dimensional density of $\mathrm{SL}_1(2)$ is $\frac{1}{2}e^{-|x|}$ and the multivariate density is given in [3, (5.2.2)] in terms of modified Bessel functions of the 3rd kind.

The symmetric Laplace distribution arises as the limit of a geometric sum. More precisely, we have the following theorem, which is elementary, using, for example, characteristic functions.

Theorem 2.7. *Let $N_p \sim \mathrm{Geo}(p)$ be independent of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, which are i.i.d. r-dimensional random vectors having mean zero and covariance matrix Σ. Then as $p \to 0$,*

$$p^{1/2} \sum_{i=1}^{N_p} \boldsymbol{X}_i \xrightarrow{d} \mathrm{SL}_r(\Sigma).$$

Here we provide a rate of convergence to a generalization of Theorem 2.7 in a metric amenable to our setting; see also [8] for a related result when $r = 1$.

Theorem 2.8. *Let $M \geqslant 1$ be a random variable with mean $\mu > 1$, independent of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, which are i.i.d. r-dimensional random vectors with zero mean, covariance matrix $\Sigma = (\Sigma_{ij})$, and finite third moments. Then there is a constant C_r depending only on r such that*

$$\begin{aligned} &d_{\mathrm{W}}\Big(\mathscr{L}\Big(\mu^{-1/2}\sum_{i=1}^{M}\boldsymbol{X}_i\Big), \mathrm{SL}_r(\Sigma)\Big) \\ &\quad\leqslant \mu^{-1/2}\Big(C_r\mu^{1/3}\,\mathbb{E}\big[\|\boldsymbol{X}_1\|_1^3\big]^{1/3} + \Big(\sum_{i=1}^{r}\sqrt{\Sigma_{ii}}\Big)d_{\mathrm{W}}\big(\mathscr{L}(M), \mathrm{Geo}(\mu^{-1})\big) + 3.5\Big). \end{aligned}$$

Proof. Let $E \sim \mathrm{Exp}(1), N \sim \mathrm{Geo}(\mu^{-1})$ and $\boldsymbol{Z} = \boldsymbol{Z}_\Sigma$ be a centered multivariate normal vector with covariance matrix Σ, with the three variables independent and independent of M and the $\boldsymbol{X}_i$. Then, since $\mathscr{L}(\sqrt{E}\boldsymbol{Z}) = \mathrm{SL}_r(\Sigma)$, the triangle inequality implies

$$\begin{aligned} d_{\mathrm{W}}\Big(\mathscr{L}\Big(\mu^{-1/2}\sum_{i=1}^{M}\boldsymbol{X}_i\Big), \mathrm{SL}_r(\Sigma)\Big) &\leqslant d_{\mathrm{W}}\Big(\mathscr{L}\Big(\mu^{-1/2}\sum_{i=1}^{M}\boldsymbol{X}_i\Big), \mathscr{L}\big(\sqrt{\mu^{-1}M}\boldsymbol{Z}\big)\Big) && (2.7)\\ &\quad + d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}M}\boldsymbol{Z}\big), \mathscr{L}\big(\sqrt{\mu^{-1}N}\boldsymbol{Z}\big)\Big) && (2.8)\\ &\quad + d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}N}\boldsymbol{Z}\big), \mathscr{L}\big(\sqrt{E}\boldsymbol{Z}\big)\Big). && (2.9) \end{aligned}$$

We use below that if X, Y, Z are random elements defined on the same space, then

$$d_{\mathrm{W}}\big(\mathscr{L}(X), \mathscr{L}(Y)\big) \leqslant \mathbb{E}\big[d_{\mathrm{W}}\big(\mathscr{L}(X|Z), \mathscr{L}(Y|Z)\big)\big], \tag{2.10}$$

which easily follows from the definition of Wasserstein distance given at (1.2).

To bound (2.7), we use the smooth function CLT, Corollary 3.2 below, which says that there is a constant C_r such that for all $m \geqslant 1$,

$$d_{\mathrm{W}}\Big(\mathscr{L}\Big(\mu^{-1/2}\sum_{i=1}^{m}\boldsymbol{X}_i\Big), \mathscr{L}\big(\sqrt{\mu^{-1}m}\boldsymbol{Z}\big)\Big) \leqslant C_r\mu^{-1/2}m^{1/3}\,\mathbb{E}\big[\|\boldsymbol{X}_1\|_1^3\big]^{1/3}.$$

Conditioning on M, applying (2.10), and using independence, we find that

$$\begin{aligned} d_{\mathrm{W}}\Big(\mathscr{L}\Big(\mu^{-1/2}\sum_{i=1}^{M}\boldsymbol{X}_i\Big), \mathscr{L}\big(\sqrt{\mu^{-1}M}\boldsymbol{Z}\big)\Big) &\leqslant C_r\mu^{-1/2}\,\mathbb{E}\big[M^{1/3}\big]\,\mathbb{E}\big[\|\boldsymbol{X}_1\|_1^3\big]^{1/3} \\ &\leqslant C_r\mu^{-1/2}\mu^{1/3}\,\mathbb{E}\big[\|\boldsymbol{X}_1\|_1^3\big]^{1/3}, \end{aligned} \tag{2.11}$$

where we have used Jensen's inequality in the last line.

To bound (2.8), we use (2.10) and the general fact that for fixed $\mathbf{z} \in \mathbb{R}^d$, and random variables X, Y,

$$d_{\mathrm{W}}\big(\mathscr{L}(X\mathbf{z}), \mathscr{L}(Y\mathbf{z})\big) \leqslant \|\mathbf{z}\|_1 d_{\mathrm{W}}\big(\mathscr{L}(X), \mathscr{L}(Y)\big). \tag{2.12}$$

Now use the dual definition of Wasserstein distance, see, for example [9] or [12], to choose a coupling between M and N such that $d_{\mathrm{W}}(\mathscr{L}(M), \mathscr{L}(N)) = \mathbb{E}\big[|M - N|\big]$. Using that $M, N \geqslant 1$, we have

$$\begin{aligned} d_{\mathrm{W}}\big(\mathscr{L}\big(\sqrt{M}\big), \mathscr{L}\big(\sqrt{N}\big)\big) &\leqslant \mathbb{E}\big|\sqrt{M} - \sqrt{N}\big| \leqslant \mathbb{E}\big[\big|\sqrt{M} - \sqrt{N}\big|\big(\sqrt{M} + \sqrt{N}\big)\big] \\ &= \mathbb{E}\big[|M - N|\big] = d_{\mathrm{W}}(\mathscr{L}(M), \mathscr{L}(N)). \end{aligned}$$

Thus, conditioning on $\boldsymbol{Z}$, using (2.10) and independence, we have that (2.12) implies

$$\begin{aligned} d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}M}\boldsymbol{Z}\big), \mathscr{L}\big(\sqrt{\mu^{-1}N}\boldsymbol{Z}\big)\Big) &\leqslant \mathbb{E}\big[\|\boldsymbol{Z}\|_1\big]\mu^{-1/2} d_{\mathrm{W}}\big(\mathscr{L}\big(\sqrt{M}\big), \mathscr{L}\big(\sqrt{N}\big)\big) \\ &\leqslant \mu^{-1/2}\Big(\sum_{i=1}^{r}\sqrt{\Sigma_{ii}}\Big) d_{\mathrm{W}}\big(\mathscr{L}(M), \mathscr{L}(N)\big), \end{aligned} \tag{2.13}$$

where the last inequality is because $\mathbb{E}\big[|Z_i|\big] \leqslant \sqrt{\Sigma_{ii}}$.

To bound (2.9), we again use (2.10), (2.12), and independence, to find

$$d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}N}\boldsymbol{Z}\big), \mathscr{L}\big(\sqrt{E}\boldsymbol{Z}\big)\Big) \leqslant \mathbb{E}\big[\|\boldsymbol{Z}\|_1\big] d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}N}\big), \mathscr{L}\big(\sqrt{E}\big)\Big).$$

To bound this last Wasserstein distance, we use the usual coupling of a geometric to an exponential. If $N' = \lceil E/(-\log(1-\mu^{-1}))\rceil$, then $\mathscr{L}(N) = \mathscr{L}(N')$, and

$$\begin{aligned} &d_{\mathrm{W}}\Big(\mathscr{L}\big(\sqrt{\mu^{-1}N}\big), \mathscr{L}\big(\sqrt{E}\big)\Big) \\ &\quad\leqslant \mathbb{E}\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big| \\ &\quad= \mathbb{E}\Big[\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big|\mathrm{I}[E < \mu^{-1}]\Big] + \mathbb{E}\Big[\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big|\mathrm{I}[E > \mu^{-1}]\Big]. \end{aligned} \tag{2.14}$$

Since $\sqrt{\mu^{-1}N'} \geqslant \mu - 1$, if $E > \mu^{-1}$, then

$$\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big| \leqslant \frac{\sqrt{\mu}}{2}\big|\mu^{-1}N' - E\big|,$$

and so

$$\begin{aligned} &\mathbb{E}\Big[\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big|\mathrm{I}[E > \mu^{-1}]\Big] \\ &\quad\leqslant \frac{\sqrt{\mu}}{2}\,\mathbb{E}\big|\mu^{-1}N' - E\big| \leqslant \frac{\sqrt{\mu}}{2}\,\mathbb{E}\Big|\frac{E}{\mu\log(1-\mu)} + \mu^{-1} - E\Big| \\ &\quad\leqslant \frac{\sqrt{\mu}}{2}\left(\Big|1 + \frac{1}{\mu\log(1-\mu)}\Big| + \mu^{-1}\right) \leqslant \mu^{-1/2}, \end{aligned} \tag{2.15}$$

since $|1 + 1/(\mu\log(1-\mu))| \leqslant 1$ for $\mu > 1$. Finally,

$$\begin{aligned} \mathbb{E}\Big[\Big|\sqrt{\mu^{-1}N'} - \sqrt{E}\Big|\mathrm{I}[E < \mu^{-1}]\Big] &\leqslant \left(\sqrt{\mu^{-1}\left(-\frac{1}{\mu\log(1-\mu)} + 1\right)} + \mu^{-1/2}\right)\mathbb{P}(E < \mu^{-1}) \\ &\leqslant \frac{\sqrt{2}+1}{\mu^{3/2}} \leqslant \frac{2.5}{\mu^{1/2}}, \end{aligned} \tag{2.16}$$

since $1 - 1/(\mu\log(1-\mu)) \leqslant 2$, $\mathbb{P}(E < \mu^{-1}) \leqslant \mu^{-1}$, and $\mu > 1$. Combining (2.14), (2.15), and (2.16) with (2.13) and (2.11) yields the result. □

2.2 Proof of Theorem 1.2

As in the proof of Theorem 1.1, we use Lemma 2.5 to move the problem to a random sum of i.i.d. random variables (with cost to the error term). Then we apply Theorem 2.8 to the random sum. Our first lemma gives some moment information for the summands, in particular, it shows the covariances converge, and the fourth moments are appropriately bounded.

Recall the notation and constructions for the BRW and modified BRW, and write $\mu_n(j) := \mathbb{E}[M_n(j)]$.

Lemma 2.9. *Assume the hypotheses of Theorem 1.2, and for $j,k \in \{1,\dots,r\}$, denote*

$$A_n(j,k) := \mathrm{Cov}\big(M_n(j) - \mu_n(j)Z_n, M_n(k) - \mu_n(k)Z_n\big).$$

Then the limits

$$\lim_{n\to\infty} A_n(j,k) =: \Sigma_{jk}$$

exist, are finite, and

$$|A_n(j,k) - \Sigma_{jk}| \leqslant cn^{2-d/3}.$$

Furthermore,

$$\lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}\big[(M_n(j) - \mu_n(j)Z_n)^4\big] = 3\sigma^2\Sigma_{jj}^2. \tag{2.17}$$

Remark 2.10. As a check on the limiting constant and linear growth of the fourth moment of $(M_n(j) - \mu_n(j)Z_n)$ given by (2.17), Theorem 1.2 suggests (assuming appropriate uniform integrability) that for $E \sim \mathrm{Exp}(1)$ and $Z_j \sim \mathrm{N}(0, \widetilde{\Sigma}_{jj})$,

$$n^{-2}\,\mathbb{E}\big[(M_n(j) - \mu_n(j)Z_n)^4 | Z_n > 0\big] \to \mathbb{E}[E^2]\,\mathbb{E}[Z_j^4] = 6\widetilde{\Sigma}_{jj}^2. \tag{2.18}$$

The left hand side of (2.18) is equal to

$$n^{-1}\frac{\mathbb{E}\big[(M_n(j) - \mu_n(j)Z_n)^4\big]}{n\,\mathbb{P}(Z_n > 0)} \sim n^{-1}\frac{\mathbb{E}\big[(M_n(j) - \mu_n(j)Z_n)^4\big]}{2/\sigma^2}. \tag{2.19}$$

From (2.28) below, we see that $\widetilde{\Sigma} = (\sigma^2/2)\Sigma$, which, with (2.18) and (2.19), agrees with (2.17).

Proof of Lemma 2.9. Write

$$A_n(j,k) = \mathbb{E}[M_n(j)M_n(k)] - \mu_n(k)\,\mathbb{E}[M_n(j)Z_n] - \mu_n(j)\,\mathbb{E}[M_n(k)Z_n] + \mu_n(j)\mu_n(k)\,\mathbb{E}[Z_n^2].$$

Let $(M_{n-1}^i(j), M_{n-1}^i(k), Z_{n-1}^i)_{i=1}^{Z_1}$ be defined as follows: The first (respectively, second) coordinate is the number of sites with j (respectively, k) particles in generation n descended from particle i in generation 1, and the third coordinate is the number of offspring in generation n descended from generation 1. Note that, given Z_1, these are i.i.d. copies of $(M_{n-1}(j), M_{n-1}(k), Z_{n-1})$. Now write

$$\begin{aligned}\mathbb{E}[M_n(j)M_n(k)] &= \mathbb{E}\bigg[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)M_{n-1}^\ell(k)\bigg] + \mathbb{E}\bigg[M_n(j)M_n(k) - \sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)M_{n-1}^\ell(k)\bigg]\\ &=: \mathbb{E}\bigg[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)M_{n-1}^\ell(k)\bigg] + e_n^{(1)}(j,k).\end{aligned}$$

The first term above is the main contribution, and the second term is a small error. For the first term,

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)M_{n-1}^\ell(k)\Big] \\
&\quad= \mathbb{E}\Big[\sum_{i,\ell=1}^{Z_1} \mathbb{E}\big[M_{n-1}^i(j)M_{n-1}^\ell(k)|Z_1\big]\Big] \\
&\quad= \mathbb{E}\Big[\sum_{i=1}^{Z_1} \mathbb{E}\big[M_{n-1}^i(j)M_{n-1}^i(k)|Z_1\big] + \sum_{i\neq j}^{Z_1} \mathbb{E}\big[M_{n-1}^i(j)M_{n-1}^j(k)|Z_1\big]\Big] \\
&\quad= \mathbb{E}[Z_1]\,\mathbb{E}\big[M_{n-1}^1(j)M_{n-1}^1(k)\big] + \mathbb{E}[Z_1(Z_1-1)]\mu_{n-1}(j)\mu_{n-1}(k) \\
&\quad= \mathbb{E}\big[M_{n-1}^1(j)M_{n-1}^1(k)\big] + \sigma^2\mu_{n-1}(j)\mu_{n-1}(k).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[M_n(j)Z_n] &= \mathbb{E}\Big[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)Z_{n-1}^\ell\Big] + \mathbb{E}\Big[Z_n\Big(M_n(j) - \sum_{i=1}^{Z_1} M_{n-1}^i(j)\Big)\Big], \\
&=: \mathbb{E}\Big[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)Z_{n-1}^\ell\Big] + e_n^{(2)}(j),
\end{aligned}$$

and

$$\mathbb{E}\Big[\sum_{i,\ell=1}^{Z_1} M_{n-1}^i(j)Z_{n-1}^\ell\Big] = \mathbb{E}\big[M_{n-1}^1(j)Z_{n-1}^1\big] + \sigma^2\mu_{n-1}(j).$$

Finally,

$$\mathbb{E}[Z_n^2] = \mathbb{E}[(Z_{n-1})^2] + \sigma^2.$$

Using that $\mathbb{E}[M_n(\ell)Z_n] \leqslant \mathbb{E}[Z_n^2] \leqslant cn$ and, from the proof of [4, Proposition 21], that $|\mu_n(\ell) - \mu_{n-1}(\ell)| \leqslant n^{-d/2}$, we can collect the work above to find

$$|A_n(j,k) - A_{n-1}(j,k)| \leqslant c\big(\big|e_n^{(1)}(j,k)\big| + \big|e_n^{(2)}(j)\big| + \big|e_n^{(2)}(k)\big| + n^{1-d/2}\big).$$

To bound the errors, first note

$$\begin{aligned}
\big|e_n^{(1)}(j,k)\big| &\leqslant \mathbb{E}\Big[M_n(j)\Big|M_n(k) - \sum_{\ell=1}^{Z_1} M_{n-1}^\ell(k)\Big|\Big] + \mathbb{E}\Big[\sum_{\ell=1}^{Z_1} M_{n-1}^\ell(k)\Big|M_n(j) - \sum_{i=1}^{Z_1} M_{n-1}^i(j)\Big|\Big] \\
&\leqslant \mathbb{E}\Big[Z_n\Big|M_n(k) - \sum_{\ell=1}^{Z_1} M_{n-1}^\ell(k)\Big|\Big] + \mathbb{E}\Big[Z_n\Big|M_n(j) - \sum_{i=1}^{Z_1} M_{n-1}^i(j)\Big|\Big] \leqslant 2\,\mathbb{E}[Z_nY_{n;1}].
\end{aligned} \tag{2.20}$$

Similarly,

$$|e_n^{(2)}(j)| \leqslant \mathbb{E}\Big[Z_n\Big|M_n(j) - \sum_{i=1}^{Z_1} M_{n-1}^i(j)\Big|\Big] \leqslant \mathbb{E}[Z_nY_{n;1}],$$

and the same inequality holds with $|e_n^{(2)}(k)|$ on the left hand side. For $\alpha > 0$, to be chosen later, we bound

$$\begin{aligned}
\mathbb{E}[Z_nY_{n;1}] &\leqslant \mathbb{E}\big[Z_nY_{n;1}\mathrm{I}[Z_n > n^{1+\alpha}]\big] + n^{1+\alpha}\,\mathbb{E}[Y_{n;1}] \\
&\leqslant \mathbb{E}\big[Z_n^2\mathrm{I}[Z_n > n^{1+\alpha}]\big] + cn^{1+\alpha-d/2},
\end{aligned} \tag{2.21}$$

where we have used [4, Corollary 20]. Now squaring both sides of the inequality in the indicator and bounding by the ratio, we have

$$\mathbb{E}\big[Z_n^2 \mathrm{I}[Z_n > n^{1+\alpha}]\big] \leqslant \frac{\mathbb{E}[Z_n^4]}{n^{2(1+\alpha)}} \leqslant cn^{1-2\alpha},$$

since $\mathbb{E}[Z_n^4] = \mathrm{O}(n^3)$. Thus, choosing $\alpha = d/6$, we obtain $\mathbb{E}[Z_n Y_{n;1}] \leqslant cn^{1-d/3}$, and thus

$$|A_n(j,k) - A_{n-1}(j,k)| \leqslant cn^{1-d/3}.$$

Since $d \geqslant 7$, $\sum_{n\geqslant 1} n^{1-d/3} < \infty$, and therefore $(A_n(j,k))_{n\geqslant 1}$ is a Cauchy sequence; denote its limit by Σ_{jk}, and observe that

$$|A_n(j,k) - \Sigma_{jk}| \leqslant cn^{2-d/3}.$$

To prove the second assertion, we fix j and drop it from the notation, e.g., writing M_n for $M_n(j)$. We follow the strategy above, but now there are higher moments, which we denote by

$$\mu_i^{(k,\ell)} := \mathbb{E}\big[(M_i)^k (Z_i)^\ell\big],$$

for $i \in \{n-1, n\}$. Now, expanding and then swapping the coefficients with powers of $\mu_n^{(0,1)}$ for those with $\mu_{n-1}^{(0,1)}$, we have

$$\begin{aligned}
\mathbb{E}\big[(M_n - \mu_n^{(1,0)} Z_n)^4\big] &= \mu_n^{(4,0)} - 4\mu_n^{(1,0)}\mu_n^{(3,1)} + 6\big(\mu_n^{(1,0)}\big)^2 \mu_n^{(2,2)} \\
&\quad - 4\big(\mu_n^{(1,0)}\big)^3 \mu_n^{(1,3)} + \big(\mu_n^{(1,0)}\big)^4 \mu_n^{(0,4)} \\
&= \mu_n^{(4,0)} - 4\mu_{n-1}^{(1,0)}\mu_n^{(3,1)} + 6\big(\mu_{n-1}^{(1,0)}\big)^2 \mu_n^{(2,2)} \\
&\quad - 4\big(\mu_{n-1}^{(1,0)}\big)^3 \mu_n^{(1,3)} + \big(\mu_{n-1}^{(1,0)}\big)^4 \mu_n^{(0,4)} + \mathrm{O}(n^{3-d/2}),
\end{aligned} \tag{2.22}$$

where the second equality follows, similar to the argument above, from

$$\big|\big(\mu_n^{(1,0)}\big)^k \mu_n^{(4-k,k)} - \big(\mu_{n-1}^{(1,0)}\big)^k \mu_n^{(4-k,k)}\big| \leqslant c\,\mathbb{E}[Z_n^4]|\mu_n^{(1,0)} - \mu_{n-1}^{(1,0)}| \leqslant cn^{3-d/2}, \tag{2.23}$$

where $k = 0, 1, \ldots, 4$. Now, denote the error made in the summands above when replacing the moments by the first step moments as

$$e_n^{(k,4-k)} := a_k\big(\mu_{n-1}^{(1,0)}\big)^k \bigg(\mu_n^{(k,4-k)} - \mathbb{E}\bigg[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^k \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^{4-k}\bigg]\bigg),$$

where a_k is $1, -4$, or 6 as appropriate. Bounding these similar to (2.20) and (2.21) above, using that $\mathbb{E}[X^{5+\lfloor 18/(d-6)\rfloor}] < \infty$, we have that for $\alpha > 0$ and $\beta = \lfloor 18/(d-6)\rfloor + 1$,

$$|e_n^{(k,4-k)}| \leqslant c\,\mathbb{E}[Z_n^3 Y_{n;1}] \leqslant c\big(\mathbb{E}[Z_n^{4+\beta}] n^{-\beta(1+\alpha)} + n^{3(1+\alpha)-d/2}\big)\big) \leqslant c(n^{3-\beta\alpha} + n^{3(1+\alpha)-d/2}).$$

Choosing $\alpha = (d-6)/6 - \varepsilon$, for $\varepsilon > 0$ small enough that $\alpha\beta > 3$, and noting that $d \geqslant 7$, we have

$$|e_n^{(k,4-k)}| = \mathrm{O}(n^{-\delta}),$$

for $\delta = \min\{\alpha\beta - 3, 3\varepsilon\} > 0$. Thus we find the fourth moment (2.22) is equal to

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^4\Big] - 4\mu_{n-1}^{(1,0)}\,\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^3 \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)\Big] \\
&\quad + 6\big(\mu_{n-1}^{(1,0)}\big)^2\,\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^2 \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^2\Big] - 4\big(\mu_{n-1}^{(1,0)}\big)^3\,\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)\Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^3\Big] \\
&\quad + \big(\mu_{n-1}^{(1,0)}\big)^4\,\mathbb{E}\Big[\Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^4\Big] + \mathrm{O}(n^{-\delta'}).
\end{aligned} \tag{2.24}$$

for $\delta' = \min\{\delta, 3 - d/2\} > 0$. As before, we expand the random sums in the expectations above and then simplify. We cover in detail only the middle term, which is the most involved, and just write the final expressions for the other terms. Write $\Sigma_{\{i,j,k\}}$ for sums over distinct indices. For the middle term,

$$\begin{aligned}
&\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^2 \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^2 \\
&\quad = \sum_{i_1,i_2,i_3,i_4=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} Z_{n-1}^{i_3} Z_{n-1}^{i_4} \\
&\quad = \sum_{i=1}^{Z_1} \big(M_{n-1}^i\big)^2 \big(Z_{n-1}^i\big)^2 + \sum_{\{i_1,i_2\}=1}^{Z_1} \big(M_{n-1}^{i_1}\big)^2 \big(Z_{n-1}^{i_2}\big)^2 \\
&\qquad + \sum_{\{i_1,i_2,i_3\}=1}^{Z_1} \big(M_{n-1}^{i_1}\big)^2 Z_{n-1}^{i_2} Z_{n-1}^{i_3} + 2 \sum_{\{i_1,i_2\}=1}^{Z_1} \big(M_{n-1}^{i_1}\big)^2 Z_{n-1}^{i_1} Z_{n-1}^{i_2} \\
&\qquad + \sum_{\{i_1,i_2,i_3,i_4\}=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} Z_{n-1}^{i_3} Z_{n-1}^{i_4} + 4 \sum_{\{i_1,i_2,i_3\}=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} Z_{n-1}^{i_1} Z_{n-1}^{i_3} \\
&\qquad + 2 \sum_{\{i_1,i_2\}=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} Z_{n-1}^{i_1} Z_{n-1}^{i_2} + 2 \sum_{\{i_1,i_2\}=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} \big(Z_{n-1}^{i_1}\big)^2 \\
&\qquad + \sum_{\{i_1,i_2,i_3\}=1}^{Z_1} M_{n-1}^{i_1} M_{n-1}^{i_2} \big(Z_{n-1}^{i_3}\big)^2.
\end{aligned}$$

For a quick parity check of this formula, note that for non-negative integer z,

$$z^4 = z(z-1)(z-2)(z-3) + 6z(z-1)(z-2) + 7z(z-1) + z.$$

Now, taking expectation by first conditioning on Z_1, writing $\gamma_k := \mathbb{E}[Z_1(Z_1-1)\cdots(Z_1-k+1)]$, $k = 3, 4$, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^2 \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^2\Big] &= \mu_{n-1}^{(2,2)} + \sigma^2 \mu_{n-1}^{(2,0)} \mu_{n-1}^{(0,2)} + \gamma_3 \mu_{n-1}^{(2,0)} + 2\sigma^2 \mu_{n-1}^{(2,1)} \\
&\quad + \gamma_4 \big(\mu_{n-1}^{(1,0)}\big)^2 + 4\gamma_3 \mu_{n-1}^{(1,1)} \mu_{n-1}^{(1,0)} + 2\sigma^2 \big(\mu_{n-1}^{(1,1)}\big)^2 \\
&\quad + 2\sigma^2 \mu_{n-1}^{(1,2)} \mu_{n-1}^{(1,0)} + \gamma_3 \big(\mu_{n-1}^{(1,0)}\big)^2 \mu_{n-1}^{(0,2)}.
\end{aligned}$$

Similar arguments shows

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^4\Big] &= \mu_{n-1}^{(4,0)} + 3\sigma^2 \big(\mu_{n-1}^{(2,0)}\big)^2 + 6\gamma_3 \mu_{n-1}^{(2,0)} \big(\mu_{n-1}^{(1,0)}\big)^2 \\
&\quad + 4\sigma^2 \mu_{n-1}^{(3,0)} \mu_{n-1}^{(1,0)} + \gamma_4 \big(\mu_{n-1}^{(1,0)}\big)^4, \\
\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)^3 \Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)\Big] &= \mu_{n-1}^{(3,1)} + 3\gamma_3 \mu_{n-1}^{(1,1)} \big(\mu_{n-1}^{(1,0)}\big)^2 + 3\sigma^2 \mu_{n-1}^{(2,1)} \mu_{n-1}^{(1,0)} \\
&\quad + 3\sigma^2 \mu_{n-1}^{(1,1)} \mu_{n-1}^{(2,0)} + \sigma^2 \mu_{n-1}^{(3,0)} + 3\gamma_3 \mu_{n-1}^{(2,0)} \mu_{n-1}^{(1,0)} + \gamma_4 \big(\mu_{n-1}^{(1,0)}\big)^3, \\
\mathbb{E}\Big[\Big(\sum_{i=1}^{Z_1} M_{n-1}^i\Big)\Big(\sum_{j=1}^{Z_1} Z_{n-1}^j\Big)^3\Big] &= \mu_{n-1}^{(1,3)} + 3\gamma_3 \mu_{n-1}^{(1,1)} + 3\sigma^2 \mu_{n-1}^{(1,2)} + 3\sigma^2 \mu_{n-1}^{(1,1)} \mu_{n-1}^{(0,2)}
\end{aligned}$$

$$+\sigma^2\mu_{n-1}^{(0,3)}\mu_{n-1}^{(1,0)}+3\gamma_3\mu_{n-1}^{(0,2)}\mu_{n-1}^{(1,0)}+\gamma_4\mu_{n-1}^{(1,0)},$$

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{Z_1}Z_{n-1}^j\Big)^4\Big]=\mu_{n-1}^{(0,4)}+3\sigma^2\big(\mu_{n-1}^{(0,2)}\big)^2+6\gamma_3\mu_{n-1}^{(0,2)}+4\sigma^2\mu_{n-1}^{(0,3)}+\gamma_4.$$

Plugging these into (2.24), we find (it's easiest to compute the coefficients for each of σ^2, γ_3, and γ_4; the last two are zero) that

$$\mathbb{E}\big[(M_n-\mu_nZ_n)^4\big]-\mathbb{E}\big[(M_{n-1}-\mu_{n-1}Z_{n-1})^4\big]=3\sigma^2A_{n-1}(j,j)^2+\mathrm{O}\big(n^{-\delta'}\big).$$

Therefore,

$$\mathbb{E}\big[(M_n-\mu_nZ_n)^4\big]=3n\sigma^2A_{n-1}(j,j)^2+\mathrm{O}\big(1+n^{1-\delta'}\big),$$

and the result easily follows. □

Proof of Theorem 1.2. Assume $d\geqslant 7$ and fix $n\geqslant 3$ and

$$1\leqslant m:=\big\lfloor n-n^{\frac{20}{3(2d+1)}}\big\rfloor<n.$$

Recall the definition and constructions for the BRW and modified BRW, and denote $\mu_{n,m}(j):=\mathbb{E}[M_{n,m}^i(j)]=\mathbb{E}[M_{n-m}(j)]$, $\widetilde{M}_{n,m}^i(j):=M_{n,m}^i(j)-\mu_{n,m}(j)Z_{n,m}^i$. Now, let $\widetilde{\Sigma}=(\sigma^2/2)\Sigma$ and use the triangle inequality to find

$$\begin{aligned}
&d_{\mathrm{W}}\bigg(\mathscr{L}\bigg(\bigg(\frac{M_n(j)-\kappa_jZ_n}{\sqrt{n}}\bigg)_{j=1}^r\,\Big|Z_n>0\bigg),\mathrm{SL}_r(\widetilde{\Sigma})\bigg)\\
&\quad\leqslant d_{\mathrm{W}}\left(\mathscr{L}\bigg(\bigg(\frac{\hat{M}_n(j)-\kappa_j\hat{Z}_n}{\sqrt{n}}\bigg)_{j=1}^r\bigg),\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}M_{n,m}^i(j)-\kappa_j\hat{Z}_n}{\sqrt{n}}\bigg)_{j=1}^r\bigg)\right) &(2.25)\\
&\qquad+d_{\mathrm{W}}\left(\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}M_{n,m}^i(j)-\kappa_j\hat{Z}_n}{\sqrt{n}}\bigg)_{j=1}^r\bigg),\right.\\
&\qquad\qquad\left.\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}M_{n,m}^i(j)-\mu_{n,m}(j)\hat{Z}_n}{\sqrt{n}}\bigg)_{j=1}^r\bigg)\right) &(2.26)\\
&\qquad+d_{\mathrm{W}}\left(\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}M_{n,m}^i(j)-\mu_{n,m}(j)\hat{Z}_n}{\sqrt{n}}\bigg)_{j=1}^r\bigg),\right.\\
&\qquad\qquad\left.\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}\big(M_{n,m}^i(j)-\mu_{n,m}(j)Z_{n,m}^i\big)}{\sqrt{n}}\bigg)_{j=1}^r\bigg)\right) &(2.27)\\
&\qquad+d_{\mathrm{W}}\left(\mathscr{L}\bigg(\bigg(\frac{\sum_{i=1}^{\hat{Z}_m}\widetilde{M}_{n,m}^i(j)}{\sqrt{n}}\bigg)_{j=1}^r\bigg),\mathrm{SL}_r\big((\sigma^2/2)\Sigma_n\big)\right) &(2.28)\\
&\qquad+\frac{\sigma}{\sqrt{2}}\,d_{\mathrm{W}}\big(\mathrm{SL}_r(\Sigma_n),\mathrm{SL}_r(\Sigma)\big), &(2.29)
\end{aligned}$$

where $\Sigma_n=(A_{n-m}(j,k))_{j,k}$ with $A_{n-m}(j,k)=\mathrm{Cov}(\widetilde{M}_{n,m}^1(j),\widetilde{M}_{n,m}^1(k))$. Using the coupling definition of Wasserstein distance and Lemma 2.5, we can bound (2.25) by noting

$$n^{-1/2}\,\mathbb{E}\left|\hat{M}_n(j)-\sum_{i=1}^{\hat{Z}_m}M_{n,m}^i(j)\right|\leqslant cn^{-\frac{2d-9}{6(2d+1)}}.$$

Similarly, (2.26) is bounded from

$$n^{-1/2}\big|\mu_{n,m}(j)-\kappa_j\big|\,\mathbb{E}[\hat{Z}_n]\leqslant cn^{-\frac{14d-3}{6(2d+1)}}\leqslant cn^{-\frac{2d-9}{6(2d+1)}},$$

where in the first inequality we have used Lemma 2.4 and that $\mathbb{E}[\hat{Z}_n] = \mathrm{O}(n)$. For (2.27), note that

$$\begin{aligned} n^{-1/2}\,\mathbb{E}\Big|\sum_{i=1}^{\hat{Z}_m} M^i_{n,m}(j) - \mu_{n,m}(j)\hat{Z}_n - \sum_{i=1}^{\hat{Z}_m}\big(M^i_{n,m}(j) - \mu_{n,m}(j)Z^i_{n,m}\big)\Big| \\ \leqslant n^{-1/2}\mu_{n,m}(j)\,\mathbb{E}\Big|\hat{Z}_n - \sum_{i=1}^{\hat{Z}_m} Z^i_{n,m}\Big| \leqslant cn^{-1/2}\,\mathbb{E}\Big[\sum_{i=m+1}^{n} R'_{n,i} + 1\Big] \\ \leqslant c(n-m) \leqslant cn^{-\frac{6d-37}{6(2d+1)}} \leqslant cn^{-\frac{2d-9}{6(2d+1)}}, \end{aligned}$$

where we used Lemma 2.4 in the second inequality, and part (iii) of Lemma 2.1 in the second to last. Noting our Wasserstein distance is with respect to L_1 distance, summing over j and k, and using the inequalities above shows that (2.25), (2.26), and (2.27) are upper bounded by $cn^{-\frac{2d-9}{6(2d+1)}}$.

To bound (2.28), we apply Theorem 2.8 with $M = \hat{Z}_m$, $\boldsymbol{X}_i = (\widetilde{M}^i_{n,m}(j))_{j=1}^r$, and $\mu = \mathbb{E}[\hat{Z}_m]$, which, using (2.4), is of strict order n. Lemma 2.9 states that

$$\mathbb{E}\Big[\big(\widetilde{M}^1_{n,m}(j)\big)^4\Big] \leqslant c(n-m),$$

and then repeated use of Hölder's and Jensen's inequalities implies that

$$\mathbb{E}\Big[\big\|\big(\widetilde{M}^1_{n,m}(j)\big)_{j=1}^r\big\|_1^3\Big]^{1/3} \leqslant c(n-m)^{1/4} \leqslant cn^{-\frac{5}{3(2d+1)}}.$$

Multiplying this by the $n^{-1/6}$ factor coming from the powers of μ in the bound from Theorem 2.8 gives a term of order $n^{-\frac{2d-9}{6(2d+1)}}$. For the remaining (nontrivial) term, the triangle inequality implies

$$\begin{aligned} &d_{\mathrm{W}}\big(\mathscr{L}(\hat{Z}_m), \mathrm{Geo}(\mu^{-1})\big) \\ &\quad\leqslant d_{\mathrm{W}}\big(\mathscr{L}(\hat{Z}_m), \mathscr{L}(Z_m|Z_m>0)\big) + d_{\mathrm{W}}\big(\mathscr{L}(Z_m|Z_m>0), \mathrm{Exp}\big(\sigma^2/(2m)\big)\big) \\ &\qquad + d_{\mathrm{W}}\big(\mathrm{Exp}\big(\sigma^2/(2m)\big), \mathrm{Exp}(\mu^{-1})\big) + d_{\mathrm{W}}\big(\mathrm{Exp}(\mu^{-1}), \mathrm{Geo}(\mu^{-1})\big), \\ &\quad\leqslant c\big((n-m) + \log(m) + \big|m(\sigma^2/2) - \mu\big|\big), \end{aligned}$$

where the second inequality uses Lemma 2.3, Theorem 1.3, and standard couplings. Using Lemmas 2.2 and 2.3, we have

$$\begin{aligned} \big|m(\sigma^2/2) - \mu\big| &\leqslant \big|m(\sigma^2/2) - \mathbb{E}[Z_m|Z_m>0]\big| + \big|\mathbb{E}[Z_m|Z_m>0] - \mathbb{E}[Z_m|Z_n>0]\big| \\ &\leqslant c\big(\log(m)^2 + (n-m)\big), \end{aligned}$$

Noting that $\mu^{-1/2}(n-m) \leqslant cn^{-\frac{1}{2}+\frac{20}{3(2d+1)}} \leqslant cn^{-\frac{2d-9}{6(2d+1)}}$, and putting these bounds into Theorem 2.8 implies that (2.28) is bounded by $cn^{-\frac{2d-9}{6(2d+1)}}$.

Finally, we bound (2.29). Using the representation $\mathscr{L}(\sqrt{E}\boldsymbol{Z}) = \mathrm{SL}_r(\mathrm{Cov}(\boldsymbol{Z}))$ for E distributed as an exponential with rate one, independent of $\boldsymbol{Z}$, an r-dimensional multivariate normal, we apply Lemma 3.3 below and Lemma 2.9, and noting that $d \geqslant 7$, we conclude that

$$d_{\mathrm{W}}\big(\mathrm{SL}_r(\Sigma_n), \mathrm{SL}_r(\Sigma)\big) \leqslant c\Big(\sum_{j,k}|A_{n-m}(j,k) - \Sigma_{jk}|\Big)^{1/2} \leqslant cn^{-\frac{10(d-6)}{9(2d+1)}} \leqslant cn^{-\frac{2d-9}{6(2d+1)}},$$

and combining the bounds above yields the theorem. □

3 CLT with error

In this section we prove a multivariate CLT with Wasserstein error for sums of i.i.d. variables that is adapted to our setting. The proof is relatively standard using Stein's method, with the complications that we are working in the Wasserstein (rather than smoother test function) metric, and that we do not demand the covariance matrix be non-singular.

In what follows, denote by $|\cdot|$ the Euclidean L_2-norm. For a k-times differential function $f : \mathbb{R}^r \to \mathbb{R}$, let

$$M_k(f) := \sup_{x\in\mathbb{R}^r} \sup_{\substack{a_1,\dots,a_k\in\mathbb{R}^r:\\ |a_1|=\dots=|a_k|=1}} \left| \sum_{i_1,\dots,i_k=1}^{r} a_{1,i_1}\cdots a_{k,i_k} \frac{\partial^k f(x)}{\partial x_{i_1}\cdots\partial x_{i_r}} \right|.$$

Clearly, for any vectors $a_1,\dots,a_k,x\in\mathbb{R}^r$, we have

$$\left| \sum_{i_1,\dots,i_r=1}^{r} a_{1,i_1}\cdots a_{k,i_k} \frac{\partial^k f(x)}{\partial x_{i_1}\cdots\partial x_{i_r}} \right| \leqslant |a_1|\cdots|a_k| M_k(f). \tag{3.1}$$

Theorem 3.1. *Let $X_1,\dots,X_n$ be i.i.d. random vectors in $\mathbb{R}^r$, with $\mathbb{E}[X_1]=0$ and $\operatorname{Var} X_1=\Sigma=(\Sigma_{uv})_{1\leqslant u,v\leqslant r}$. Let $W=n^{-1/2}\sum_{i=1}^n X_i$, and let Z have a standard multivariate normal distribution, and let $Z_\Sigma=\Sigma^{1/2}Z$. Then, for any differentiable function $h:\mathbb{R}^r\to\mathbb{R}$,*

$$|\mathbb{E}[h(W)]-\mathbb{E}[h(Z_\Sigma)]| \leqslant 2M_1(h)\left(\frac{2r\,\mathbb{E}\big[|X_1|^3\big]}{\sqrt{n}}\right)^{1/3}.$$

Proof. We first replace h by h_ε, which is defined as

$$h_\varepsilon(x)=\mathbb{E}[h(x+\varepsilon Z')], \qquad x\in\mathbb{R}^r,$$

where Z' has a standard multivariate normal distribution, independent of all else. The error introduced by replacing h by h_ε is at most

$$|\mathbb{E}[h(W)]-\mathbb{E}[h(W+\varepsilon Z')]| \leqslant \varepsilon M_1(h)\,\mathbb{E}\big[|Z'|\big] \leqslant \varepsilon M_1(h)\sqrt{r},$$

and the same bound holds when W is replaced by Z_Σ. Following [6, Lemma 1], since h_ε is infinitely differentiable, there is a function f on $\mathbb{R}^n$ such that

$$\sum_{u,v=1}^{r}\Sigma_{uv}f_{uv}(x)-\sum_{u=1}^{r}x_u f_u(x)=h_\varepsilon(x)-\mathbb{E}[h_\varepsilon(Z_\Sigma)], \qquad x\in\mathbb{R}^r,$$

and from [6, Lemma 2] and then [10, Lemma 4.6], this function satisfies

$$M_3(f)\leqslant M_3(h_\varepsilon)/3\leqslant \varepsilon^{-2}M_1(h)/3.$$

Now, letting $Y_i=n^{-1/2}X_i$, and using Taylor's expansion about $W^i=W-Y_i$, we have that

$$\begin{aligned}
&\mathbb{E}\sum_{u=1}^{r}W_u f_u(W)\\
&\quad=\mathbb{E}\sum_{i=1}^{n}\sum_{u=1}^{r}Y_{i,u}f_u(W)\\
&\quad=\mathbb{E}\sum_{i=1}^{n}\sum_{u=1}^{r}Y_{i,u}\left[f_u(W^i)+\sum_{v=1}^{r}Y_{i,v}f_{uv}(W^i)+\sum_{v,w=1}^{r}Y_{i,v}Y_{i,w}\int_0^1(1-s)f_{uvw}(W^i+sY_i)ds\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\sum_{i=1}^{n}\sum_{u,v=1}^{r} n^{-1}\Sigma_{uv} f_{uv}(W^i) + \mathbb{E}\sum_{i=1}^{n}\sum_{u,v,w=1}^{r} Y_{i,u}Y_{i,v}Y_{i,w}\int_0^1 (1-s) f_{uvw}(W^i + sY_i)ds \\
&= \mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv} f_{uv}(W^1) + n\,\mathbb{E}\sum_{u,v,w=1}^{r} Y_{1,u}Y_{1,v}Y_{1,w}\int_0^1 (1-s) f_{uvw}(W^1 + sY_1)ds \\
&= \mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv} f_{uv}(W^1) + n\int_0^1 (1-s)\biggl(\mathbb{E}\sum_{u,v,w=1}^{r} Y_{1,u}Y_{1,v}Y_{1,w} f_{uvw}(W^1 + sY_1)\biggr)ds.
\end{aligned}
$$

Applying (3.1) to the quantity inside the second expectation, we have

$$
\biggl|\sum_{u,v,w=1}^{r} Y_{1,u}Y_{1,v}Y_{1,w} f_{uvw}(W^1 + sY_1)\biggr| \leqslant |Y_1|^3 M_3(f).
$$

Now, let Y_1' be an independent copy of Y_1, and note that we have $\Sigma_{uv} = n\,\mathbb{E}(Y_{1,u}'Y_{1,v}')$. Hence,

$$
\begin{aligned}
\mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv} f_{uv}(W) &= \mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv}\biggl[f_{uv}(W^1) + \sum_{w=1}^{r} Y_{1,w}\int_0^1 f_{uvw}(W^1 + sY_1)ds\biggr] \\
&= \mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv} f_{uv}(W^1) + \mathbb{E}\sum_{u,v,w=1}^{r}\Sigma_{uv}Y_{1,w}\int_0^1 f_{uvw}(W^1 + sY_1)ds \\
&= \mathbb{E}\sum_{u,v=1}^{r}\Sigma_{uv} f_{uv}(W^1) + n\int_0^1 \mathbb{E}\sum_{u,v,w=1}^{r} Y_{1,u}'Y_{1,v}'Y_{1,w} f_{uvw}(W^1 + sY_1)ds.
\end{aligned}
$$

Applying again (3.1) to the quantity inside the second expectation, we have

$$
\biggl|\sum_{u,v,w=1}^{r} Y_{1,u}'Y_{1,v}'Y_{1,w} f_{uvw}(W^1 + sY_1)\biggr| \leqslant |Y_1'|^2|Y_1| M_3(f).
$$

Subtracting one from the other, it follows that

$$
|\mathbb{E}[h_\varepsilon(W)] - \mathbb{E}[h_\varepsilon(Z_\Sigma)]| \leqslant \frac{\mathbb{E}\bigl[|X_1|^3\bigr]}{2\sqrt{n}} M_3(f) + \frac{\mathbb{E}\bigl[|X_1|^2\bigr]\,\mathbb{E}\bigl[|X_1|\bigr]}{\sqrt{n}} M_3(f) \leqslant \frac{M_1(h)\,\mathbb{E}\bigl[|X_1|^3\bigr]}{2\varepsilon^2\sqrt{n}}.
$$

Thus,

$$
|\mathbb{E}[h(W)] - \mathbb{E}[h(Z_\Sigma)]| \leqslant 2\varepsilon M_1(h)\sqrt{r} + \frac{M_1(h)\,\mathbb{E}\bigl[|X_1|^3\bigr]}{2\varepsilon^2\sqrt{n}},
$$

and choosing

$$
\varepsilon = \biggl(\frac{\mathbb{E}\bigl[|X_1|^3\bigr]}{4\sqrt{rn}}\biggr)^{1/3}
$$

yields the final bound. □

We also have the following easy corollary to fit our setup above.

Corollary 3.2. *Under the notation and assumptions of Theorem 3.1, if h is differentiable and 1-Lipschitz with respect to the 1-norm, then there is a constant C_r depending only on r, such that*

$$
|\mathbb{E}[h(W)] - \mathbb{E}[h(Z_\Sigma)]| \leqslant C_r\biggl(\frac{\mathbb{E}\bigl[\|X_1\|_1^3\bigr]}{\sqrt{n}}\biggr)^{1/3}.
$$

Proof. By equivalence of norms in $\mathbb{R}^r$, there is a constant q_r depending only on r such that for any $a \in \mathbb{R}^r$, $q_r^{-1}|a| \leqslant \|a\|_1 \leqslant q_r|a|$. Therefore, for any $a, x \in \mathbb{R}^r$,

$$\left|\sum_{i=1}^{r} \frac{a_i}{|a|} \frac{\partial h(x)}{\partial x_i}\right| \leqslant \frac{\|a\|_1}{|a|} \leqslant q_r,$$

and $\mathbb{E}\big[|X|^3\big] \leqslant q_r^3\, \mathbb{E}\big[\|X\|_1^3\big]$. The result now follows from Theorem 3.1. □

Finally, we state a simple lemma used to compare centered multivariate normal distributions.

Lemma 3.3. *Let Σ and Σ' be two non-negative semi-definite $(r \times r)$ matrices for $r \geqslant 1$. Let $X = (X_1, \ldots, X_r)$, respectively $\boldsymbol{Y} = (Y_1, \ldots, Y_r)$, be a centered multivariate random normal vector with covariance matrix Σ, respectively Σ'. Then*

$$d_{\mathrm{W}}\big(\mathscr{L}(\boldsymbol{X}), \mathscr{L}(\boldsymbol{Y})\big) \leqslant C\bigg(\sum_{u,v=1}^{r} \big|\Sigma_{uv} - \Sigma'_{uv}\big|\bigg)^{1/2}$$

for some constant C that only depends on r.

Proof. Using Stein's identity for the multivariate normal, for any twice-differentiable function f, we have

$$\begin{aligned}\mathbb{E}\bigg\{\sum_{u,v=1}^{r} \Sigma_{uv} f_{uv}(Y) - \sum_{u=1}^{r} X_u f_u(Y)\bigg\} &= \mathbb{E}\bigg\{\sum_{u,v=1}^{r} \big(\Sigma_{uv} - \Sigma'_{uv}\big) f_{uv}(Y)\bigg\} \\ &\leqslant M_2(f) \sum_{u,v=1}^{r} \big|\Sigma_{uv} - \Sigma'_{uv}\big|.\end{aligned}$$

The result now follows by using the smoothing argument and Stein's method as in the proof of Theorem 3.1; we omit the details. □

References

[1] Donald A. Dawson. Introductory lectures on stochastic population systems. Technical report. Available at https://arxiv.org/abs/1705.03781, 2017.

[2] Jochen Geiger. A new proof of Yaglom's exponential limit law. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 245–249. Birkhäuser, Basel, 2000. MR-1798303

[3] Samuel Kotz, Tomasz J. Kozubowski, and Krzysztof Podgórski. *The Laplace distribution and generalizations*. Birkhäuser Boston, Inc., Boston, MA, 2001. A revisit with applications to communications, economics, engineering, and finance. MR-1935481

[4] Steven P. Lalley and Xinghua Zheng. Occupation statistics of critical branching random walks in two or higher dimensions. *Ann. Probab.*, 39(1):327–368, 2011. MR-2778804

[5] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995. MR-1349164

[6] Elizabeth Meckes. On Stein's method for multivariate normal approximation. In *High dimensional probability V: the Luminy volume*, volume 5 of *Inst. Math. Stat. (IMS) Collect.*, pages 153–178. Inst. Math. Statist., Beachwood, OH, 2009. MR-2797946

[7] Erol A. Peköz and Adrian Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011. MR-2789507

[8] John Pike and Haining Ren. Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 11(1):571–587, 2014. MR-3283586

[9] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory. MR-1619171

[10] Martin Raič. A multivariate central limit theorem for Lipschitz and smooth test functions. Preprint https://arxiv.org/abs/1812.08268, 2018.

[11] Vladimir A. Vatutin and Andrei M. Zubkov. Branching processes. I. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 23*, Itogi Nauki i Tekhniki, pages 3–67, 154. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1985. MR-0810404

[12] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new. MR-2459454

[13] Akiva M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 56:795–798, 1947. MR-0022045

Acknowledgments. We thank the referees for their helpful comments. AR was supported by Singapore Ministry of Education Academic Research Fund Tier 2 grant MOE2018-T2-2-076, and AR and NR were supported by Australian Research Council research grant DP150101459.