Theoretical Population Biology 124 (2018) 81–92

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

The Wright–Fisher site frequency spectrum as a perturbation of the coalescent's

Andrew Melfi, Divakar Viswanath *

Department of Mathematics, University of Michigan, United States

ARTICLE INFO

Article history:
Received 7 June 2018
Available online 9 October 2018

Keywords:
Sample frequency spectrum
Wright–Fisher
Coalescent
Multiple mergers

ABSTRACT

The first terms of the Wright–Fisher (WF) site frequency spectrum that follow the coalescent approximation are determined precisely, with a view to understanding the accuracy of the coalescent approximation for large samples. The perturbing terms show that the probability of a single mutant in the sample (singleton probability) is elevated in WF but the rest of the frequency spectrum is lowered. A part of the perturbation can be attributed to a mismatch in rates of merger between WF and the coalescent. The rest of it can be attributed to the difference in the way WF and the coalescent partition children between parents. In particular, the number of children of a parent is approximately Poisson under WF and approximately geometric under the coalescent. Whereas the mismatch in rates raises the probability of singletons under WF, its offspring distribution being approximately Poisson lowers it. The two effects are of opposite sense everywhere except at the tail of the frequency spectrum. The WF frequency spectrum begins to depart from that of the coalescent only for sample sizes that are comparable to the population size. These conclusions are confirmed by a separate analysis that assumes the sample size n to be equal to the population size N. Partly thanks to the canceling effects, the total variation distance of WF minus coalescent is $0.12/\log N$ for a population sized sample with $n = N$, which is only 1% for $N = 2 \times 10^4$. The coalescent remains a good approximation for the site frequency spectrum of-large samples.

1. Introduction

An attractive aspect of genealogical analysis is that it begins with current samples whose sequence data are directly measured. Wright–Fisher (WF) and the coalescent are two theoretical models used to make deductions about the genealogies of the current samples (Durrett, 2008).

The coalescent was derived and justified by Kingman (1982a) as an approximation of the WF model. Kingman's analysis and extensions by other authors (Möhle, 2000; Möhle and Sagitov, 2001) assume the current sample size n to be fixed as the haploid population size N becomes large.

In view of the rapid increase in sample sizes in human genetics (see Karczewski et al., 2016, for example), it is worth asking how close the WF and coalescent models are for large samples. A key property of the coalescent is that the genealogy is constructed entirely using binary mergers. In earlier work (Melfi and Viswanath, 2018), we have shown that for sample sizes $n = o\left(N^{1/3}\right)$, WF genealogies involve only binary mergers with probability tending to 1. A more precise result derived here states that if the sample size is given by $n = \alpha N^{1/3}$, the probability that the WF genealogy involves only binary mergers is $\exp\left(-\frac{\alpha^3}{12}\right)$ in the large N limit. To understand the onset of the deviation of WF from the coalescent, Fu (2006) as well as Bhaskar et al. (2014) looked at triple mergers, where three individuals merge into a common parent over a single WF generation. Among other results, we show that if the sample size is $n = \alpha N^{1/2}$, the expected number of triple mergers in the WF genealogy is $\alpha^2/6 + \left(\exp(-\alpha^2/2) - 1\right)/3$. This last result is in agreement with the $N^{1/2}$ scaling deduced in Melfi and Viswanath (2018).

With regard to sequence data, such results are perhaps too exacting. Detailed agreement in the genealogy is essential to reproduce the Kingman partition distribution (Kingman, 1982b) at each step of the genealogy. However, summary statistics such as the site frequency spectrum are not so refined. The site frequency spectrum of a sample of size n, which may be directly obtained from sequence data, consists of the probability that j of the samples are mutants and $n-j$ are ancestral, $j = 1, \ldots, n-1$, at a base pair. The site is assumed to be polymorphic with a single mutation at some individual that is an ancestor of some but not all samples.

The site frequency spectrum has been widely used for making demographic inferences (see Excoffier et al., 2013; Fu et al., 2013; Griffiths and Tavaré, 1998; Gutenkunst et al., 2009; Kamm et al., 2017; Keinan and Clark, 2012; Lukic and Hey, 2012; Wakeley and

* Corresponding author.
E-mail addresses: melfi@umich.edu (A. Melfi), divakar@umich.edu (D. Viswanath).

https://doi.org/10.1016/j.tpb.2018.09.005
0040-5809/

Hey, 1997, for example) and is therefore a good basis to understand the difference between WF and the coalescent with regard to sequence data. If the genealogy is given by the coalescent and if μ is the mutation rate per site per generation, the probability that j out of n samples are mutants is

$$\frac{1/j}{\mathcal{H}_{n-1}}, \tag{1}$$

where $\mathcal{H}_{n-1} = 1 + \frac{1}{2} + \cdots + \frac{1}{n-1}$ is the harmonic number,[1] assuming μ to be so small that μN is negligible and assuming the sample to be polymorphic at the site. We will derive the first perturbing terms that follow (1) under the assumption of WF genealogy.

The elegant formula (1) for the probability of j mutants has a long and complicated history. Fisher (1922) stated that the correlation between heights of fathers and sons was 0.5 and attempted to obtain a Mendelian explanation of that correlation. He was thus led to a consideration of "gene ratios", which is equivalent to counting the number of mutants. He derived the numerator of (1) some years later (Fisher, 1930). Wright (1931, p. 120) had contacted Fisher earlier, noting (among other discrepancies) that he obtained $2N \log(1.8N)$ for the size of the genealogy, whereas Fisher (1922) had obtained $\sqrt{\pi} N^{3/2}$.[2] There can be little doubt that Fisher was aided by Wright (1931) in coming up with the arguments that led him to the numerator of (1) as well as another result we will review shortly.

The size of the genealogy under WF is equal to the number of ancestors (with the current sample included) with $1, \ldots, n-1$ (but not n) descendants in the sample; in other words, the number of ancestors who would make the sample polymorphic if hit with a mutation. The b-branch length of the genealogy is the number of ancestors with exactly b descendants for $b = 1, \ldots, n-1$. The size of the genealogy and the b-branch length are defined analogously for the coalescent, with the difference that the number of generations a lineage survives is no longer an integer. Ancestors of the current sample will be referred to as ancestral samples. Ancestors of the current sample in the same generation will be referred to as an ancestral sample. An ancestral sample induces a partition of the current sample, and for the coalescent, the partition follows the Kingman partition distribution (Durrett, 2008; Kingman, 1982b; Griffiths and Tavaré, 1998; Melfi and Viswanath, 2018).

Kimura (1955) (also see Kimura, 1964, p. 222) solved the diffusion equation for gene frequencies. From that point, (1) can be derived, although an argument connecting mutant frequencies in the population to that in the sample (such as the argument in Durrett, 2008, p. 51) would be needed. The first such argument was given by Ewens (1972), who also introduced the "frequency spectrum" terminology. A coalescent derivation of (1) was given by Fu (1995), as a consequence of the expectations and variances of b-branch lengths of the coalescent genealogy and was preceded by the treatment of special cases (Fu and Li, 1993; Tajima, 1989). A mathematically complete treatment, allowing for varying population sizes, is due to Griffiths and Tavaré (1998), Polanski et al. (2003), and Polanski and Kimmel (2003). A concise and elegant approach to the main ideas of Polanski et al. was obtained recently by Waltoft and Hobolth (2018).

If the genealogy is given by WF, we show that the probability of j mutants out of n is given by

$$\frac{1/j}{\mathcal{H}_{n-1}} - \frac{1}{6N\mathcal{H}_{n-1}(n-1)} - \frac{(j-1)}{6N\mathcal{H}_{n-1}(n-1)(n-j)} + \frac{1}{6N\mathcal{H}_{n-1}j} - \frac{n}{12N\mathcal{H}_{n-1}^2 j} + \frac{n[j=1]}{12N\mathcal{H}_{n-1}} + \cdots, \tag{2}$$

where $[j = 1]$ is 1 if the assertion $j = 1$ is true and 0 otherwise and where $j = 1, \ldots, n-1$. The result (2) is perturbative in that it gives the N^{-1} terms but not the N^{-2} terms.

The main point in calculating the first terms of the perturbation series, shown in (2), is to understand the onset of deviations. Under WF, children are split between parents according to the multinomial distribution. Under the coalescent, the split is uniform (Melfi and Viswanath, 2018). The uniform split is intuitively unreasonable and appears implausible. For example, if two parents have ten children the splits $9+1$ and $5+5$ are equally likely. The assumption of at most a single binary merger per generation breaks down for sample sizes that are as small as $N^{1/3}$. Yet the first terms of the perturbative series (2) show that the deviation in the site frequency spectrum sets in only for sample sizes n that are order of the population size N.

The first few terms in the perturbative series cannot be a good approximation to the total deviation except for small n (however, see Figs. 2 and 3). It is well-known that the first neglected term in a power series often gives a good idea of the error. In the same way, the $\frac{1}{N}$ terms in (2) give an idea of the various phenomena at work in making the WF frequency spectrum differ from that of the coalescent. As noted by earlier authors (Bhaskar et al., 2014; Fu, 2006; Wakeley and Takahashi, 2003), the WF frequency spectrum elevates the probability of singletons ($j = 1$ mutants) and lowers the probability of j mutants for each $j > 1$. Such a movement in mutant probabilities may be verified explicitly from the last two terms of (2), which are the only terms that increase with n. The last two terms increase approximately linearly with sample size. For population sized samples, (2) yields an estimate of $1/12\mathcal{H}_{N-1}$ (or $1/12 \log N$) for the amount by which singleton probability is raised under WF. Even with later terms in the perturbative series not taken into account, that estimate is off only by a factor of 3/2.

In principle, better approximations can be obtained by calculating more terms of the perturbative series. However, the extension of our method to calculate even the N^{-2} terms, which are presumably of the form n^2/N^2, appears difficult. Therefore, we give a separate analysis of population sized samples with $n = N$, with the work of Wakeley and Takahashi (2003) being our starting point. Fisher (1930) gave an ingenious derivation of b-branch lengths of WF genealogies with $n = N$, although some of his arguments are not entirely clear.[3] Wakeley and Takahashi (2003) gave a different and more transparent argument for the $b = 1$ case, which we extend to $b > 1$.

If p_j and q_j are two probability distributions over $j = 1, \ldots, n-1$, the total variation (TV) distance between them is $\frac{1}{2}\sum_{j=1}^{n-1} |p_j - q_j|$. The total variation distance is the maximum difference in probabilities of any possible event under the two distributions (Brémaud, 1999, p. 126) and is therefore a quite robust way to compare probability distributions. The total variation distance between the frequency spectrums under WF and the coalescent for a population sized sample with $n = N$ is approximately

$$\frac{0.1204}{\mathcal{H}_{N-1}} - \frac{0.1124}{\mathcal{H}_{N-1}^2} + \cdots,$$

[1] In sources such as Fu (1995), the harmonic number $\mathcal{H}_{n-1}$ is denoted by a_n. The notation we use is from Graham et al. (1994).

[2] Wright's result is the same as the modern coalescent estimate of the expected size of the genealogy, with 1.8 being his approximation to e^{γ}, where γ is Euler's constant. Wright's $2N$ is the same as our N and his μ is the same as our 2μ. We have modified his formulas accordingly.

[3] Specifically, in deriving the functional equation $\phi(e^{x-1}) - \phi(x) = 1 - x$, Fisher (1930, p. 209) assumes silently that most mutations do not become fixed in the population after assuming the probability of fixation to be 1/2 one paragraph back. That most mutations do not become fixed was known to Wright (1931), and Kimura later proved the probability of a neutral mutation becoming fixed to be $1/N$.

with a slight change in the approximation for $N > 6.8 \times 10^5$. For $N = 2 \times 10^4$, the baseline assumption in human genetics (Durrett, 2008), the total variation distance is only around 1%.

Fu (2006) has connected the greater speed of mergers under the coalescent to the elevation of singleton probability under WF. As we will explain, the coalescent is indeed faster for $n \ll N^{1/2}$ but for $n \approx N$, the picture is not so clear. We refer to the same phenomenon as a mismatch in rates of merger to cover both cases. Another difference between the models is in the way children are partitioned between the parents (Melfi and Viswanath, 2018). In particular, the offspring distribution is approximately Poisson for WF but approximately geometric for the coalescent.

To disentangle the two effects, we define an intermediate model called the discrete coalescent. In the discrete coalescent, the number of parents of a sample of size n has exactly the same distribution as in WF. However, once the number of parents is determined, the children are split between the parents according to Kingman's partition distribution (Kingman, 1982b). The intermediate model shows that the effect of the mismatch in rates is twice as great as the effect of the difference in the way children are split between parents. The two effects are of opposite sense and combine to cause a reduction in overall error.

2. Poisson approximations to Wright–Fisher genealogies

In a backward WF step, each haploid individual chooses one out of N parents with equal probability and independently of all other individuals in its generation. The WF genealogy of a sample is built up using backward WF steps. The coalescent (Kingman, 1982a) may be thought of as a rate varying Poisson approximation of WF genealogies.

Other Poisson approximations may be used to capture more detailed information about WF genealogies. The clumping heuristic is a general method for deriving Poisson approximations (Aldous, 1989). Applications of the heuristic require greater sophistication when the "clumps" are disconnected. In the case of WF, the clumps have a relatively simple form and the heuristic is not difficult to apply.

For the most part, the following basic fact is all that we will need. Suppose the probability of occurrence of an event (such as a thunderstorm) in the interval $(u, u+du)$ is $\lambda(u)\,du$. Then the total number of occurrences of the event in the domain $[a, b]$ has Poisson distribution with rate $\Lambda = \int_a^b \lambda(u)\,du$. In particular, the probability of k occurrences is $\frac{\Lambda^k}{k!}e^{-\Lambda}$. If an event is rare in every neighborhood, the total number of occurrences is approximately Poisson with the rate obtained by summing over the domain.

Let n be the number of samples and N the size of the parental generation. If δ is the number of samples lost due to mergers in a single backward WF step, the number of parental samples is $n - \delta$ and we have

$$\begin{aligned}\mathbb{E}\,\delta &= n - N + N\left(1 - \frac{1}{N}\right)^n\\ \operatorname{Var}\delta &= N\left(\left(1-\frac{1}{N}\right)^n - \left(1-\frac{2}{N}\right)^n\right)\\ &\quad + N^2\left(\left(1-\frac{2}{N}\right)^n - \left(1-\frac{1}{N}\right)^{2n}\right)\end{aligned} \tag{3}$$

(Watterson, 1975). When n is fixed, $\mathbb{E}\,\delta = \frac{n(n-1)}{2N} - \binom{n}{3}\frac{1}{N^2} + \cdots$ and $\operatorname{Var}\delta - \mathbb{E}\,\delta = -2n^3/3N^2 + \cdots$, suggesting a Poisson approximation for small n which turns out to be the Kingman coalescent. More generally, if $n = N^{\mathfrak{a}}$, where $\mathfrak{a} \in [0, 1)$, we have $\mathbb{E}\,\delta - \operatorname{Var}\delta = \mathcal{O}(N^{3\mathfrak{a}-2}) = o(N^{\mathfrak{a}})$, suggesting a Poisson approximation to δ for $\mathfrak{a} < 1$.

For $i = 1, \ldots, k$, the ith sample has the same parent as the $k+1$st sample with probability $1/N$, which is a rare event for $N \gg 1$. The accumulated rate is k/N. Thus, the probability the $k+1$st sample has the same parent as one of the prior samples is approximately $1 - \exp(-k/N)$.

Suppose the number of samples is n. The probability that the $i+1$st sample merges with one of the prior samples is $1 - \exp(-i/N)$ approximately, which is a rare event for $i \ll N$. For $n \ll N$, the cumulative rate $\sum_{i=1}^{n-1}(1 - \exp(-i/N))$ is the left hand Riemann sum of the integral

$$\begin{aligned}\int_1^n \left(1 - e^{-\frac{u}{N}}\right) du &= n - 1 + N\left(e^{-\frac{n}{N}} - e^{-\frac{1}{N}}\right)\\ &= n + N(e^{-\frac{n}{N}} - 1) + \cdots\end{aligned}$$

We may correct for an error that occurs in replacing the sum by an integral by taking $\lambda_\delta(n) = n + N(\exp(-n/N) - 1) - n/2N$. (The sum is now approximated to order N^{-2}.)

For $n \ll N$, δ approximately follows a Poisson distribution of rate $\lambda_\delta(n)$. Thus, $\mathbb{P}(\delta = k) \approx \exp(-\lambda_\delta(n)) \times \lambda_\delta(n)^k/k!$. In fact, $\mathbb{E}\delta = \lambda_\delta(n) + \epsilon$, where $\epsilon = n^2/2N^2 + \cdots$ is of the same order as the error in the Poisson approximation.

2.1. Non-binary mergers

If the sample size is small enough, mergers in any generation are likely to be single binary mergers as in the Kingman coalescent. As the sample size increases, multiple binary mergers may appear with some likelihood and then triple mergers and so on Melfi and Viswanath (2018).

The probability of something other than a binary merger conditional on $\delta \geq 1$ is

$$\frac{1 - e^{-\lambda_\delta(n)} - \lambda_\delta(n)e^{-\lambda_\delta(n)}}{1 - e^{-\lambda_\delta(n)}}.$$

For $n \ll N^{1/2}$, $\lambda_\delta(n) = n^2/2N$ is a good approximation. Let $\Lambda_{2^2}(n)$ be the accumulated rate of occurrence of a double binary merger in some generation of the WF genealogy of a sample of size n. Because non-binary mergers at onset are double binary mergers (Melfi and Viswanath, 2018), we have the cumulative rate of double binary mergers (or non-binary mergers) to be

$$\begin{aligned}\Lambda_{2^2}(n) &= \int_0^n \frac{1 - e^{-x^2/2N} - (x^2/2N)e^{-x^2/2N}}{1 - e^{-x^2/2N}}\,dx\\ &= \int_0^n \frac{e^{x^2/2N} - 1 - x^2/2N}{e^{x^2/2N} - 1}\,dx\\ &= \frac{n^3}{12N} + \cdots,\end{aligned} \tag{4}$$

where the last step is from a power series expansion of $e^{x^2/2N}$. If $n = \alpha N^{1/3}$, $\Lambda_{2^2}(n) = \alpha^3/12$ implying the probability of coalescence with only binary mergers to be $1 - \exp\left(-\alpha^3/12\right)$ (see Fig. 1) and the probability of exactly k binary mergers in the genealogy to be $\exp(-\beta)\beta^k/k!$, with $\beta = \alpha^3/12$.

2.2. Simultaneous binary mergers

The rate $\Lambda_{2^p}(n)$ for p simultaneous binary mergers is obtained similarly. A p-fold simultaneous binary merger occurs during a single backward WF step conditional on $\delta \geq 1$ with probability

$$\begin{aligned}\frac{1 - \sum_{k=0}^{p-1}\exp(-\lambda_\delta(n))\lambda_\delta(n)^k/k!}{1 - e^{-\lambda_\delta(n)}} &= \frac{\lambda_\delta(n)^{p-1}}{p!} + \cdots\\ &= \frac{1}{p!}\left(\frac{n^2}{2N}\right)^{p-1} + \cdots\end{aligned}$$

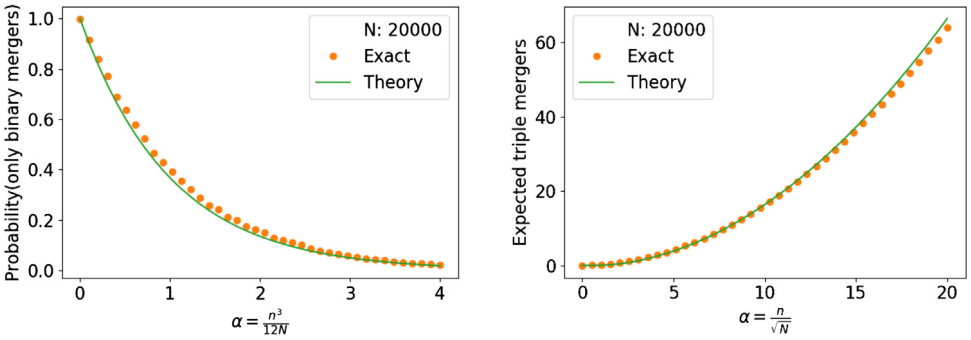


Fig. 1. Plots verifying the approximations implied by (4) and (5). The exact numbers are from computer programs described in Bhaskar et al. (2014) and Melfi and Viswanath (2018).

The accumulated rate over the entire genealogy is

$$\Lambda_{2^p}(n) = \frac{1}{p!(2N)^{p-1}} \int_0^n x^{2p-2}\, dx = \frac{n^{2p-1}}{(2p-1)p!(2N)^{p-1}}.$$

Thus, the correct scaling for the onset of p-fold binary mergers is $n = \alpha N^{\frac{p-1}{2p-1}}$. The scaling was given in earlier work (Melfi and Viswanath, 2018) but not the Poisson approximation.

2.3. Triple mergers

The reasoning for triple mergers is slightly different. We need to first obtain the rate of triple mergers during a single backward WF step. Consider the $m+1$st sample. Each of the first m samples has the same parent as the $m+1$st sample with probability $1/N$, a rare event. Thus, the number of samples out of the first m that have the same parent as the $m+1$st sample is Poisson with rate m/N. The probability that two of them have the same parent as $m+1$, causing a triple merger, is

$$\frac{1}{2!}\left(\frac{m}{N}\right)^2 e^{-m/N}$$

approximately. Therefore, the accumulated rate of triple mergers over a single generation is

$$\lambda_3(n) = \frac{1}{2}\int_0^n \left(\frac{x}{N}\right)^2 e^{-x/N}\, dx = \frac{n^3 \exp(-n/N)}{4N^2} + \cdots$$

Triple mergers are a rare event for $n \ll N^{2/3}$. However, when accumulating the rate of triple mergers over the entire genealogy, it is essential to account for the WF genealogy skipping some sample sizes.

The expected δ when the sample size is n, given that $\delta \geq 1$, is $\lambda_\delta(n)/(1-\exp(-\lambda_\delta(n)))$. Thus, for $m \leq n$, we take the probability that m is reached to be $(1-\exp(-\lambda_\delta(m)))/\lambda_\delta(m)$.

For the accumulated rate of triple mergers, we obtain

$$\Lambda_3(n) = \int_0^n \lambda_3(x)\frac{1-\exp(-\lambda_\delta(x))}{\lambda_\delta(x)}dx$$

We may set $n = \alpha N^{1/2}$ and then use the approximation $\lambda_\delta(n) = n^2/2N$ to obtain

$$\Lambda_3(n) = \frac{\alpha^2}{6} + \frac{e^{-\alpha^2/2}-1}{3}. \tag{5}$$

We may then use the Poisson distribution and approximate the expected number of triple mergers in the genealogy as $\Lambda_3(n)$ (see Fig. 1) or calculate the probability of k triple mergers in the WF genealogy of the sample. For example, if $n = \alpha N^{1/2}$, the expected number of triple mergers in the genealogy is $\alpha^2/6 + \exp(-\alpha^2/2)/3 - 1/3$ in the limit of large N. The $N^{1/2}$ scaling of triple mergers was determined in earlier work (Melfi and Viswanath, 2018).

3. Perturbative analysis of the WF site frequency spectrum

The manner in which coalescent and WF genealogies differ may be inferred from (4), (5), and other similar results. Such differences in genealogy are a part of modeling and are not directly observable from sequence data. The question becomes to what extent the genealogical differences show up in sequence data.

In this section, we will outline the main ideas in obtaining the WF frequency spectrum. The leading term of course is the coalescent answer, which is $1/j\mathcal{H}_{n-1}$. We will calculate the following N^{-1} terms.

The first perturbing terms, which we calculate, suggest that the correct scaling for the divergence of WF frequency spectrum from that of the coalescent is $n = \alpha N$. Although not a proof, the suggestion is almost surely correct and we verify it from another angle later. The scaling for the onset of simultaneous binary mergers and triple mergers is $N^{1/3}$ and $N^{1/2}$ (from (4) and (5)). The fact that the divergence in the frequency spectrum sets in for much larger samples means that the frequency spectrum is not very sensitive to multiple mergers in the genealogy.

If the WF genealogy of a sample of size n progresses through sample sizes as in

$$n \to n-1 \to \cdots \to 2 \to 1$$

without skipping any sample size in-between n and 1, we denote that no-skip event by $\mathcal{S}_0$. If the WF genealogy skips from a sample size of $m+2$ to m, omitting $m+1$, we denote such a skip-to-m event by $\mathcal{S}_m$ for $m = n-2, \ldots, 1$. The sample size of $m+1$ is the only omission in $\mathcal{S}_m$.

Other patterns of skipping are possible. However, the probability of such events is $\mathcal{O}(N^{-2})$. For an $\mathcal{O}(N^{-1})$ calculation, we only need to consider $\mathcal{S}_0$ and $\mathcal{S}_m$.

The WF frequency spectrum is calculated under the assumption of exactly one mutation in the genealogy of the sample. Therefore, we define the event $\mathcal{S}_0^\mu$ to be $\mathcal{S}_0$ and exactly one mutation in the genealogy of the sample. The event $\mathcal{S}_m^\mu$ is defined analogously.

3.1. Coalescent propagators

The general approach to derive the WF frequency spectrum is to first determine the probability that a mutation occurs in the genealogy when the sample size is m for $m = n, \ldots, 2$. That would mean that 1 out of m ancestral samples is a mutant at some point in the genealogy. That probability is then propagated to the current

sample size of n. We begin by studying propagation under the coalescent.

Suppose an ancestral sample of size m has $i \geq 1$ mutants and $(m-i)$ non-mutants. The genealogy from the current sample of size n to the ancestral sample of size m is assumed to involve only binary mergers, with no mutations in-between. As we will presently show, the probability that the current sample has $j \geq i$ mutants is given by

$$\frac{(j-1)^{\underline{(i-1)}}}{(i-1)!} \cdot \frac{(m-1)^{\underline{i}}(n-m)^{\underline{j-i}}}{(n-1)^{\underline{j}}}, \tag{6}$$

where $j^{\underline{i}}$ is the falling power $j(j-1)\dots(j-i+1)$ (Graham et al., 1994). We adopt the convention that $j^{\underline{0}}$ is 1, even for $j = 0$.

The Kingman partition distribution may be used to obtain (6). However, we will give a more direct argument. Suppose an ancestral sample of size $\mathfrak{m}$ has $\mathfrak{i}$ mutants. Suppose that an ancestral sample of size $\mathfrak{m}+1$ is related to it through a single binary merger. Then the probability that the sample of $\mathfrak{m}+1$ has $\mathfrak{i}+1$ mutants is $\mathfrak{i}/\mathfrak{m}$ because each sample out of $\mathfrak{m}$ is equally like to "split" and one of the $\mathfrak{i}$ mutants will split with probability $\mathfrak{i}/\mathfrak{m}$. Similarly, the probability that the sample of $\mathfrak{m}+1$ has $\mathfrak{i}$ mutants is $(\mathfrak{m}-\mathfrak{i})/\mathfrak{m}$ (in this case, one of the $\mathfrak{m}-\mathfrak{i}$ non-mutants has to split).

From here, we can write down the probability that a sample of n has j mutants when it is descended through binary splits from a sample of size m with i mutants to be

$$\binom{n-m}{j-i} \frac{((m-i)\dots(n-j-1))\,(i\dots(j-1))}{m\dots n-1}.$$

The argument for this expression is as follows. There are $n-m$ splits from n to m. The binomial coefficient chooses $j-i$ of those splits to be ones that increase the number of mutants. The denominator of the fraction in the expression steps from the sample size of m to the sample size of $n-1$ because those are the sample sizes that split. The numerator has the factor $(m-i)\dots(n-j-1)$ to account for splits of non-mutants. The other factor $i\dots(j-1)$ accounts for splits of mutants. The above expression is simplified to obtain (6).

When $i = 1$, (6) reduces to

$$\frac{(n-m)^{\underline{j-1}}}{(n-1)^{\underline{j}}}(m-1), \tag{7}$$

a useful special case. Setting $j = 1$, we find the probability of a single mutant in the sample of n given a single mutant in the ancestral sample to be

$$\frac{m-1}{n-1}, \tag{8}$$

which is another useful special case.

3.2. WF propagators

Suppose next that a sample of size n is descended from a sample of size m with i mutants through a single backward WF step. It is assumed that there are no mutations during this descent. The probability of j mutants in the current sample is then given by

$$\binom{n}{j}\left(\left\{{j \atop i}\right\} i!\right)\left(\left\{{n-j \atop m-i}\right\}(m-i)!\right) \Big/ \left\{{n \atop m}\right\} m!. \tag{9}$$

That is because in the current sample, we can choose j individuals to be mutants in $\binom{n}{j}$ ways. That being done, the j mutants in the current sample can be assigned to i mutants in the parental sample, with each parent receiving at least one child, in $\left\{{j \atop i}\right\} i!$ ways: the j samples can be partitioned into i in $\left\{{j \atop i}\right\}$ ways and then can be permuted in $i!$ ways. The last bracketed factor in the numerator is the number of ways to assign $(n-j)$ not mutants to $(m-i)$ non-mutants in the parental sample. The denominator is the number of ways to assign n children to m parents, with each parent receiving at least one child.

The Stirling numbers (of the second kind) $\left\{{\mathfrak{n} \atop 1}\right\}$, $\left\{{\mathfrak{n} \atop \mathfrak{n}-1}\right\}$, and $\left\{{\mathfrak{n} \atop \mathfrak{n}-2}\right\}$ are given by 1, $\mathfrak{n}(\mathfrak{n}-1)/2$, and $\mathfrak{n}(\mathfrak{n}-1)(\mathfrak{n}-2)(3\mathfrak{n}-5)/24$, respectively (Graham et al., 1994). Using (9) along with those formulas, we obtain the probabilities that a sample of size $m+2$ has i, $i+1$, $i+2$ mutants when it is descended from an ancestral sample of size m, with i of them being mutants, in a single WF generation to be

$$\frac{(m-i)(m-i+1)}{m(m+1)} - \frac{2i(m-i)}{m(m+1)(3m+1)}, \tag{10a}$$

$$\frac{2i(m-i)}{m(m+1)} + \frac{4i(m-i)}{m(m+1)(3m+1)}, \tag{10b}$$

$$\frac{i(i+1)}{m(m+1)} - \frac{2i(m-i)}{m(m+1)(3m+1)}, \tag{10c}$$

respectively.

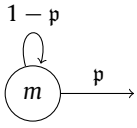
Suppose that a sample of size $m+2$ changes into a parental sample of size m under a single backward WF step. Given that the parental sample of m has only a single mutant, the probability that the sample of size $m+2$ has only a single mutant is

$$\frac{m-1}{m+1} - \frac{2(m-1)}{m(m+1)(3m+1)}, \tag{11}$$

which is obtained by setting $i = 1$ in (10a). Comparing against (8), we find that skipping a step under WF reduces the factor that propagates the probability of a single mutant.

3.3. Probability of mutation at m

What is the probability of a mutation event at m assuming that the sample size m is visited? Consider the following picture:



The picture is showing that an ancestral sample size of m remains m under a backward WF step with probability $1-\mathfrak{p}$ and exits to a lower sample size with probability $\mathfrak{p}$. Neglecting μ^2 terms, the probability that a sample of size m will be hit with a mutation is

$$\sum_{k=0}^{\infty} k(m\mu)\mathfrak{p}(1-\mathfrak{p})^k,$$

where k is the number of returns from m to m.

Thus, with μ^2 terms neglected, the probability of being hit with a mutation at m is equal to $m\mu/\mathfrak{p}$. We may take

$$\begin{aligned}\mathfrak{p} &= 1 - \prod_{k=1}^{m-1}\left(1-\frac{k}{N}\right)\\ &= \frac{m(m-1)}{2N} - \frac{m(m-1)(m-2)(3m-1)}{24N^2} + \cdots\end{aligned}$$

by ignoring terms after N^{-2}. We get the probability of being hit with a mutation at m to be

$$(2N\mu)\left(\frac{1}{m-1} + \frac{(m-2)(3m-1)}{12N(m-1)} + \mathcal{O}(N^{-2})\right) + \mathcal{O}(\mu^2).$$

Neglecting μ^2 and N^{-2} terms, we denote the probability of being hit with a mutation at m by

$$(2N\mu)\left(\frac{1}{m-1} + \frac{\mu_m}{12N}\right), \tag{12}$$

where $\mu_m = (m-2)(3m-1)/(m-1) + \mathcal{O}(N^{-1})$.

In fact, because we are neglecting μ^2 terms, (12) gives the probability that there is a single mutation in the entire genealogy with that mutation occurring when the ancestral sample size is m.

3.4. Probability that $m+2$ skips to m

Suppose the ancestral sample size is $m+2$. What is the probability that the ancestral sample size skips over $m+1$ and goes directly to m under WF? The ancestral sample size could skip over both $m+1$ and m, but because we are neglecting N^{-2} terms, those possibilities may be ignored.

The probability that a backward WF applied to a sample of size $m+2$ results in a sample of size m, conditioned on a merger, is

$$\frac{\left(\frac{1}{N^2}\binom{m+2}{3} + \frac{3}{N^2}\binom{m+2}{4}\right)(1-1/N)\ldots(1-(m-1)/N)}{1-(1-1/N)\ldots(1-(m+1)/N)}.$$

There are $\binom{m+2}{3}$ possible triple mergers and $3\binom{m+2}{4}$ possible double binary mergers. The first factor in the numerator accounts for the probabilities of those. In both a triple merger and a double binary merger, a total of m parents must be chosen distinctly, which occurs with probability $(1-1/N)\ldots(1-(m-1)/N)$. That accounts for the second factor in the numerator. The denominator is the probability that the number of parents of $m+2$ samples is fewer than $m+2$ in a single backward WF step.

Simplifying the above expression, we obtain the probability of skipping to m as

$$\frac{m(3m+1)}{12N}, \tag{13}$$

with N^{-2} terms neglected. If $s_m = m(3m+1)$, this probability can be taken to be $s_m/12N$.

3.5. The event $\mathcal{S}_0^\mu$ and $\mathbb{P}\left(j\middle|\mathcal{S}_0^\mu\right)$

From (13), it follows that the probability of $\mathcal{S}_0$, which visits each ancestral sample size in $\{1,\ldots,n\}$ is $\Pi_{m=1}^{n-2}(1-s_m/12N)$. Using (12), the probability of a single mutation in the genealogy is $(2N\mu)\left(\sum_{m=2}^{n}(1/(m-1)+\mu_m/12N)\right)$, with μ^2 terms ignored and with N^{-2} terms ignored in the coefficient of $2N\mu$. The summation over $m=2,\ldots,n$ sums over the probability of the single mutation occurring when the ancestral sample size is one of $2,\ldots,n$.

Thus, the probability of $\mathcal{S}_0^\mu$ is $\Pi_{m=1}^{n-2}(1-s_m/12N)\times(2N\mu)\left(\sum_{m=2}^{n}(1/(m-1)+\mu_m/12N)\right)$. Simplifying and omitting N^{-2} terms in the coefficient of $2N\mu$, we get $\mathbb{P}(\mathcal{S}_0^\mu) = (2N\mu)W_0 + \mathcal{O}(\mu^2)$ with

$$\begin{aligned} W_0 &= \mathcal{H}_{n-1} - \frac{\mathcal{H}_{n-1}}{12N}\sum_{m=1}^{n-2} m(3m+1) + \frac{1}{12N}\sum_{m=2}^{n}\mu_m \\ &= \mathcal{H}_{n-1} - \frac{\mathcal{H}_{n-1}n(n^2-4n+5)}{12N} + \frac{(n-1)(3n-2)}{24N} \end{aligned} \tag{14}$$

and with N^{-2} terms neglected in W_0. The last step in (14) is gotten after a routine simplification. At this point, we can think of $\mathbb{P}(\mathcal{S}_0^\mu)$ as proportional to the weight W_0.

Let $\mathcal{M}_m$ be the event that a mutation occurs in the genealogy of the sample of size n when the ancestral sample size is m. From (12), we know that the probability that a mutation occurs at m but nowhere else in the genealogy is proportional to $1/(m-1)+\mu_m/12N$. Therefore,

$$\mathbb{P}\left(\mathcal{M}_m\middle|\mathcal{S}_0^\mu\right) = \frac{\frac{1}{m-1}+\frac{\mu_m}{12N}}{\mathcal{H}_{n-1}+\frac{1}{12N}\sum_{m=2}^{n}\mu_m},$$

where the denominator is obtained by summing over $m = 2,\ldots,n$. The right hand side above can be simplified to obtain

$$\mathbb{P}\left(\mathcal{M}_m\middle|\mathcal{S}_0^\mu\right) = \frac{1}{(m-1)\mathcal{H}_{n-1}} + \frac{3m-4}{12N\mathcal{H}_{n-1}} - \frac{(n-1)(3n-2)}{24N\mathcal{H}_{n-1}^2(m-1)} + \cdots$$

with N^{-2} terms ignored and in the limit $\mu\to 0$.

The number of mutants in the current sample of size n is always denoted by j. The next step is to calculate $\mathbb{P}(j\big|\mathcal{M}_m,\mathcal{S}_0^\mu)$. For $j=1$, we can use (8) to propagate a single mutant from an ancestral sample of size m to the current sample of size n and get

$$\mathbb{P}\left(j=1\middle|\mathcal{M}_m,\mathcal{S}_0^\mu\right) = \frac{m-1}{n-1}.$$

More generally, the probability $\mathbb{P}(j\big|\mathcal{M}_m,\mathcal{S}_0^\mu)$, where j stands for j mutants in the current sample of n, is given by (7). By writing $(3m-4)(m-1)$ as $3(m-1)^2-(m-1)$, we obtain

$$\begin{aligned} &\mathbb{P}(j\big|\mathcal{M}_m,\mathcal{S}_0^\mu)\mathbb{P}(\mathcal{M}_m\big|\mathcal{S}_0^\mu) \\ &= \frac{(n-m)^{\underline{j-1}}}{\mathcal{H}_{n-1}(n-1)^{\underline{j}}} + \frac{(m-1)^2(n-m)^{\underline{j-1}}}{4N\mathcal{H}_{n-1}(n-1)^{\underline{j}}} - \frac{(m-1)(n-m)^{\underline{j-1}}}{12N\mathcal{H}_{n-1}(n-1)^{\underline{j}}} \\ &\quad - \frac{(n-1)(3n-2)(n-m)^{\underline{j-1}}}{24N\mathcal{H}_{n-1}^2(n-1)^{\underline{j}}}, \end{aligned}$$

with N^{-2} terms ignored and in the limit $\mu\to 0$. We then have

$$\begin{aligned} \mathbb{P}(j\big|\mathcal{S}_0^\mu) &= \sum_{m=2}^{n}\mathbb{P}(j\big|\mathcal{M}_m,\mathcal{S}_0^\mu)\mathbb{P}(\mathcal{M}_m\big|\mathcal{S}_0^\mu) \\ &= \frac{1}{\mathcal{H}_{n-1}j} + \frac{n(2n-j)}{j(j+1)(j+2)} - \frac{n}{12N\mathcal{H}_{n-1}j(j+1)} \\ &\quad - \frac{(n-1)(3n-2)}{24N\mathcal{H}_{n-1}^2 j} \end{aligned}$$

after simplification, with N^{-2} terms ignored and in the limit $\mu\to 0$. The simplification is effected using the following identities:

$$\sum_{m=2}^{n}(n-m)^{\underline{j-1}} = (n-1)^{\underline{j}}/j,$$

$$\sum_{m=2}^{n}(m-1)(n-m)^{\underline{j-1}} = n(n-1)^{\underline{j}}/j(j+1),$$

$$\sum_{m=2}^{n}(m-1)^2(n-m)^{\underline{j-1}} = n(2n-j)(n-1)^{\underline{j}}/j(j+1)(j+2),$$

for $j=1,2,\ldots,$ each of which is easily proved by induction on n. Another method of proof is to begin with the difference identity $(n+1)^{\underline{j}} - n^{\underline{j}} = jn^{\underline{j}}$.

The event $\mathcal{S}_m^\mu$ and $\mathbb{P}\left(j\middle|\mathcal{S}_m^\mu\right)$

From (13), $\mathbb{P}(S_m)$, which is the probability the genealogy skips from sample size $m+2$ to m, is

$$\frac{m(3m+1)}{12N}\times\prod_{\ell\in\{1\ldots n-2\}-\{m,m+1\}}\left(1-\frac{\ell(3\ell+1)}{12N}\right)$$

or simply $m(3m+1)/12N$ with N^{-2} terms neglected.

Because $\mathbb{P}(S_m)$ leads with an N^{-1} term, we may simplify (12) and take the probability that a mutation hits when the ancestral sample size is ℓ to be $(2N\mu)/(\ell-1)$. It follows that $\mathbb{P}\left(\mathcal{S}_m^\mu\right) = (2N\mu)W_m + \mathcal{O}(\mu^2)$

$$W_m = \frac{m(3m+1)}{12N}\left(\mathcal{H}_{n-1} - \frac{1}{m}\right). \tag{15}$$

with N^{-2} terms neglected in W_m. At this point, we can take $\mathbb{P}(\mathcal{S}_m^\mu)$ to be proportional to W_m.

To calculate $\mathbb{P}(j|\mathcal{S}_m^\mu)$, we use a shortcut that greatly simplifies the algebra. Under the condition $\mathcal{S}_m^\mu$ and by (12) (with the $\mu_m/12N$ term ignored because W_m leads with a N^{-1} term), the probability of a mutation at ℓ is proportional to $1/(\ell-1)$ for $\ell \in \{2,\dots,n\}-\{m+1\}$. Therefore the probability of a mutation at ℓ under the condition $\mathcal{S}_m^\mu$ is equal to

$$\frac{1/(\ell-1)}{\mathcal{H}_{n-1}-1/m},$$

in the limit $\mu \to 0$ and with N^{-1} terms ignored. Now for the shortcut, suppose we can ignore the WF corrections to the propagators, namely, the latter terms in the WF propagators (10a), (10b), and (10c). We can then obtain the probability of j mutants in the current sample of n to be

$$\frac{1}{\mathcal{H}_{n-1}-1/m}\sum_{\ell\in\{2,\dots,n\}-\{m+1\}}\frac{1}{(\ell-1)}\times\frac{(\ell-1)(n-\ell)^{\underline{j-1}}}{(n-1)^{\underline{j}}},$$

where the single mutant at ℓ is propagated to n using the coalescent propagator (7) before summing over ℓ. This expression can be simplified to get

$$\frac{1}{\mathcal{H}_{n-1}-1/m}\left(\frac{1}{j}-\frac{(n-m-1)^{\underline{j-1}}}{(n-1)^{\underline{j}}}\right), \tag{16}$$

which is the probability of j mutants except for the corrections given by the latter terms in the WF propagators (10a)–(10c).

We will now calculate the corrections separately. Let $\mathcal{M}_{2\dots m}$ denote $\mathcal{M}_2\cup\cdots\cup\mathcal{M}_m$, in words, the event where a mutation occurs when the ancestral sample size is $2,\dots,m$. The probability that a mutation strikes when the sample size is ℓ is proportional to $1/(\ell-1)$. Therefore,

$$\mathbb{P}\left(\mathcal{M}_{2\dots m}\middle|\mathcal{S}_m^\mu\right)=\frac{\mathcal{H}_{m-1}}{\mathcal{H}_{n-1}-1/m},$$

with all N^{-1} and μ terms ignored. The latter terms in the WF propagators (10a), (10b), and (10c) will be activated only when the condition $\mathcal{M}_{2\dots m}$ holds in addition to $\mathcal{S}_m^\mu$.

Conditioning on $\mathcal{M}_{2\dots m}$ and $\mathcal{S}_m^\mu$, the frequency spectrum of ancestral sample of size m is given by

$$1/i\mathcal{H}_{m-1}$$

for the probability of i mutants, $i=1,\dots,m-1$ (in the limit $\mu\to 0$ and with N^{-1} terms neglected). To obtain the correction, this frequency spectrum must first be propagated to $m+2$ samples using the latter terms of the WF propagators (10a), (10b), and (10c) because the condition $\mathcal{S}_m^\mu$ stipulates a skip from sample size $m+2$ to sample size m. Propagating the probabilities to $m+2$, we get the corrections to the probability of i mutants in a sample of $m+2$ under the conditions $\mathcal{M}_{2\dots m}$ and $\mathcal{S}_m^\mu$ to be

$$\begin{aligned}
&\frac{-2(m-1)}{\mathcal{H}_{m-1}m(m+1)(3m+1)} && \text{for } i=1,\\
&\frac{2m}{\mathcal{H}_{m-1}m(m+1)(3m+1)} && \text{for } i=2,\\
&\frac{-2}{\mathcal{H}_{m-1}m(m+1)(3m+1)} && \text{for } i=m+1,
\end{aligned}$$

and zero for all other $i\in\{1,\dots,m+1\}-\{1,2,m+1\}$. Multiplying these numbers with the coalescent propagator (6) with $m\leftarrow m+2$ and $i\leftarrow 1,2,m+1$, respectively, we get the corrections to the probability of j mutants in the current sample of n under the conditions $\mathcal{M}_{2\dots m}$ and $\mathcal{S}_m^\mu$ to be

$$\begin{aligned}
&\frac{-2(m-1)(n-m-2)^{\underline{j-1}}}{\mathcal{H}_{m-1}m(3m+1)(n-1)^{\underline{j}}},\\
&\frac{2(j-1)m(n-m-2)^{\underline{j-2}}}{\mathcal{H}_{m-1}(3m+1)(n-1)^{\underline{j}}},\\
&\frac{-2(j-1)^{\underline{m}}(n-m-2)^{\underline{j-m-1}}}{\mathcal{H}_{m-1}m(3m+1)(n-1)^{\underline{j}}}.
\end{aligned}$$

Multiplying these terms by $\mathbb{P}(\mathcal{M}_{2\dots m}|\mathcal{S}_m^\mu)$ and adding to (16), we get

$$\begin{aligned}
\mathbb{P}(j|\mathcal{S}_m^\mu)=&\frac{1}{\mathcal{H}_{n-1}-1/m}\left(\frac{1}{j}-\frac{(n-m-1)^{\underline{j-1}}}{(n-1)^{\underline{j}}}\right)\\
&-\frac{2(m-1)(n-m-2)^{\underline{j-1}}}{(\mathcal{H}_{n-1}-1/m)m(3m+1)(n-1)^{\underline{j}}}\\
&+\frac{2(j-1)m(n-m-2)^{\underline{j-2}}}{(\mathcal{H}_{n-1}-1/m)(3m+1)(n-1)^{\underline{j}}}\\
&-\frac{2(j-1)^{\underline{m}}(n-m-2)^{\underline{j-m-1}}}{\mathcal{H}_{m-1}m(3m+1)(n-1)^{\underline{j}}},
\end{aligned}$$

in the limit $\mu\to 0$ and with N^{-1} terms ignored.

3.6. WF sample frequency spectrum

The sum $\sum_{m=1}^{n-2}W_m\mathbb{P}(j|\mathcal{S}_m^\mu)$ may be simplified to get

$$\begin{aligned}
&\frac{(n-2)(n-1)^2}{12Nj}+\frac{(3n-2)[j=1]}{12N}-\frac{n}{12Nj(j+1)}\\
&-\frac{n(2n-j)}{4Nj(j+1)(j+2)}\\
&-\frac{(n-j-2)(n-j-1)}{6Nj(j+1)(n-1)}+\frac{(2n-j-1)[j\ge 2]}{6Nj(j+1)}\\
&-\frac{(j-1)}{6N(n-1)(n-j)},
\end{aligned} \tag{17}$$

where the second line accounts for WF corrections to the coalescent propagators. The simplification uses the identities

$$\begin{aligned}
\sum_{m=1}^{n-2}(n-m-2)^{\underline{j-1}}&=(n-2)^{\underline{j}}\Big/j && \text{for } j=1,2,\dots\\
\sum_{m=1}^{n-2}(m-1)(n-m-2)^{\underline{j-1}}&=(n-2)^{\underline{j+1}}\Big/j(j+1) && \text{for } j=1,2,\dots\\
\sum_{m=1}^{n-2}m^2(n-m-2)^{\underline{j-2}}&=(2n-j-1)(n-1)^{\underline{j}}\\
&\qquad\Big/(j-1)j(j+1) && \text{for } j=2,3,\dots\\
\sum_{m=1}^{n-2}(j-1)^{\underline{m}}(n-m-2)^{\underline{j-m-1}}&=(j-1)(n-2)^{\underline{j-2}} && \text{for } j=1,2,\dots
\end{aligned}$$

In the last identity, $a^{\underline{b}}$ is assumed to be 1 if $b\le 0$. All these identities may be verified by induction on n.

The WF frequency spectrum (2) is obtained by simplifying

$$\frac{W_0\mathbb{P}(j|\mathcal{S}_0^\mu)+\sum_{m=1}^{n-2}W_m\mathbb{P}(j|\mathcal{S}_m^\mu)}{W_0+\sum_{m=1}^{n-2}W_m}. \tag{18}$$

If we look at the sequence of steps building up to this point, the difference in the way WF and the coalescent partition children between parents first comes up in the latter term of (11) as well as (10a), (10b), (10c). Those terms propagate to the second line of (17).

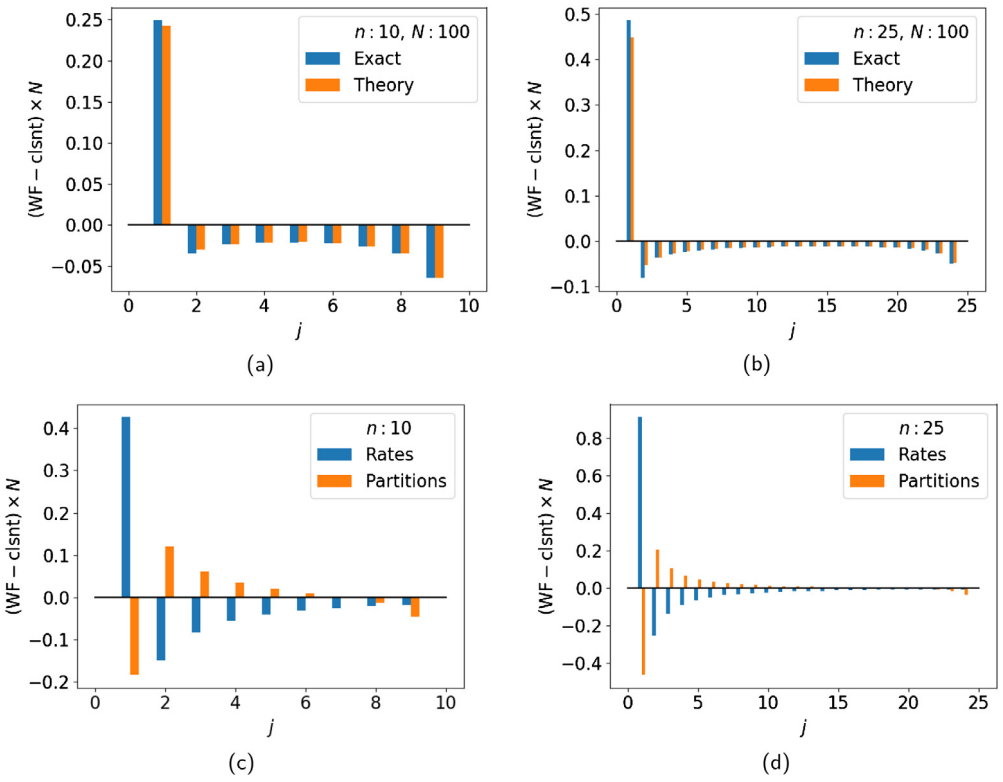


Fig. 2. (a) and (b) WF minus coalescent computed using (2) minus (1) (theory) is compared with a computation using the program of Bhaskar et al. (2014) (exact). (c) and (d) (2) minus (1) minus (19) (rates) is compared with (19) (partitions).

Thus the N^{-1} terms in the WF frequency spectrum (2) due to differences in partitioning between WF and the coalescent are given by

$$-\frac{(n-j-2)(n-j-1)}{6N\mathcal{H}_{n-1}(n-1)j(j+1)}+\frac{(2n-j-1)[j\geq 2]}{6N\mathcal{H}_{n-1}j(j+1)}-\frac{j-1}{6N\mathcal{H}_{n-1}(n-1)(n-j)}. \tag{19}$$

Evaluating with $j=1$ and retaining only the dominant term, we get $-n/12N\mathcal{H}_{n-1}$ to be the effect on singleton probability of the difference in partitioning distributions. Evaluating (2) with $j=1$ and retaining only the dominant term, we obtain $n/12N\mathcal{H}_{n-1}$ as the amount by which the WF singleton probability exceeds that of the coalescent. Therefore, the effect of the mismatch in rates of merger must be $n/6N\mathcal{H}_{n-1}$.

Fig. 2 shows that the WF singleton probabilities are elevated and the rest of the frequency spectrum is depressed, as may be inferred from the last two terms of (2). The figure also illustrates the correction due to rates being twice as high as the correction due to differences in the way children are partitioned between parents.

Because $j=1$ singleton probabilities are elevated under WF and other probabilities are lowered, we may obtain the total variation distance between WF and the coalescent by simply taking the difference in $j=1$ probabilities. Thus, the perturbative estimate for the total variation distance between the WF frequency spectrum and that of the coalescent is $n/12N\mathcal{H}_{n-1}$. This estimate is qualitatively correct even for $n=N$, and even quantitatively it is not unreasonable, being about 2/3rds of a better estimate we will presently derive. Fig. 3 shows that the total variation distance increases with n and decreases with N.

If the number of samples is $n=3$, the exact WF frequency spectrum is given by

$$\frac{2N-1}{3N-2},\quad \frac{N-1}{3N-2}.$$

If $n=4$, the exact WF frequency spectrum is given by

$$\frac{2\left(9N^3-20N^2+16N-4\right)}{33N^3-82N^2+73N-22},\quad \frac{3\left(N^2-2N+1\right)}{11N^2-20N+11},\quad \frac{2\left(N-1\right)\left(3N^2-6N+4\right)}{(3N-2)\left(11N^2-20N+11\right)}.$$

The perturbative WF frequency spectrum (2) may be checked against these exact answers.

4. Population sized samples

Suppose the sample size is $n=\alpha N$. For an individual among the parental population of N, the probability that any given sample is a child is $1/N$, a rare event. The accumulated probability over the sample of size αN is α. Therefore, by the Poisson clumping heuristic, we may approximate the number of children of an individual in the parental generation by the Poisson distribution with rate α. The probability that an individual has k children among the αN samples is approximately $\exp(-\alpha)\alpha^k/k!$. The generating function $\sum_{k=0}^{\infty}\mathfrak{p}_k x^k$ with $\mathfrak{p}_k=\exp(-\alpha)\alpha^k/k!$ is $\exp(\alpha(x-1))$.

The only individuals in the parental generation that appear in the genealogy are ones who have at least one child among the samples. Therefore, it is natural to look at the Poisson distribution under the condition of having one child. Under that condition, the probability of having k children is $\mathfrak{p}_k/(1-\exp(-\alpha))$ and the generating function is $(\exp(\alpha x)-1)/(\exp(\alpha)-1)$.

Let $G_1(\alpha N)=g_1(\alpha)N$ be the expected b-branch length with $b=1$ of the WF genealogy of a sample of size αN. By (3), the expected number of parents is $N(1-\exp(-\alpha))$. Thus, we may write

$$g_1(\alpha)N=N\alpha+\mathfrak{f}Ng_1(1-\exp(-\alpha))$$

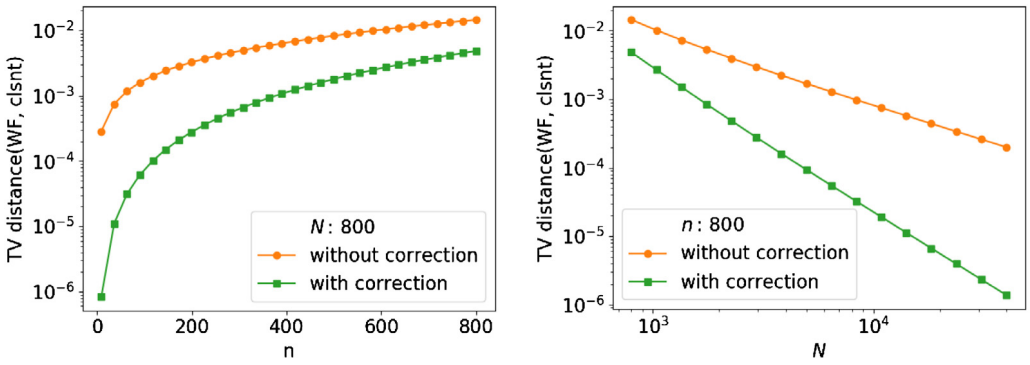


Fig. 3. Total variation distance between WF frequency spectrum (Bhaskar et al., 2014) and that of the coalescent given by (1) (without correction) or with correction as given by (2).

because the current samples $N\alpha$ all contribute to the 1-branch length and with the understanding that $\mathfrak{f}$ is the probability that a branch with a single descendant in the genealogy of the parental sample of size $(1 - \exp(-\alpha))N$ remains a branch with a single descendant in the genealogy of the current sample of size αN. That probability $\mathfrak{f}$ is the same as the probability of a parent having a single child, which is $\alpha/(\exp(\alpha) - 1)$. Therefore, we have

$$g_1(\alpha) = \alpha + \frac{\alpha}{\exp(\alpha) - 1} g_1\left(1 - \exp(-\alpha)\right), \tag{20}$$

which is a result of Wakeley and Takahashi (2003) derived essentially using their arguments.

The generating function for the number of children of a parents is approximately

$$\left(\frac{\exp(\alpha x) - 1}{\exp(\alpha) - 1}\right)^a. \tag{21}$$

The coefficient of x^b in $(\exp(\alpha x) - 1)^a/(\exp(\alpha) - 1)^a$ is given by

$$\frac{(-1)^a \alpha^b}{(\exp(\alpha) - 1)^a b!} \sum_{j=0}^{a} \binom{a}{j} (-1)^j j^b.$$

Using (Graham et al., 1994, (6.19), p. 265) to evaluate the sum, the probability that a parents have b children is found to be

$$\frac{\alpha^b a!}{(\exp(\alpha) - 1)^a b!} \left\{ {b \atop a} \right\} \tag{22}$$

for $b = a, a + 1, \ldots$ Here $\left\{ {b \atop a} \right\}$ is a Stirling number of the second kind (Graham et al., 1994). Using the same argument as above and taking the b-branch length with αN samples to be $G_b(\alpha N) = N g_b(\alpha)$, we get the recurrence

$$g_b(\alpha) = \sum_{a=1}^{b} g_a(1 - \exp(-\alpha)) \times \frac{\alpha^b a!}{(\exp(\alpha) - 1)^a b!} \left\{ {b \atop a} \right\} \tag{23}$$

for $b = 2, 3, \ldots$

By solving the recurrences for $g_b(\alpha)$ and taking $\alpha = 1$, we can obtain approximations to the WF frequency spectrum with $n = N$ and compare it to (1), which is the coalescent frequency spectrum. However, we seek to separate the difference into a part due to the mismatch in rates of mergers and a part due to the difference in the way children are partitioned among parents.

To do so, we turn to the discrete coalescent, which is a model intermediate between the coalescent and WF. To obtain the manner in which αN children are split between βN parents under the discrete coalescent, which uses the Kingman partition distribution, we may fix an orange at the left most position and permute $\beta N - 1$ identical oranges and $\alpha N - \beta N$ identical apples after it. The number of children of the ith parent can be taken to be the number of apples between the ith and $i + 1$st orange plus one (thus counting the ith orange) (Durrett, 2008; Griffiths and Tavaré, 1998; Melfi and Viswanath, 2018).

Table 1
The expected b-branch length of the WF genealogy of $n = N$ samples is $(2/b + \epsilon_b)N$. For the discrete coalescent, whose merger rates match WF but which partitions children between parents like the coalescent, it is $(2/b + \tilde{\epsilon}_b)N$.

b	ϵ_b	$\tilde{\epsilon}_b$
1	0.240917257	0.418035261
2	−0.046223840	−0.100136471
3	0.005196946	−0.032826669
4	0.001095702	−0.017086273
5	−0.000238278	−0.011181848
6	−0.000114882	−0.008036411
7	−0.000004091	−0.006053860

The probability that a parent has k children is approximately $\gamma(1 - \gamma)^k$, with $\gamma = \beta/\alpha = (1 - \exp(-\alpha))/\alpha$ for a sample of size αN. The generating function of this geometric distribution is $\gamma x/(1 - (1 - \gamma)x)$. The generating function for the number of children of a parents is approximately $(\gamma x/(1 - (1 - \gamma)x))^a$. By extracting the coefficient of x^b, we find the probability of a parents having b children under the discrete coalescent to be

$$\binom{b-1}{a-1} \gamma^a (1 - \gamma)^{b-a}$$

approximately.

If $\tilde{G}_b(\alpha N)$ denotes the b-branch length of the discrete coalescent genealogy of αN samples, we may set $\tilde{G}_b(\alpha N) = N\tilde{g}_b(\alpha)$ and obtain the recurrences

$$\tilde{g}_1(\alpha) = \alpha + \frac{1 - \exp(-\alpha)}{\alpha} \tilde{g}_1\left(1 - \exp(-\alpha)\right)$$

$$\tilde{g}_b(\alpha) = \sum_{a=1}^{b} \tilde{g}_a(1 - \exp(-\alpha)) \times \binom{b-1}{a-1} \gamma^a (1 - \gamma)^{b-a}, \tag{24}$$

where $b = 2, 3, \ldots$

In the Appendix, we show how to solve (20), (23) and (24) accurately using Chebyshev polynomials. For the coalescent, the expected b-branch length for a sample of size $n = N$ is $2N/b$. Therefore, we set $g_b(1) = (2/b + \epsilon_b)N$ and $\tilde{g}_b(1) = (2/b + \tilde{\epsilon}_b)N$ and report ϵ_b and $\tilde{\epsilon}_b$ in Table 1.

The first column of the table agrees very well with Fisher (1922, p. 214). To obtain the total size of the WF genealogy of $n = N$ samples, we use

$$\sum_{b=1}^{N-1} G_b(N) = N\left(2\mathcal{H}_{N-1} + \sum_{b=1}^{N-1} \epsilon_b\right)$$
$$= N\left(2\mathcal{H}_{N-1} + \Delta\right).$$

We estimate Δ to be 0.200645075 by summing ϵ_b over $1 \le b \le 20$. The size of the discrete coalescent genealogy is the same as that of WF genealogy by definition. Our value for Δ agrees with Fisher's except in the last decimal place.

The probability of j mutants in the WF spectrum of $n = N$ samples is estimated to be

$$\frac{G_b(N)}{N(2\mathcal{H}_{N-1}+\Delta)} = \frac{1}{j\mathcal{H}_{N-1}} + \frac{\epsilon_j\mathcal{H}_{N-1} - \Delta/j}{\mathcal{H}_{N-1}(2\mathcal{H}_{N-1}+\Delta)}.$$

The estimated probability of $j = 1$ under WF exceeds $1/j\mathcal{H}_{N-1}$ because $\epsilon_1 > \Delta$. For $j = 3$, the term $\epsilon_j\mathcal{H}_{N-1} - \Delta/j$ is negative as long as $N < 6.8 \times 10^5$ but flips sign around $N = 6.8 \times 10^5$. For $j = 4$, $\epsilon_j\mathcal{H}_{N-1} - \Delta/j$ turns positive only around $N = 10^{20}$. Thus, with minor caveats, the WF frequency spectrum is elevated at $j = 1$ but depressed slightly for $j > 1$.

We can approximate the total variation distance between WF and coalescent frequency spectrums for $n = N$ as

$$\frac{\epsilon_j\mathcal{H}_{N-1} - \Delta/j}{\mathcal{H}_{N-1}(2\mathcal{H}_{N-1}+\Delta)} = \frac{0.1204}{\log N} - \frac{0.1819}{(\log N)^2} + \cdots, \tag{25}$$

a result that is a direct consequence of Fisher (1922). The first plot of Fig. 4 shows this estimate to be quite good. The figure also shows $g_1(\alpha)$ is quite accurate for even $N = 100$ and small α, although $g_4(\alpha)$ has visible errors for $N = 100$.

From Table 1, it is evident that the excess of the discrete coalescent's j mutant probability over that of the coalescent

$$\frac{\tilde{\epsilon}_j\mathcal{H}_{N-1} - \Delta/j}{\mathcal{H}_{N-1}(2\mathcal{H}_{N-1}+\Delta)}$$

is positive for $j = 1$ and negative for $j > 1$. The discrete coalescent and the coalescent differ only with respect to their rates of merger. Both of them follow the Kingman partition distribution. The effect due to difference in the rates of merger alone is twice as great because $\tilde{\epsilon}_1$ is nearly twice ϵ_1.

When $n \ll N^{1/2}$, we can say that the coalescent is faster than WF (Fu, 2006) because $n(n-1)/2N \ge \mathbb{E}\delta$ (see Appendix). However, when $n \gg N^{1/2}$, there can be several mergers in the same generation and rates of merger cannot be compared so directly. Although, the coalescent begins with a higher rate it adjusts its rate downwards with every binary merger.

During a single backward WF step a sample size of $n = \alpha N$ changes on an average to $m = (1 - \exp(-\alpha))N$. On an average the coalescent takes $2N(1/m - 1/n)$ generations to go from n samples to m. In fact, $2((1 - \exp(-\alpha))^{-1} - \alpha^{-1}) = 1 + \alpha/6 + \cdots > 1$ (see Appendix), and the coalescent is actually slower.

However, the 1-branch length of the coalescent in going from n samples to m is equal to $2N(n-m)/(n-1)$. It may be shown that $2N(n-m)/(n-1) < N\alpha$ when $n = N\alpha$ and $m = N(1 - \exp(-\alpha))$ (see Appendix). Therefore, although the coalescent may take a little more than a generation to go from n samples to m, its 1-branch length is lower as a consequence of repeated binary mergers over slightly more than a generation.

5. Discussion

WF deviates from the assumptions of the coalescent for even small sample sizes. Simultaneous binary mergers appear in WF genealogies for sample sizes of only $\alpha N^{1/3}$ with appreciable probability. Triple mergers appear for samples sizes of $\alpha N^{1/2}$.

However, the effect of such deviations on the site frequency spectrum is minimal. Deviations in the site frequency spectrum set in only for sample sizes αN. Even for population sized samples the deviation is only around 1%. The effect is so small because the coalescent is self-correcting. The rate of mergers under the coalescent is faster, but the coalescent lowers the rate with every merger. The coalescent limits itself to binary mergers. As a result, the offspring distribution under the coalescent is geometric, whereas it is Poisson under WF. The geometric and Poisson distributions are not far enough apart to cause a major effect. In addition, the effect of differing offspring distributions partly cancels the effect of differing rates of merger.

Population substructure is perhaps the major reason to look for more sophisticated models than the coalescent (Durrett, 2008; Wakeley, 2009). Skewed offspring distributions are another reason (Eldon and Wakeley, 2006; Matuszewski et al., 2018). In the setting of skewed offspring distributions, it is known that the skew has to be comparable to the population size for deviations to show up (Eldon and Wakeley, 2006). Thus, in that setting too, the coalescent is a robust model.

It is known that increasing skewness of offspring distribution raises the probability of singletons and lowers the probabilities of j mutants for $j > 1$ (Eldon and Wakeley, 2006). The Poisson offspring distribution of WF has a lower variance than the geometric offspring distribution of the coalescent (see Appendix). Our finding that the effect of differing offspring distributions is to lower the singleton probability under WF is consistent with this point of view.

As far as the single site frequency spectrum is concerned, the coalescent appears to be a robust and reliable model relative to WF, and it will perhaps remain so until the SNP determination errors fall below a percent. However, what if multiple sites are considered, possibly allowing for recombination between sites? Our conjecture is that the total variation distance between WF and the coalescent will still be of the order $C/\log N$ for population sized samples. However, the constant C may increase with the number of sites. In that regard, we mention the availability of software to efficiently simulate WF genealogies under very general conditions (Palamara, 2016).

Acknowledgment

The authors thank the editor and one of the reviewers for their comments and suggestions.

Appendix

In this appendix, we explain how to solve (20), (23), and (24) using Chebyshev polynomials. In addition, a few elementary inequalities used in the text are proved.

For convenience, we restate the recurrence for g_1:

$$g_1(\alpha) = \alpha + \frac{\alpha}{\exp(\alpha) - 1} g_1\left(1 - \exp(-\alpha)\right).$$

Begin with $g_1(\alpha) = C_0 + C_1\alpha + C_2\alpha^2 + C_3\alpha^3 + \cdots$ and expand each term of the recurrence to obtain all terms up to the α^3 term. We then obtain

$$\begin{aligned} C_0 + C_1\alpha + C_2\alpha^2 + C_3\alpha^3 = C_0 &+ (1 - C_0/2 + C_1)\,\alpha \\ &+ (C_0/12 - C_1 + C_2)\,\alpha^2 \\ &+ (C_1/2 - 3C_2/2 + C_3)\alpha^3. \end{aligned}$$

If follows that $g_1(\alpha) = 2 + \alpha/6 + \alpha^2/18 + \cdots$ as in Wakeley and Takahashi (2003). Using the same method, we get $g_2(\alpha) = 1 - \alpha^2/36 + \mathcal{O}(\alpha^3)$ and $g_b(\alpha) = 2/b + \mathcal{O}(\alpha^3)$ for $b > 2$.

To solve the recurrence for $g_1(\alpha)$, we set $g_1(\alpha) = 2 + \alpha/6 + \alpha^2/18 + \mathfrak{g}_1(\alpha)$. The resulting recurrence of $\mathfrak{g}_1(\alpha)$ is

$$\begin{aligned} \mathfrak{g}_1(\alpha) = \alpha - 2 - \alpha/6 - \alpha^2/18 \\ + \frac{\alpha}{\exp(\alpha) - 1}(2 + \beta/6 + \beta^2/18 + \mathfrak{g}_1(\beta)), \end{aligned}$$

where $\beta = 1 - \exp(-\alpha)$. It is solved by iteration at each of 32 Chebyshev points in $\alpha \in [0, 1]$. The function $g_1(\alpha)$ may then be

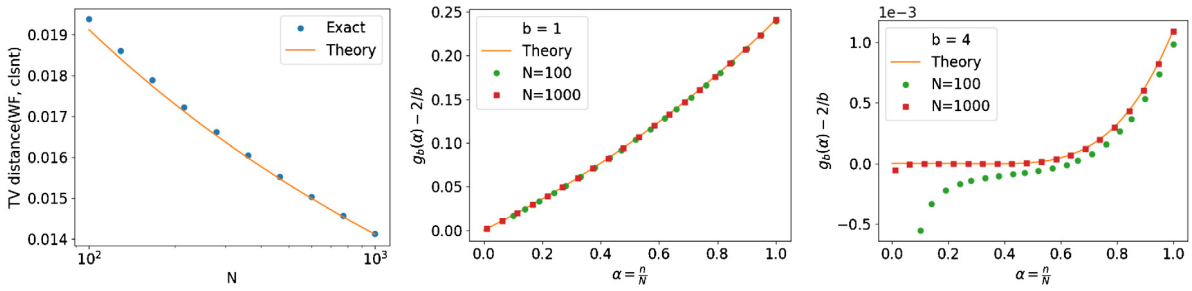


Fig. 4. The first plot demonstrates the accuracy of (25). The next two plots examine the accuracy of $g_b(\alpha)$. In all cases, the exact computations use the computer program of Bhaskar et al. (2014).

obtained with 10+ digits of accuracy at any $\alpha \in [0, 1]$ using the barycentric Lagrange interpolant (Trefethen, 2013). The functions $g_b(\alpha)$, with $b = 2, 3, \ldots, 20$ are calculated using the same method.

For the functions $\tilde{g}_b(\alpha)$, $b = 1, \ldots, 20$, we begin with $\tilde{g}_1(\alpha) = 2 + \alpha/3 + 2\alpha^2/27 + \tilde{\mathfrak{g}}_1(\alpha)$, $\tilde{g}_2(\alpha) = 1 - \alpha/18 - 19\alpha^2/432 + \tilde{\mathfrak{g}}_2(\alpha)$, and $\tilde{g}_b(\alpha) = 2/b - \alpha/3b(b+1) - \alpha^2/18b(b+1)(b+2) + \tilde{\mathfrak{g}}_b(\alpha)$ for $b > 2$. The rest of the method is the same.

The inequality $n(n-1)/2 \geq \mathbb{E}\delta$

To verify that $n(n-1)/2 \geq \mathbb{E}\delta = n - N + N(1-1/N)^n$, set

$$(1-1/N)^n = 1 - n/N + n(n-1)/2N^2 - \mathfrak{r}/6N^3$$

and use the Lagrange form of the Taylor series remainder to deduce $\mathfrak{r} > 0$.

The inequality $2((1-\exp(-\alpha))^{-1} - \alpha^{-1}) > 1$

The inequality $2((1-\exp(-\alpha))^{-1} - \alpha^{-1}) > 1$ is equivalent to

$$e^\alpha > \frac{e^\alpha - 1}{2} + \frac{e^\alpha - 1}{\alpha},$$

which is proved by verifying that the series for the left hand side majorizes the series for the right hand side.

The inequality $2N(n-m)/(n-1) < N\alpha$

To show that $2N(n-m)/(n-1) < N\alpha$ when $n = N\alpha$ and $m = N(1-\exp(-\alpha))$, first observe the inequality follows from $2(1-m/n) < \alpha$ for N large. Now $2(1-m/n) < \alpha$ is equivalent to $e^{-\alpha} < 1 - \alpha + \alpha^2/2$, which can be verified using the Lagrange form of the Taylor series remainder.

The inequality $\sigma_G > \sigma_P$

Suppose the sample size is αN with the parental population size being N as usual. Conditional on an individual of the parental generation being a parent of one of the samples and assuming N large, its number of children (among the samples) is given by the generating function $(\exp(\alpha x) - 1)/(\exp(\alpha) - 1)$. It follows that the expectation of the number of children is α and the variance is

$$\sigma_P = \frac{\alpha}{1-\exp(-\alpha)} + \frac{\alpha^2}{1-\exp(-\alpha)} - \frac{\alpha^2}{(1-\exp(-\alpha))^2}.$$

If the αN children are split among their parents according to the Kingman partition distribution, the generating function for the number of children is $\gamma x/(1-(1-\gamma)x)$ with $\gamma = (1-\exp(-\alpha))/\alpha$. The expectation is again α and the variance is

$$\sigma_G = \frac{\alpha}{1-\exp(-\alpha)} + \frac{\alpha^2}{(1-\exp(-\alpha))^2} - 2.$$

One may verify that $\sigma_G > \sigma_P$ by plotting a graph. Alternatively,

$$\sigma_G - \sigma_P = \frac{\alpha^2}{(\exp(\alpha)-1)^2}\left(2e^\alpha\frac{(e^\alpha+1)}{2} - 2\left(\frac{e^\alpha-1}{\alpha}\right)^2\right)$$

must be positive because the power series of both e^α and $(e^\alpha+1)/2$ majorize the power series of $(e^\alpha - 1)/\alpha$.

Intuitively, we expect $\sigma_G > \sigma_P$ because the geometric distribution has exponential decay, whereas the Poisson distribution has super-exponential decay.

References

Aldous, D., 1989. Probability Approximations via the Poisson Clumping Heuristic. Springer, New York.

Bhaskar, A., Clark, A.G., Song, Y.S., 2014. Distortion of genealogical properties when the sample is very large. Proc. Natl. Acad. Sci. 111, 2385–2390.

Brémaud, P., 1999. Markov Chains. Springer.

Durrett, R., 2008. Probability Models for DNA Sequence Evolution. Springer Science & Business Media.

Eldon, B., Wakeley, J., 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172, 2621–2633.

Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3, 87–112.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M., 2013. Robust demographic inference from genomic and SNP data. PLos Genet. 9 (10).

Fisher, R.A., 1922. On the dominance ratio. Proc. Roy. Soc. Edinburgh 42, 321–341.

Fisher, R.A., 1930. The distribution of gene ratios for rare mutations. Proc. Roy. Soc. Edinburgh 50, 204–219.

Fu, Y.-X., 1995. Statistical properties of segregating sites. Theor. Popul. Biol. 48, 172–197.

Fu, Y., 2006. Exact coalescent for the Wright-Fisher model. Theor. Popul. Biol. 69, 385–394.

Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Altshuler, D., Shendure, J., Nickerson, D.A., Bamshad, M.J., Akey, J.M., 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493 (7431), 216–220.

Graham, R.L., Knuth, D.E., Patashnik, O., 1994. Concrete Mathematics, second ed. Addison-Wesley, NJ.

Griffiths, R.C., Tavaré, S., 1998. The age of a mutation in a general coalescent tree. Stoch. Models 14, 273–295.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLos Genet. 5.

Kamm, J.A., Terhorst, J., Song, Y.S., 2017. Efficient computation of the joint sample frequency spectra for multiple populations. J. Comput. Graph. Statist. 26, 182–194.

Karczewski, K.J., Weisburd, B., Thomas, B., et al., 2016. The ExAC browser: displaying reference data information from over 60000 exomes. Nucleic Acids Res. 45, D840–D845.

Keinan, A., Clark, A.G., 2012. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 336 (6082), 740–743.

Kimura, M., 1955. Solution of a process of random genetic drift with a continous model. Proc. Nat. Acad. Sci. 41, 144–150.

Kimura, M., 1964. Diffusion models in population genetics. J. Appl. Probab. 1, 177–232.

Kingman, J.F.C., 1982a. On the genealogy of large populations. J. Appl. Probab. 19, 27–43.

Kingman, J.F.C., 1982b. The coalescent. Stochastic Process. Appl. 13, 235–248.

Lukic, S., Hey, J., 2012. Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-africa expansion. Genetics 192, 619–639.
Matuszewski, S., Hildebrandt, M.E., Achaz, G., Jensen, J.D., 2018. Coalescent processes with skewed offspring distributions and nonequilibrium demography. Genetics 208, 323–338.
Melfi, A., Viswanath, D., 2018. Single and simultaneous binary mergers in Wright-Fisher genealogies. Theor. Popul. Biol. 121, 60–71.
Möhle, M., 2000. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. Adv. Appl. Probab. 32, 983–993.
Möhle, M., Sagitov, S., 2001. A classification of coalescent processes for haploid exchangeable population models. Ann. Probab. 29, 1547–1562.
Palamara, P.F., 2016. ARGON: Fast, whole-genome simulation of the discrete time Wright-Fisher process. Bioinformatics 32 (19), 3032–3034.
Polanski, A., Bobrowski, A., Kimmel, M., 2003. A note on distributions of times to coalescence, under time-dependent population size. Theor. Popul. Biol. 63 (1), 33–40.
Polanski, A., Kimmel, M., 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. Genetics 165, 427–436.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Trefethen, L.N., 2013. Approximation Theory and Approximation Practice. SIAM.
Wakeley, J., 2009. Coalescent Theory. W.H. Freeman.
Wakeley, J., Hey, J., 1997. Estimating ancestral population parameters. Genetics 145, 847–855.
Wakeley, J., Takahashi, T., 2003. Gene genealogies when the sample size exceeds the effective size of the population. Mol. Biol. Evol. 20, 208–213.
Waltoft, B.L., Hobolth, A., 2018. Non-parametric estimation of population size changes from the site frequency spectrum. Stat. Appl. Genet. Mol. Biol. 17.
Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.