PADÉ APPROXIMANTS AND EXACT TWO-LOCUS SAMPLING DISTRIBUTIONS

By Paul A. Jenkins* and Yun S. Song*,†

University of California, Berkeley

July 26, 2010

For population genetics models with recombination, obtaining an exact, analytic sampling distribution has remained a challenging open problem for several decades. Recently, a new perspective based on asymptotic series has been introduced to make progress on this problem. Specifically, closed-form expressions have been derived for the first few terms in an asymptotic expansion of the two-locus sampling distribution when the recombination rate ρ is moderate to large. In this paper, a new computational technique is developed for finding the asymptotic expansion to an *arbitrary order*. Computation in this new approach can be automated easily. Furthermore, it is proved here that only a finite number of terms in the asymptotic expansion is needed to recover (via the method of Padé approximants) the *exact* two-locus sampling distribution as an analytic function of ρ; this function is exact for all values of $\rho \in [0, \infty)$. It is also shown that the new computational framework presented here is flexible enough to incorporate natural selection.

1. Introduction. Central to many applications in genetic analysis is the notion of sampling distribution, which describes the probability of observing a sample of DNA sequences randomly drawn from a population. In the one-locus case with special models of mutation such as the infinite-alleles model or the finite-alleles parent-independent mutation model, closed-form sampling distributions have been known for many decades (Ewens, 1972; Wright, 1949). In contrast, for multi-locus models with finite recombination rates, finding a closed-form sampling distribution has so far remained a challenging open problem. To make progress on this long-standing issue, we recently proposed a new approach based on asymptotic expansion and showed that it is possible to obtain useful analytic results when the recom-

1

bination rate is moderate to large (Jenkins and Song, 2009, 2010). More precisely, our previous work can be summarized as follows.

Consider a two-locus Wright-Fisher model, with the two loci denoted by A and B. In the standard coalescent or diffusion limit, let θ_A and θ_B denote the population-scaled mutation rates at loci A and B, respectively, and let ρ denote the population-scaled recombination rate between the two loci. Given a sample configuration $\boldsymbol{n}$ (defined later in the text), consider the following asymptotic expansion of the sampling probability $q(\boldsymbol{n} \mid \theta_A, \theta_B, \rho)$ for large ρ:

$$(1.1) \quad q(\boldsymbol{n} \mid \theta_A, \theta_B, \rho) = q_0(\boldsymbol{n} \mid \theta_A, \theta_B) + \frac{q_1(\boldsymbol{n} \mid \theta_A, \theta_B)}{\rho} + \frac{q_2(\boldsymbol{n} \mid \theta_A, \theta_B)}{\rho^2} + \cdots ,$$

where the coefficients $q_0(\boldsymbol{n} \mid \theta_A, \theta_B), q_1(\boldsymbol{n} \mid \theta_A, \theta_B), q_2(\boldsymbol{n} \mid \theta_A, \theta_B), \ldots$, are independent of ρ. The zeroth-order term q_0 corresponds to the contribution from the completely unlinked (i.e., $\rho = \infty$) case, given simply by a product of marginal one-locus sampling distributions. Until recently, higher-order terms $q_M(\boldsymbol{n} \mid \theta_A, \theta_B)$, for $M \geq 1$, were not known. In Jenkins and Song (2010), assuming the infinite-alleles model of mutation at each locus, we used probabilistic and combinatorial techniques to derive a closed-form formula for the first-order term q_1, and showed that the second-order term q_2 can be decomposed into two parts, one for which we obtained a closed-form formula and the other which satisfies a simple recursion that can be easily evaluated using dynamic programming. We later extended these results to an arbitrary finite-alleles model and showed that the same functional form of q_1 is shared by all mutation models, a property which we referred to as *universality* [see Jenkins and Song (2009) for details]. Importantly, we also performed an extensive study of the accuracy of our results and showed that they may be accurate even for moderate values of ρ, including a range that is of biological interest.

Given the above findings, one is naturally led to ask several important follow-up questions. In particular, the following are of both theoretical and practical interest:

1. Is it possible to compute the higher-order coefficients $q_M(\boldsymbol{n} \mid \theta_A, \theta_B)$ for $M > 2$?
2. For a given finite $\rho > 0$, does the series in (1.1) converge as more terms are added?
3. If not, how should one make use of the asymptotic expansion in practice?
4. Is it possible to incorporate into the asymptotic expansion framework other important mechanisms of evolution such as natural selection?

In this paper, we develop a new computational technique to answer the above questions. Our previous method requires rewriting complex recursion relations into more structured forms, followed by laborious computation of the expectation of increasingly complicated functions of multivariate hypergeometric random variables. Generalizing that method to go beyond the second order (i.e., $M > 2$) seems unwieldy. In contrast, our new method is based on the diffusion process and it utilizes the diffusion generator to organize computation in a simple, transparent fashion. Moreover, the same basic procedure, which is purely algebraic, applies to all orders and the computation can be completely automated; we have, in fact, made such an implementation.

To recapitulate, we propose here a method of computing the asymptotic expansion (1.1) to an arbitrary order. That is, for any given positive integer M, our method can be used to compute the coefficients $q_k(\boldsymbol{n} \mid \theta_A, \theta_B)$ for all $k \leq M$; Theorem 3.1 summarizes this result. As discussed in Section 6.2, however, one can find examples for which the series (1.1) diverges for finite, non-zero ρ. To get around this problem, we employ the method of Padé approximants. The key idea behind Padé approximants is to approximate the function of interest by a rational function. Although (1.1) may diverge, we show that the sequence of Padé approximants converges for all values of $\rho > 0$. In fact, for every sample configuration $\boldsymbol{n}$, we show that there exists a finite positive integer $C(\boldsymbol{n})$, such that the Padé approximant of the asymptotic expansion up to order $\geq C(\boldsymbol{n})$ is equal to the *exact* two-locus sampling distribution. Hence, our result implies that only a finite number of terms in the asymptotic expansion need to be computed to recover (via the Padé approximant) the exact sampling distribution as an analytic function of ρ; this function is exact for all values of ρ, including 0. Theorem 4.1 and the surrounding discussion lay out the details. Lastly, we also show in this paper that our new framework is flexible enough to incorporate a general model of diploid selection. This extension is detailed in Section 5.

The above-mentioned convergence result is theoretically appealing. For practical applications, however, one needs to bear in mind that the value of $C(\boldsymbol{n})$ generally grows with sample size, thus implying that obtaining an exact, analytic sampling distribution may be impracticable for large samples. A possible remedy, which works well in practice, is to compute the asymptotic expansion only up to some reasonable order $M < C(\boldsymbol{n})$, and use the corresponding Padé approximant as an approximate sampling distribution. We show in Section 6 that using $M = 10$ or so produces quite accurate results.

An important advantage of our method over Monte Carlo-based meth-

ods is that, for a given mutation model, the bulk of the computation in our approach needs to be carried out only once. Specifically, the coefficients $q_k(\boldsymbol{n} \mid \theta_A, \theta_B)$ need to be computed only once, and the same coefficients can be used to evaluate the sampling distribution at different values of the recombination rate ρ. We expect this aspect of our work to have important practical implications. For example, in the composite likelihood method for estimating fine-scale recombination rates (Hudson, 2001; McVean et al., 2004), one needs to generate exhaustive lookup tables containing two-locus sampling probabilities for a wide range of discrete ρ values. An alternative approach would be to store the coefficients $q_k(\boldsymbol{n} \mid \theta_A, \theta_B)$ instead of generating an exhaustive lookup table using importance sampling, which is computationally expensive.

The rest of this paper is organized as follows. In Section 2, we lay out the notational convention adopted throughout this paper and review our previous work on asymptotic expansion of the two-locus sampling distribution up to second order. Our new technique for obtaining an arbitrary-order asymptotic expansion is described in Section 3, where we focus on the selectively neutral case. In Section 4, we present the method of Padé approximants and describe the aforementioned result on convergence to the exact sampling distribution. In Section 5, we describe how natural selection can be incorporated into our new framework. Finally, we summarize in Section 6 our empirical study of the accuracy of various approximate sampling distributions and provide in Section 7 proofs of the main theoretical results presented in this paper.

2. Notation and review of previous work. In this section, we introduce some notation and briefly review previous results on asymptotic sampling distributions. Initial results were obtained for the infinite-alleles model of mutation (Jenkins and Song, 2010) and later generalized to an arbitrary finite-alleles model (Jenkins and Song, 2009). In this paper we focus on the latter case.

The set of nonnegative integers is denoted by $\mathbb{N}$. Given a positive integer k, $[k]$ denotes the set $\{1, \ldots, k\}$. For a nonnegative real number z and a positive integer n, $(z)_n := z(z+1) \ldots (z+n-1)$ denotes the nth ascending factorial of z. We use $\boldsymbol{0}$ to denote either a vector or a matrix of all zeroes; it will be clear from context which is intended. Throughout, we consider the diffusion limit of a haploid exchangeable model of random mating with constant population size $2N$. We refer to the haploid individuals in the population as gametes. Initially we shall assume that the population is selectively neutral; we extend to the nonneutral case in Section 5.

2.1. *One-locus sampling distribution.* The sample configuration at a locus is denoted by a vector $\boldsymbol{n} = (n_1, \ldots, n_K)$, where n_i denotes the number of gametes with allele i at the locus, and K denotes the number of distinct possible alleles. We use $n = \sum_{i=1}^{K} n_i$ to denote the total sample size. Let u denote the probability of mutation at the locus per gamete per generation. Then, in the diffusion limit, $N \to \infty$ and $u \to 0$ with the population-scaled mutation rate $\theta = 4Nu$ held fixed. Mutation events occur according to a Poisson process with rate $\theta/2$, and allelic changes are described by a Markov chain with transition matrix $\boldsymbol{P} = (P_{ij})$; i.e., when a mutation occurs to an allele i, it mutates to allele j with probability P_{ij}.

We denote by $p(\boldsymbol{n})$ the probability of obtaining the *unordered* sample configuration $\boldsymbol{n}$. When writing sampling probabilities, we suppress the dependence on parameters for ease of notation. By exchangeability, the probability of any *ordered* configuration corresponding to $\boldsymbol{n}$ is invariant under all permutations of the sampling order. We may therefore use $q(\boldsymbol{n})$ without ambiguity to denote the probability of any *particular* ordered configuration consistent with $\boldsymbol{n}$. The two probabilities are related by

$$p(\boldsymbol{n}) = \frac{n!}{n_1! \ldots n_K!} q(\boldsymbol{n}).$$

Throughout this paper, we express our results in terms of ordered samples for convenience.

Consider an infinite population specified by the population-wide allele frequencies $\boldsymbol{x} = (x_i, \ldots, x_K)$, evolving according to a Wright-Fisher diffusion on the simplex

$$(2.1) \qquad \Delta_K = \left\{ \boldsymbol{x} = (x_i) \in [0,1]^K : \sum_{i=1}^{K} x_i = 1 \right\}.$$

We assume that a sample is drawn from the population at stationarity. No closed-form expression for $q(\boldsymbol{n})$ is known except in the special case of parent-independent mutation (PIM), in which $P_{ij} = P_j$ for all i. In the PIM model, the stationary distribution of $\boldsymbol{x}$ is Dirichlet with parameters $(\theta P_1, \ldots, \theta P_K)$ (Wright, 1949), and so $q(\boldsymbol{n})$ is obtained by drawing an ordered sample from this population:

(2.2)
$$q(\boldsymbol{n}) = \mathbb{E}\left[\prod_{i=1}^{K} X_i^{n_i} \right] = \Gamma(n + \theta) \int_{\Delta_K} \prod_{i=1}^{K} \frac{x_i^{n_i + \theta P_i - 1}}{\Gamma(n_i + \theta P_i)} \, d\boldsymbol{x} = \frac{1}{(\theta)_n} \prod_{i=1}^{K} (\theta P_i)_{n_i}.$$

This sampling distribution can also be obtained by coalescent arguments (Griffiths and Tavaré, 1994).

2.2. *Two loci.* We now extend the above notation to two loci, which we refer to as A and B. Denote the probability of a recombination event between the two loci per gamete per generation by r. In the diffusion limit, as $N \to \infty$ we let $r \to 0$ such that the population-scaled recombination parameter $\rho = 4Nr$ is held fixed. Suppose there are K possible alleles at locus A and L possible alleles at locus B, with respective population-scaled mutation parameters θ_A and θ_B, and respective mutation transition matrices $\boldsymbol{P}^A$ and $\boldsymbol{P}^B$. The two-locus sample configuration is denoted by $\boldsymbol{n} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, where $\boldsymbol{a} = (a_1, \ldots, a_K)$ with a_i being the number of gametes with allele i at locus A and unspecified alleles at locus B; $\boldsymbol{b} = (b_1, \ldots, b_L)$ with b_j being the number of gametes with unspecified alleles at locus A and allele j at locus B; and $\boldsymbol{c} = (c_{ij})$ is a $K \times L$ matrix with c_{ij} being the multiplicity of gametes with allele i at locus A and allele j at locus B. We also define

$$a = \sum_{i=1}^{K} a_i, \quad c_{i\cdot} = \sum_{j=1}^{L} c_{ij}, \quad c = \sum_{i=1}^{K} \sum_{j=1}^{L} c_{ij},$$
$$b = \sum_{j=1}^{L} b_j, \quad c_{\cdot j} = \sum_{i=1}^{K} c_{ij}, \quad n = a + b + c,$$

and use $\boldsymbol{c}_A = (c_{i\cdot})_{i \in [K]}$ and $\boldsymbol{c}_B = (c_{\cdot j})_{j \in [L]}$ to denote the marginal sample configurations of $\boldsymbol{c}$ restricted to locus A and locus B, respectively. Notice the distinction between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, which represent gametes with alleles specified at only one of the two loci, and the vectors $\boldsymbol{c}_A$ and $\boldsymbol{c}_B$, which represent the one-locus marginal configurations of gametes with both alleles observed.

When we consider the ancestry of a sample backward in time, a gamete may undergo recombination between the two loci, with each of its two parents transmitting genetic material at only one of the two loci. We allow the nontransmitting locus to remain unspecified as we trace the ancestry further back in time.

Denote by $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ the sampling probability of an ordered sample with configuration $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, again suppressing the dependence on parameters for ease of notation. Sampling is now from a two-dimensional Wright-Fisher diffusion with population allele frequencies $\boldsymbol{x} = (x_{ij})_{(i \in [K], j \in [L])}$, evolving on the state space

$$(2.3) \qquad \Delta_{K \times L} = \left\{ \boldsymbol{x} = (x_{ij}) \in [0,1]^{K \times L} : \sum_{i=1}^{K} \sum_{j=1}^{L} x_{ij} = 1 \right\}.$$

As before, $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is specified by drawing an ordered sample from the

population at stationarity: $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \mathbb{E}[F(\boldsymbol{X}; \boldsymbol{n})]$, where

$$(2.4) \qquad F(\boldsymbol{x}; \boldsymbol{n}) = \left(\prod_{i=1}^{K} x_{i\cdot}^{a_i} \right) \left(\prod_{j=1}^{L} x_{\cdot j}^{b_j} \right) \left(\prod_{i=1}^{K} \prod_{j=1}^{L} x_{ij}^{c_{ij}} \right),$$

with $x_{i\cdot} = \sum_{j=1}^{L} x_{ij}$, and $x_{\cdot j} = \sum_{i=1}^{K} x_{ij}$. In the two-locus model with $0 \leq \rho < \infty$, the stationary distribution, and hence the sampling distribution, is not known in closed-form even when the mutation process is parent-independent. However, when $\rho = \infty$, the two loci become independent and $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is simply the product of the two marginal one-locus sampling distributions. More precisely, denoting the one-locus sampling distributions at A and B by q^A and q^B, respectively, we have

$$\lim_{\rho \to \infty} q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = q^A(\boldsymbol{a} + \boldsymbol{c}_A) q^B(\boldsymbol{b} + \boldsymbol{c}_B),$$

for all mutation models (Ethier, 1979). In particular, if mutation is parent-independent, then we do have a closed-form formula for $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ when $\rho = \infty$, since from (2.2) we know that

$$(2.5) \quad q^A(\boldsymbol{a}) = \frac{1}{(\theta_A)_a} \prod_{i=1}^{K} (\theta_A P_i^A)_{a_i}, \quad \text{and} \quad q^B(\boldsymbol{b}) = \frac{1}{(\theta_B)_b} \prod_{j=1}^{L} (\theta_B P_j^B)_{b_j}.$$

2.3. *Asymptotic sampling formula.* As mentioned in Introduction, although a closed-form formula for $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is not known for an arbitrary ρ, previously we (Jenkins and Song, 2009, 2010) were able to make progress by posing, for large ρ, an asymptotic expansion of the form

$$(2.6) \qquad q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) + \frac{q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho} + \frac{q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho^2} + \cdots,$$

where the coefficients $q_k(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, for all $k \geq 0$, are independent of ρ. We summarize our previous results in the following theorem, specialized to the case of finite-alleles mutation, which is our interest here:

THEOREM 2.1 (Jenkins and Song, 2009). *In the asymptotic expansion (2.6) of the neutral two-locus sampling formula, the zeroth order term is given by*

$$(2.7) \qquad q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = q^A(\boldsymbol{a} + \boldsymbol{c}_A) q^B(\boldsymbol{b} + \boldsymbol{c}_B),$$

and the first order term is given by

$$q_1(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = \binom{c}{2} q^A(\boldsymbol{a}+\boldsymbol{c}_A) q^B(\boldsymbol{b}+\boldsymbol{c}_B)$$

$$- q^B(\boldsymbol{b}+\boldsymbol{c}_B) \sum_{i=1}^{K} \binom{c_{i\cdot}}{2} q^A(\boldsymbol{a}+\boldsymbol{c}_A-\boldsymbol{e}_i)$$

$$- q^A(\boldsymbol{a}+\boldsymbol{c}_A) \sum_{j=1}^{L} \binom{c_{\cdot j}}{2} q^B(\boldsymbol{b}+\boldsymbol{c}_B-\boldsymbol{e}_j)$$

$$(2.8) \qquad + \sum_{i=1}^{K}\sum_{j=1}^{L} \binom{c_{ij}}{2} q^A(\boldsymbol{a}+\boldsymbol{c}_A-\boldsymbol{e}_i) q^B(\boldsymbol{b}+\boldsymbol{c}_B-\boldsymbol{e}_j),$$

where $\boldsymbol{e}_i$ is a unit vector with a 1 at the ith entry and 0s elsewhere. Furthermore, the second order term can be decomposed as

$$(2.9) \qquad q_2(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = \sigma(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) + q_2(\boldsymbol{a}+\boldsymbol{c}_A,\boldsymbol{b}+\boldsymbol{c}_B,\boldsymbol{0}),$$

where $\sigma(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ is known analytically and $q_2(\boldsymbol{a},\boldsymbol{b},\boldsymbol{0})$ satisfies the recursion relation

$$[a(a+\theta_A-1) + b(b+\theta_B-1)] q_2(\boldsymbol{a},\boldsymbol{b},\boldsymbol{0}) =$$

$$\sum_{i=1}^{K} a_i(a_i-1) q_2(\boldsymbol{a}-\boldsymbol{e}_i,\boldsymbol{b},\boldsymbol{0}) + \sum_{j=1}^{L} b_j(b_j-1) q_2(\boldsymbol{a},\boldsymbol{b}-\boldsymbol{e}_j,\boldsymbol{0})$$

$$+ \theta_A \sum_{i=1}^{K} a_i \sum_{t=1}^{K} P_{ti}^A q_2(\boldsymbol{a}-\boldsymbol{e}_i+\boldsymbol{e}_t,\boldsymbol{b},\boldsymbol{0})$$

$$+ \theta_B \sum_{j=1}^{L} b_j \sum_{t=1}^{L} P_{tj}^B q_2(\boldsymbol{a},\boldsymbol{b}-\boldsymbol{e}_j+\boldsymbol{e}_t,\boldsymbol{0})$$

$$+ 4\sum_{i=1}^{K}\sum_{j=1}^{L} a_i b_j [(a-1)(b-1) q^A(\boldsymbol{a}) q^B(\boldsymbol{b})$$

$$- (b-1)(a_i-1) q^A(\boldsymbol{a}-\boldsymbol{e}_i) q^B(\boldsymbol{b})$$

$$- (a-1)(b_j-1) q^A(\boldsymbol{a}) q^B(\boldsymbol{b}-\boldsymbol{e}_j)$$

$$(2.10) \qquad + (a_i-1)(b_j-1) q^A(\boldsymbol{a}-\boldsymbol{e}_i) q^B(\boldsymbol{b}-\boldsymbol{e}_j)],$$

with boundary conditions $q_2(\boldsymbol{e}_i,\boldsymbol{0},\boldsymbol{0}) = 0$, $q_2(\boldsymbol{0},\boldsymbol{e}_j,\boldsymbol{0}) = 0$, and $q_2(\boldsymbol{e}_i,\boldsymbol{e}_j,\boldsymbol{0}) = 0$ for all $i \in [K]$ and $j \in [L]$.

The expression for $\sigma(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ can be found in Jenkins and Song (2009) and we do not reproduce it here. Notice that $q_0(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ and $q_1(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ exhibit

universality: their dependence on the model of mutation is subsumed entirely into the one-locus sampling distributions.

The proof of Theorem 2.1 used coalescent arguments. By considering the most recent event back in time in the coalescent process for the sample, it is possible to write down a recursion relation for the sampling distribution. In a two-locus, finite-alleles model, the appropriate recursion is a simple modification of the one introduced by Golding (1984) for the *infinite*-alleles model, also studied in detail by Ethier and Griffiths (1990). By substituting (2.6) into this recursion, after some lengthy algebraic manipulation one can obtain the expressions given in Theorem 2.1 [see Jenkins and Song (2009, 2010) for details].

3. Arbitrary-order asymptotic expansion. The approach described in Section 2.3 does not generalize easily. In what follows, we introduce a new approach based on the diffusion approximation. This new method is more transparent and more easily generalizable than the one used previously, and we illustrate this point by developing a method for computing the asymptotic expansion (2.6) to an arbitrary order.

3.1. *Diffusion approximation of the two-locus model.* Our approach is based on the diffusion process that is dual to the coalescent process. The generator for the 2-locus finite-alleles diffusion process is

$$(3.1) \quad \mathscr{L} = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{L} \left[\sum_{k=1}^{K} \sum_{l=1}^{L} x_{ij} (\delta_{ik} \delta_{jl} - x_{kl}) \frac{\partial}{\partial x_{kl}} + \theta_A \sum_{k=1}^{K} x_{kj} (P_{ki}^A - \delta_{ik}) \right.$$
$$\left. + \theta_B \sum_{l=1}^{L} x_{il} (P_{lj}^B - \delta_{jl}) + \rho(x_{i.} x_{.j} - x_{ij}) \right] \frac{\partial}{\partial x_{ij}},$$

where δ_{ik} is the Kronecker delta. For notational convenience, henceforth where not specified otherwise the indices i and k are assumed to take value in $[K]$, while the indices j and l are assumed to take value in $[L]$.

In what follows, we change to a new set of variables that capture the decay of dependence between the two loci, an approach originally due to Ohta and Kimura (1969a,b). Specifically, the key quantity of interest is the following:

DEFINITION 3.1. *The linkage disequilibrium (LD) between allele i at locus A and allele j at locus B is given by*

$$d_{ij} = x_{ij} - x_{i.} x_{.j}.$$

The collection of $(K+1)(L+1) - 1$ new variables is

$$(x_{1\cdot}, \ldots, x_{K\cdot}; x_{\cdot 1}, \ldots, x_{\cdot L}; d_{11}, \ldots, d_{KL}).$$

The diffusion is then constrained to the $(KL-1)$-dimensional simplex $\Delta_{K \times L}$ by imposing the conditions

$$(3.2) \qquad \sum_{i=1}^{K} x_{i\cdot} = 1; \quad \sum_{j=1}^{L} x_{\cdot j} = 1; \quad \sum_{i=1}^{K} d_{ij} = 0, \forall j; \quad \sum_{j=1}^{L} d_{ij} = 0, \forall i.$$

3.2. *Rescaling LD.* Since we are interested in the large ρ limit, we expect each d_{ij} to be small. We introduce the rescaling $\tilde{d}_{ij} = \sqrt{\rho}\, d_{ij}$. The reason for this choice should become clear from equation (3.3) below; in the resulting generator, there should be a nontrivial interaction between recombination and genetic drift, i.e. they should both act on the fastest timescale. The leading order contribution to the generator, denoted below by $L^{(2)}$, has contributions from both of these biological processes if and only if we use this choice for $\tilde{d}_{ij}$. See Song and Song (2007) for another example of this rescaling. By substituting for the new variables in (3.1) and using (3.2) for extensive simplification, the generator can be expressed as

$$(3.3) \qquad \widetilde{\mathscr{L}} = \frac{1}{2} \left[\rho L^{(2)} + \sqrt{\rho} L^{(1)} + L^{(0)} + \frac{1}{\sqrt{\rho}} L^{(-1)} \right],$$

where the operators in (3.3) are given by

$$L^{(2)} = \sum_{i,j} \left\{ x_{i\cdot} x_{\cdot j} \left[\sum_{k,l} (\delta_{ik} - x_{k\cdot})(\delta_{jl} - x_{\cdot l}) \frac{\partial}{\partial \tilde{d}_{kl}} \right] - \tilde{d}_{ij} \right\} \frac{\partial}{\partial \tilde{d}_{ij}},$$

$$L^{(1)} = \sum_{i,j,k,l} \left[2\tilde{d}_{ij}(\delta_{ik} - x_{k\cdot})(\delta_{jl} - x_{\cdot l}) - \delta_{ik}\delta_{jl}\tilde{d}_{ij} + 2\tilde{d}_{il}x_{k\cdot}x_{\cdot j} \right] \frac{\partial^2}{\partial \tilde{d}_{kl} \partial \tilde{d}_{ij}},$$

$$L^{(0)} = -\sum_{i,j,k,l} \tilde{d}_{ij}\tilde{d}_{kl} \frac{\partial^2}{\partial \tilde{d}_{kl} \partial \tilde{d}_{ij}} + 2\sum_{i,k,l} \left[(\delta_{ik} - x_{i\cdot})\tilde{d}_{kl} - \tilde{d}_{il}x_{k\cdot} \right] \frac{\partial^2}{\partial \tilde{d}_{kl} \partial x_{i\cdot}}$$

$$+ 2\sum_{j,k,l} \left[(\delta_{jl} - x_{\cdot j})\tilde{d}_{kl} - \tilde{d}_{kj}x_{\cdot l} \right] \frac{\partial^2}{\partial \tilde{d}_{kl} \partial x_{\cdot j}}$$

$$+ \sum_{i,k} x_{i\cdot}(\delta_{ik} - x_{k\cdot}) \frac{\partial^2}{\partial x_{k\cdot} \partial x_{i\cdot}} + \sum_{j,l} x_{\cdot j}(\delta_{jl} - x_{\cdot l}) \frac{\partial^2}{\partial x_{\cdot l} \partial x_{\cdot j}}$$

$$+ \sum_{i,j} \left[\theta_A \sum_k \tilde{d}_{kj}(P_{ki}^A - \delta_{ik}) + \theta_B \sum_l \tilde{d}_{il}(P_{lj}^B - \delta_{jl}) - 2\tilde{d}_{ij} \right] \frac{\partial}{\partial \tilde{d}_{ij}}$$

$$+ \frac{\theta_A}{2} \sum_{i,k} x_{k\cdot}(P^A_{ki} - \delta_{ik})\frac{\partial}{\partial x_{i\cdot}} + \frac{\theta_B}{2} \sum_{j,l} x_{\cdot l}(P^B_{lj} - \delta_{jl})\frac{\partial}{\partial x_{\cdot j}},$$

$$L^{(-1)} = 2\sum_{i,j} \widetilde{d}_{ij} \frac{\partial^2}{\partial x_{i\cdot}\partial x_{\cdot j}}.$$

This generator extends that of Ohta and Kimura (1969a) from a (2×2)- to a $(K \times L)$-allele model. Note that ours differs from that of Ohta and Kimura (1969a) by a factor of two; one unit of time corresponds to $2N$ (rather than N) generations in our convention.

Recall that our interest is in calculating the expectation at stationarity of the function $F(\boldsymbol{x}; \boldsymbol{n})$ shown in (2.4), which is now viewed as a function of

$$\widetilde{\boldsymbol{x}} = (x_{1\cdot}, \ldots, x_{K\cdot}; x_{\cdot 1}, \ldots, x_{\cdot L}; \widetilde{d}_{11}, \ldots, \widetilde{d}_{KL}).$$

In the same way that the multiplicity matrix $\boldsymbol{c}$ represents multinomial samples from a population with frequencies $(x_{ij})_{i\in[K],j\in[L]}$, we introduce an analogous matrix $\boldsymbol{r} = (r_{ij})_{i\in[K],j\in[L]}$ associated with the variables $(\widetilde{d}_{ij})_{i\in[K],j\in[L]}$. We further define the marginal vectors $\boldsymbol{r}_A = (r_{i\cdot})_{i\in[K]}$ and $\boldsymbol{r}_B = (r_{\cdot j})_{j\in[L]}$, where $r_{i\cdot} = \sum_j r_{ij}$ and $r_{\cdot j} = \sum_i r_{ij}$, analogous to $\boldsymbol{c}_A$ and $\boldsymbol{c}_B$. In this notation, the function $F(\boldsymbol{x}; \boldsymbol{n})$ becomes

$$F(\widetilde{\boldsymbol{x}}; \boldsymbol{n}) = \left(\prod_{i=1}^{K} x_{i\cdot}^{a_i}\right)\left(\prod_{j=1}^{L} x_{\cdot j}^{b_j}\right)\left(\prod_{i=1}^{K}\prod_{j=1}^{L}\left[\frac{\widetilde{d}_{ij}}{\sqrt{\rho}} + x_{i\cdot}x_{\cdot j}\right]^{c_{ij}}\right)$$

$$(3.4) \qquad = \sum_{m=0}^{c} \frac{1}{\rho^{\frac{m}{2}}} \sum_{\boldsymbol{r}\in\mathcal{P}_m} \left[\prod_{i,j}\binom{c_{ij}}{r_{ij}}\right] G^{(m)}(\widetilde{\boldsymbol{x}}; \boldsymbol{a} + \boldsymbol{c}_A - \boldsymbol{r}_A, \boldsymbol{b} + \boldsymbol{c}_B - \boldsymbol{r}_B, \boldsymbol{r}),$$

where

$$(3.5) \qquad G^{(m)}(\widetilde{\boldsymbol{x}}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = \left(\prod_{i=1}^{K} x_{i\cdot}^{a_i}\right)\left(\prod_{j=1}^{L} x_{\cdot j}^{b_j}\right)\left(\prod_{i=1}^{K}\prod_{j=1}^{L} \widetilde{d}_{ij}^{r_{ij}}\right),$$

and the inner summation in (3.4) is over all $K \times L$ matrices $\boldsymbol{r}$ of nonnegative integers whose entries sum to m:

$$\mathcal{P}_m = \left\{\boldsymbol{r} \in \mathbb{N}^{K\times L} : \sum_{i=1}^{K}\sum_{j=1}^{L} r_{ij} = m\right\}.$$

Note that only those matrices which form "subsamples" of $\boldsymbol{c}$ have nonzero coefficient in (3.4); i.e., $0 \leq r_{ij} \leq c_{ij}$ for all i and j.

3.3. *The key algorithm.* We now pose an asymptotic expansion for the expectation $\mathbb{E}[G^{(m)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})]$:

$$(3.6) \quad \mathbb{E}[G^{(m)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})] = g_0^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) + \frac{g_1^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})}{\sqrt{\rho}} + \frac{g_2^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})}{\rho} + \cdots,$$

so that, using (3.4), the quantity of interest is given by

$$q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \mathbb{E}[F(\widetilde{\boldsymbol{X}}; \boldsymbol{n})]$$

$$(3.7) \quad = \sum_{m=0}^{c} \sum_{\boldsymbol{r} \in \mathcal{P}_m} \left[\prod_{i=1}^{K} \prod_{j=1}^{L} \binom{c_{ij}}{r_{ij}} \right] \sum_{u=0}^{\infty} \frac{g_u^{(m)}(\boldsymbol{a} + \boldsymbol{c}_A - \boldsymbol{r}_A, \boldsymbol{b} + \boldsymbol{c}_B - \boldsymbol{r}_B, \boldsymbol{r})}{\rho^{\frac{1}{2}(m+u)}}.$$

We also have the boundary conditions

$$(3.8) \qquad q(\boldsymbol{e}_i, \boldsymbol{0}, \boldsymbol{0}) = \pi_i^A, \quad q(\boldsymbol{0}, \boldsymbol{e}_j, \boldsymbol{0}) = \pi_j^B, \quad q(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{0}) = \pi_i^A \pi_j^B,$$

where $\boldsymbol{\pi}^A = (\pi_i^A)_{i \in [K]}$ and $\boldsymbol{\pi}^B = (\pi_j^B)_{j \in [L]}$ are the stationary distributions of $\boldsymbol{P}^A$ and $\boldsymbol{P}^B$, respectively.

Using equations (3.7) and (3.8), we can also assign boundary conditions for each $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$:

$$(3.9) \quad \begin{aligned} g_0^{(0)}(\boldsymbol{e}_i, \boldsymbol{0}, \boldsymbol{0}) &= \pi_i^A, & g_u^{(0)}(\boldsymbol{e}_i, \boldsymbol{0}, \boldsymbol{0}) &= 0, \ \forall u \geq 1, \\ g_0^{(0)}(\boldsymbol{0}, \boldsymbol{e}_j, \boldsymbol{0}) &= \pi_j^B, & g_u^{(0)}(\boldsymbol{0}, \boldsymbol{e}_j, \boldsymbol{0}) &= 0, \ \forall u \geq 1, \\ g_0^{(0)}(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{0}) &= \pi_i^A \pi_j^B, & g_u^{(0)}(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{0}) &= 0, \ \forall u \geq 1. \end{aligned}$$

We have reduced the problem of computing an asymptotic expansion for $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ to one of computing $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ for each m and u. Consider arranging these quantities in a $c \times \mathbb{N}$ array, as illustrated in Figure 1. Refer to entries on the ℓth anti-diagonal, such that $m + u = \ell$, as residing on the ℓth *level*. As is clear from (3.7), the contribution in the expansion for $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ of order $\rho^{-\ell/2}$ is comprised of entries on level ℓ. For convenience, we define $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = 0$ if $u < 0$, or $m < 0$, or if any entry $a_i, b_j, r_{ij} < 0$. Then the following theorem, proved in Section 7.1, enables the level-wise computation of each $g_u^{(m)}$:

THEOREM 3.1. *The term $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ in the right hand side of (3.6) is determined as follows.*

(i) *For $m = 0$ and $u = 0$, $g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) = q^A(\boldsymbol{a}) q^B(\boldsymbol{b})$. (Recall that $q^A(\boldsymbol{a})$ and $q^B(\boldsymbol{b})$ are the respective one-locus sampling distributions at locus A and locus B.)*
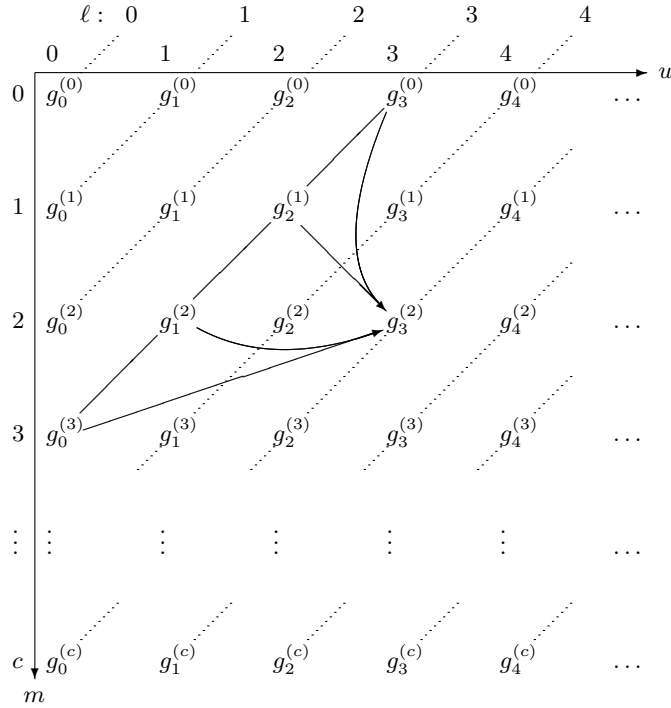
FIG 1. *Computation of $g_u^{(m)}$ for each $m = 0, \ldots, c$, and $u = 0, 1, \ldots$. When $m > 0$, the term $g_u^{(m)}$ residing on level $\ell = m + u$ is determined by up to four entries on level $\ell - 2$. This is illustrated for the example $g_3^{(2)}$.*

(ii) For $m > 0$ and $u \geq 0$, $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ on level ℓ is determined by at most four entries on level $\ell - 2$; these are $g_u^{(m-2)}$, $g_{u-1}^{(m-1)}$, $g_{u-2}^{(m)}$, and $g_{u-3}^{(m+1)}$. This relationship is given explicitly (equation (7.2)) in Section 7.

(iii) For $m = 0$ and $u \geq 1$, $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ on level ℓ is determined by $g_{u-1}^{(1)}$, also on level ℓ, and other similar terms $g_u^{(0)}(\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{0})$, where $a' \leq a$, $b' \leq b$. This relationship is given explicitly (equation (7.3)) in Section 7.

For odd levels (i.e., $\ell = 1, 3, 5, \ldots$), the above theorem implies the following vanishing result, which we prove in Section 7.2:

COROLLARY 3.1. *If $m + u = \ell$ is odd, then $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = 0$, for all configurations $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$.*

Incidentally, Corollary 3.1 implies that only integral powers of ρ have nonzero coefficients in (3.7).

Given the entries on level $\ell - 2$, Theorem 3.1 provides a method for computing each of the entries on level ℓ. They can be computed in any order, apart from the slight complication of (iii), which requires knowledge of $g_{u-1}^{(1)}$ as a prerequisite to $g_u^{(0)}$. The expression for $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ is only known recursively in $\boldsymbol{a}$ and $\boldsymbol{b}$, and we do not have a closed-form solution for this recursion for $u \geq 4$ and even. Equation (2.10) provides one example, which turns out to be the recursion for $g_4^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$. If the marginal one-locus sampling distributions are known, the complexity of computing $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ for $m > 0$ does not depend on the sample configuration $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$. In contrast, the complexity of computing $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ depends on $\boldsymbol{a}$ and $\boldsymbol{b}$, and the running time generally grows with sample size. However, in Section 6 we show that ignoring $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ generally leads to little loss of accuracy in practice.

To illustrate the method, entries on the first two even levels are summarized in Table 1. These recapitulate part of the results given in Theorem 2.1. The last column of Table 1 gives the "contribution" to (2.6) from $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ (for fixed u and m and summing over the relevant $\boldsymbol{r}$). According to (3.7), this quantity is

$$(3.10) \qquad \sum_{\boldsymbol{r} \in \mathcal{P}_m} \left[\prod_{i=1}^{K} \prod_{j=1}^{L} \binom{c_{ij}}{r_{ij}} \right] g_u^{(m)}(\boldsymbol{a} + \boldsymbol{c}_A - \boldsymbol{r}_A, \boldsymbol{b} + \boldsymbol{c}_B - \boldsymbol{r}_B, \boldsymbol{r}).$$

We have also checked that the total contribution from entries on level $\ell = 4$ is equal to $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, as given in Theorem 2.1. We note in passing that Theorem 3.1 makes it transparent why $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is *not* universal in the sense described in Section 2.3: expressions on level $\ell = 4$ depend directly on $L^{(0)}$, which in turn depends upon the model of mutation. By contrast, the nonzero contribution to $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, for example, is determined by $L^{(2)}$, which does not depend on the model of mutation.

It is important to emphasize that $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ is a function of the vectors $\boldsymbol{a}$, $\boldsymbol{b}$, the matrix $\boldsymbol{r}$, and (implicitly) the parameters θ_A and θ_B. The relationships given in Theorem 3.1 are thus *functional*, and only need to be computed once. In other words, all of these arguments of $g_u^{(m)}$ can remain arbitrary. It is not necessary to redo any of the algebraic computations for each particular choice of sample configuration, for example. Moreover, the solutions to each $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ are expressed concisely in terms of the marginal one-locus sampling distributions q^A and q^B; this fact follows inductively from the solution for $g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$. Unlike the method of Jenkins and Song (2009, 2010), the iterative procedure here is essentially the same at every step.

<div align="center">Table 1</div>
<div align="center">Entries on levels $\ell = 0, 2$ of the array $(g_u^{(m)})$.</div>

ℓ	u	m	$g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$	Contribution to (2.6)
0	0	0	$\mathbb{E}\left[\prod_i X_{i\cdot}^{a_i} \prod_j X_{\cdot j}^{b_j}\right]$	$q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$
2	0	2	$\mathbb{E}\left[X_{i\cdot}X_{\cdot j}(\delta_{ik} - X_{k\cdot})(\delta_{jl} - X_{\cdot l})\prod_u X_{u\cdot}^{a_u} \prod_v X_{\cdot v}^{b_v}\right],$ where $\boldsymbol{r} = \boldsymbol{e}_{ij} + \boldsymbol{e}_{kl}$	$q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$
	1	1	0	0
	2	0	0	0

4. Partial sums, optimal truncation, and Padé approximants.
In principle, the procedure described in the previous section provides a method of computing an arbitrary number of terms in the asymptotic expansion (2.6), for any sample configuration. Suppose the computation has been carried out up to level $\ell = 2M$ and consider the partial sum

$$(4.1) \qquad q_{\mathrm{PS}}^{(M)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) + \frac{q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho} + \cdots + \frac{q_M(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho^M}.$$

Since we do not know its radius of convergence, we should be prepared for this sum to diverge eventually as M increases. An important question that we address in this section is: How many terms should we use to maximize the accuracy of the approximation?

4.1. *Optimal truncation.* As mentioned above, simply adding more and more terms to the asymptotic expansion may decrease the accuracy beyond a certain point. Optimal truncation is a rule of thumb for truncating the partial sum at a point that is expected to maximize its accuracy. More precisely, it is defined as follows.

DEFINITION 4.1 (Optimal truncation rule). *Given the first $M + 1$ terms $q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, ..., $q_M(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, in the asymptotic expansion, let M' be the index such that $\left|q_{M'}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})/\rho^{M'}\right| < \left|q_{M''}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})/\rho^{M''}\right|$, for all $M'' \neq M'$, where $M', M'' \leq M$. Then, the optimal truncation rule (OTR) suggests truncating the sum at order M':*

$$q_{\mathrm{OTR}}^{(M)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) + \frac{q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho} + \cdots + \frac{q_{M'}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\rho^{M'}}.$$

The motivation for this rule is that the magnitude of the ith term in the expansion is an estimate of the magnitude of the remainder. More sophisticated versions of the OTR are available (e.g. Dingle, 1973, Chapter XXI), but for simplicity we focus on the definition given above.

There are two issues with the OTR. First, it minimizes the error only approximately, and so, despite its name, it is not be guaranteed to be optimal. For example, the magnitude of the first few terms in the series can behave very irregularly before a pattern emerges. Second, it may use only the first few terms in the expansion and discard the rest. As we discuss later, for some sample configurations and parameter values of interest, the OTR might truncate very early, even as early as $M' = 2$. This is unrelated to the first issue, since the series may indeed begin to diverge very early. Below, we discuss a better approximation scheme with a provable convergence property.

4.2. *Padé approximants.*   The key idea behind Padé approximants is to approximate the function of interest by a rational function. In contrast to the OTR, Padé approximants make use of *all* of the computed terms in the expansion, even when the expansion diverges rapidly. More precisely, the $[U/V]$ Padé approximant of a function is defined as follows.

DEFINITION 4.2 ($[U/V]$ Padé approximant).   *Given a function f and two nonnegative integers U and V, the $[U/V]$ Padé approximant of f is a rational function of the form*

$$[U/V]_f(x) = \frac{A_0 + A_1 x + \cdots + A_U x^U}{B_0 + B_1 x + \cdots + B_V x^V},$$

such that $B_0 = 1$ and

$$f(x) - [U/V]_f(x) = O(x^{U+V+1}).$$

That is, the first $U + V + 1$ terms in a Maclaurin series of the Padé approximant $[U/V]_f(x)$ matches the first $U + V + 1$ terms in a Maclaurin series of f.

Our goal is to approximate the sampling distribution $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ by the Padé approximant

$$(4.2) \qquad q_{\text{Padé}}^{[U/V]}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = [U/V]_{q(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}\left(\frac{1}{\rho}\right),$$

such that the first $U + V + 1$ terms in a Maclaurin series of $[U/V]_{q(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}(\frac{1}{\rho})$ agrees with (4.1), where $M = U + V$. (In this notation, $[U/V]_{q(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}(\frac{1}{\rho})$ is

an implicit function of the mutation parameters.) As more terms in (4.1) are computed (i.e., as M increases), a sequence of Padé approximants can be constructed. This sequence often has much better convergence properties than (4.1) itself (Baker and Graves-Morris, 1996).

For a given M, there is still some freedom over the choice of U and V. As M increases, we construct the following "staircase" sequence of Padé approximants: $[0/0], [0/1], [1/1], [1/2], [2/2], \ldots$. This scheme is motivated by the following lemma, proved in Section 7.3:

LEMMA 4.1. *Under a neutral, finite-alleles model, the sampling distribution $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is a rational function of $1/\rho$, and the degree of the numerator is equal to the degree of the denominator.*

This simple yet powerful observation immediately leads to a convergence result for the Padé approximants in the following manner:

THEOREM 4.1. *Consider a neutral, finite-alleles model. For every given two-locus sample configuration $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, there exists a finite nonnegative integer U_0 such that for all $U \geq U_0$ and $V \geq U_0$, the Padé approximant $q_{\mathrm{Padé}}^{[U/V]}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is exactly equal to $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ for all $\rho \geq 0$.*

A proof of this theorem is provided in Section 7.4. Note that the staircase sequence is the "quickest" to reach the true sampling distribution $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. Although Theorem 4.1 provides very strong information about the convergence of the Padé approximants, in practice U and V will have to be intractably large for such convergence to take place. The real value of Padé summation derives from the empirical observation that the approximants exhibit high accuracy even *before* they hit the true sampling distribution. The staircase sequence also has the advantage that it exhibits a continued fraction representation, which enables their construction to be made computationally more efficient (Baker and Graves-Morris, 1996, Chapter 4).

5. Incorporating selection. We now incorporate a model of diploid selection into the results of Section 3. Suppose that a diploid individual is composed of two haplotypes $(i, j), (k, l) \in [K] \times [L]$, and that, without loss of generality, selective differences exist at locus A. We denote the fitness of this individual by $1 + s_{ik}^A$, and consider the diffusion limit in which $\sigma_{ik}^A = 4N s_{ik}^A$ is held fixed for each $i, k \in [K]$ as $N \to \infty$.

In what follows, we use a subscript "s" to denote selectively nonneutral versions of the quantities defined above. Results will be given in terms of the nonneutral one-locus sampling distribution q_{s}^A at locus A and the neutral one-locus sampling distribution q^B at locus B.

5.1. *One-locus sampling distribution under selection.* For the infinite-alleles model, one-locus sampling distributions under symmetric selection have been studied by Grote and Speed (2002), Handa (2005), and Huillet (2007). In the case of a parent-independent finite-alleles model, the stationary distribution of the one-locus selection model is known to be a *weighted* Dirichlet distribution (Wright, 1949):

$$\pi_{\mathrm{s}}^A(\boldsymbol{x}) = D \left(\prod_{i=1}^K x_i^{\theta_A P_i^A - 1} \right) \exp \left(\frac{1}{2} \sum_{i=1}^K \sum_{k=1}^K \sigma_{ik}^A x_i x_k \right),$$

where $\boldsymbol{x} \in \Delta_K$ [see (2.1)] and D is a normalizing constant. The one-locus sampling distribution at stationarity is then obtained by drawing a multinomial sample from this distribution:

$$(5.1) \qquad q_{\mathrm{s}}^A(\boldsymbol{a}) = \mathbb{E}_{\mathrm{s}} \left[\prod_{i=1}^K X_i^{a_i} \right].$$

Thus, under a diploid selection model with parent-independent mutation, we are able to express the one-locus sampling distribution at least in integral form. There are two integrals that need to be evaluated: one for the expectation and the other for the normalizing constant. In practice, these integrals must be evaluated using numerical methods (Donnelly et al., 2001).

5.2. *Two-locus sampling distribution with one locus under selection.* To incorporate selection into our framework, we first introduce some further notation.

DEFINITION 5.1. *Given two-locus population-wide allele frequencies $\boldsymbol{x} \in \Delta_{K \times L}$ [see (2.3)], the mean fitness of the population at locus A is*

$$\bar{\sigma}^A(\boldsymbol{x}_A) = \sum_{i,k} \sigma_{ik}^A x_{i\cdot} x_{k\cdot}.$$

Selection has an additive effect on the generator of the process, which is now given by

$$\mathscr{L}_{\mathrm{s}} = \mathscr{L} + \frac{1}{2} \sum_{i,j} x_{ij} \left[\sum_k \sigma_{ik}^A x_{k\cdot} - \bar{\sigma}^A(\boldsymbol{x}_A) \right] \frac{\partial}{\partial x_{ij}},$$

where $\mathscr{L}$ is the generator (3.1) of the neutral diffusion process [see, for example, Ethier and Nagylaki (1989)]. Rewriting $\mathscr{L}_{\mathrm{s}}$ in terms of the LD variables and then rescaling d_{ij} as before, we obtain

$$\widetilde{\mathscr{L}_{\mathrm{s}}} = \widetilde{\mathscr{L}} + \frac{1}{2} \left[\rho L_{\mathrm{s}}^{(2)} + \sqrt{\rho} L_{\mathrm{s}}^{(1)} + L_{\mathrm{s}}^{(0)} + \frac{1}{\sqrt{\rho}} L_{\mathrm{s}}^{(-1)} \right],$$

where $\widetilde{\mathscr{L}}$ is as in (3.3), and the new contributions are

$$L_{\mathrm{s}}^{(2)} = 0,$$
$$L_{\mathrm{s}}^{(1)} = 0,$$
$$L_{\mathrm{s}}^{(0)} = \sum_{i,j} \left[d_{ij} \left(\sum_k \sigma_{ik}^A x_{k\cdot} - \bar{\sigma}^A(\boldsymbol{x}_A) \right) - \sum_{k,k'} d_{kj}\sigma_{kk'}^A x_{i\cdot} x_{k'\cdot} \right] \frac{\partial}{\partial d_{ij}}$$
$$+ \sum_i x_{i\cdot} \left(\sum_k \sigma_{ik}^A x_{k\cdot} - \bar{\sigma}^A(\boldsymbol{x}_A) \right) \frac{\partial}{\partial x_{i\cdot}},$$
$$L_{\mathrm{s}}^{(-1)} = \sum_{i,j,k} \widetilde{d}_{ij}\sigma_{ik}^A x_{k\cdot} \frac{\partial}{\partial x_{\cdot j}}.$$

In addition, we replace the asymptotic expansion (3.6) with

$$\mathbb{E}_{\mathrm{s}}[G^{(m)}(\widetilde{\boldsymbol{X}};\boldsymbol{a},\boldsymbol{b},\boldsymbol{r})] = h_0^{(m)}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{r}) + \frac{h_1^{(m)}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{r})}{\sqrt{\rho}} + \frac{h_2^{(m)}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{r})}{\rho} + \cdots,$$

the corresponding expansion for the expectation of $G^{(m)}(\widetilde{\boldsymbol{X}};\boldsymbol{a},\boldsymbol{b},\boldsymbol{r})$ with respect to the stationary distribution *under selection* at locus A. Finally, the boundary conditions (3.9) become the following (Fearnhead, 2003):

$$\begin{array}{ll}
h_0^{(0)}(\boldsymbol{e}_i,\boldsymbol{0},\boldsymbol{0}) = \phi_i^A, & h_u^{(0)}(\boldsymbol{e}_i,\boldsymbol{0},\boldsymbol{0}) = 0, \ \forall u \geq 1, \\
h_0^{(0)}(\boldsymbol{0},\boldsymbol{e}_j,\boldsymbol{0}) = \pi_j^B, & h_u^{(0)}(\boldsymbol{0},\boldsymbol{e}_j,\boldsymbol{0}) = 0, \ \forall u \geq 1, \\
h_0^{(0)}(\boldsymbol{e}_i,\boldsymbol{e}_j,\boldsymbol{0}) = \phi_i^A \pi_j^B, & h_u^{(0)}(\boldsymbol{e}_i,\boldsymbol{e}_j,\boldsymbol{0}) = 0, \ \forall u \geq 1,
\end{array}$$
(5.2)

where $\boldsymbol{\phi}^A = (\phi_i^A)_{i \in [K]}$ is the stationary distribution for drawing a sample of size one from a single selected locus.

With only minor modifications to the arguments of Section 3, each term in the array for $h_u^{(m)}$ can be computed in a manner similar to Theorem 3.1. In particular, entries on odd levels are still zero. Furthermore, as proved in Section 7.5, we can update Theorem 2.1 as follows:

THEOREM 5.1. *Suppose locus A is under selection, while locus B is selectively neutral. Then, in the asymptotic expansion (2.6) of the two-locus sampling distribution, the zeroth and first order terms are given by (2.7) and (2.8), respectively, with $q_{\mathrm{s}}^A(\boldsymbol{a})$ in place of $q^A(\boldsymbol{a})$. Furthermore, the second order term (2.9) may now be decomposed into two parts:*

$$q_{2,\mathrm{s}}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = q_{2,\mathrm{s}}(\boldsymbol{a}+\boldsymbol{c}_A,\boldsymbol{b}+\boldsymbol{c}_B,\boldsymbol{0}) + \sigma_{\mathrm{s}}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}),$$
(5.3)

*where $\sigma_{\mathrm{s}}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is given by a known analytic expression and $q_{2,\mathrm{s}}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ sat-
isfies a slightly modified version of the recursion relation (2.10) for $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$.
(These expressions are omitted for brevity.)*

We remark that the above arguments can be modified to allow for locus B
also to be under selection, provided the selection is independent, with no
epistatic interactions, and provided one can substitute ϕ_j^B for π_j^B in (5.2).
Then, one could also simply substitute $q_{\mathrm{s}}^B(\boldsymbol{b})$ for $q^B(\boldsymbol{b})$ in the expressions
for $q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ and $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. However, for the boundary conditions (5.2) to
be modified in this way we would need to extend the result of Fearnhead
(2003) to deal with *two* nonneutral loci, and we are unaware of such a result
in the literature.

6. Empirical study of accuracy. In this section, we study empiri-
cally the accuracy of the approximate sampling distributions discussed in
Section 4.

6.1. *Computational details.* As discussed earlier, a major advantage of
our technique is that, given the first M terms in the asymptotic expansion
(2.6), the $(M + 1)$th term can be found and has to be computed only once.
There are two complications to this statement: First, as mentioned in the
discussion following Theorem 3.1, the Mth order term q_M for $M \geq 2$ has
a contribution (namely, $g_{2M}^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$) that is *not* known in closed form, and
is only given recursively. (Recall that the $M = 1$ case is an exception, with
$g_2^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) = 0$ for all $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$.) In Jenkins and Song (2009), it was ob-
served that the contribution of $g_4^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ to $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is generally very
small, but that its burden in computational time increases with sample size.
Extrapolating this observation to higher order terms, we consider making
the following approximation:

APPROXIMATION 6.1. *For all $M \geq 2$, assume*

$$g_{2M}^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) \approx 0,$$

for all configurations $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$.

As we show presently, adopting this approximation has little effect on the
accuracy of asymptotic sampling distributions. In what follows, we use the
symbol "∘" to indicate when the above approximation is employed. For ex-
ample, the partial sum $q_{\mathrm{PS}}^{(M)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ in (4.1) becomes $\mathring{q}_{\mathrm{PS}}^{(M)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ under
Approximation 6.1.

Upon making the above approximation, it is then possible to construct a closed-form expression for each subsequent term $\mathring{q}_M(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. However, there is a second issue: Given the effort required to reach the complicated expression for $\mathring{q}_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ (Jenkins and Song, 2009), performing the computation by hand for $M > 2$ does not seem tractable. Symbolic computation using computer software such as *Mathematica* is a feasible option, but we defer this for future work. Here, we are interested in comparing the accuracy of asymptotic sampling distributions with the true likelihood. Therefore, we have implemented an automated *numerical* computation of each subsequent term in the asymptotic expansion, for a given fixed sample configuration and fixed mutation parameters. For the samples investigated below, this did not impose undue computational burden, even when repeating this procedure across *all* samples of a given size. Exact numerical computation of the true likelihood is possible for only small sample sizes (say, up to thirty), so we restrict our study to those cases.

For simplicity, we assume in our empirical study that all alleles are selectively neutral. Furthermore, we assume a symmetric, PIM model so that $\boldsymbol{P}^A = \boldsymbol{P}^B = \left(\begin{smallmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{smallmatrix} \right)$, and take $\theta_A = \theta_B = 0.01$. This is a reasonable approximation for modeling neutral single nucleotide polymorphism (SNP) data in humans (e.g. McVean et al., 2002). For the PIM model, recall that the marginal one-locus sampling distributions are available in closed-form, as shown in (2.5).

6.2. *Rate of convergence: an example.* To compare the convergence of the sequence of partial sums (4.1) with that of the sequence of Padé approximants (4.2), we re-examine in detail an example studied previously. The sample configuration for this example is $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$, $\boldsymbol{c} = \left(\begin{smallmatrix} 10 & 7 \\ 2 & 1 \end{smallmatrix} \right)$. In Jenkins and Song (2009), we were able to compute the first three terms in the asymptotic expansion, obtaining the partial sum $q_{\mathrm{PS}}^{(2)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. Applying the new method described in this paper, we computed $q_{\mathrm{PS}}^{(M)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ for $M \leq 11$ (including the recursive terms $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ discussed above), and also the corresponding staircase Padé approximants $q_{\mathrm{Padé}}^{[U/V]}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. Results are illustrated in Figure 2, in which we compare various approximations of the likelihood curve for ρ with the true likelihood curve. Here, the likelihood of a sample is defined simply as its sampling distribution $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ treated as a function of ρ, with θ_A and θ_B fixed at 0.01.

Figure 2(a) exhibits a number of features which we also observed more generally for many other samples. For fixed ρ in the range illustrated, the sequence $(q_{\mathrm{PS}}^{(0)}, q_{\mathrm{PS}}^{(1)}, q_{\mathrm{PS}}^{(2)}, \ldots)$ of partial sums diverges eventually, and, for many realistic choices of ρ, this divergence can be surprisingly early. Arguably
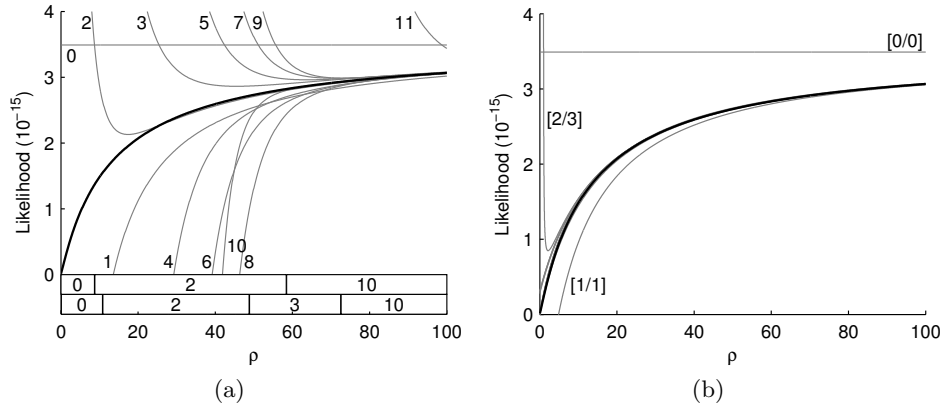
FIG 2. *Likelihood curves for ρ, comparing different levels of truncation, for the sample $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$, $\boldsymbol{c} = \left(\begin{smallmatrix} 10 & 7 \\ 2 & 1 \end{smallmatrix} \right)$. A symmetric PIM model with $\theta_A = \theta_B = 0.01$ is assumed. The true likelihood is shown as a thick black line. Approximate likelihood curves for various levels of truncation, $M = 0, 1, \ldots, 11$ are shown as thinner gray lines. (a) Partial sums $q_{\mathrm{PS}}^{(M)}$, labeled by M. The bottom row of indices records M', the level of truncation in $q_{\mathrm{OTR}}^{(11)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ recommended by the OTR. The row above records the actual level of truncation that minimizes the unsigned relative error. (b) Staircase Padé approximants $q_{\mathrm{Padé}}^{[U/V]}$, some labeled by $[U/V]$. To the naked eye, $q_{\mathrm{Padé}}^{[U/V]}$ for most $[U/V]$ are indistinguishable from the true likelihood curve.*

the best overall fit in the figure is $q_{\mathrm{PS}}^{(2)}$, though for $\rho \geq 70$ good accuracy is maintained by $q_{\mathrm{PS}}^{(M)}$ for all $1 \leq M \leq 10$. Divergence is extremely rapid if we add any further terms; witness the curve for $q_{\mathrm{PS}}^{(11)}$. Of course, in real applications the true likelihood will be unavailable and we might rely on the aforementioned optimal truncation rule to guide the truncation point. Here, it performs reasonably well, correctly picking the most accurate index across most of the range [compare the bottom two rows of indices in Figure 2(a)].

In contrast, Figure 2(b) shows that there is much less risk in using "too many" terms in constructing the Padé approximants. They approximate the true likelihood very accurately and very quickly; to the naked eye, Padé approximants $q_{\mathrm{Padé}}^{[U/V]}$ for most $[U/V]$ are indistinguishable from the true likelihood curve. Indeed, this example suggests that there is very little gain in accuracy beyond $q_{\mathrm{Padé}}^{[0/1]}$, but that there is no significant loss beyond it either. (It should be pointed out, however, that achievement of such high accuracy so early in the sequence of approximants does not seem to be a common occurrence across samples. Usually, several more terms are required; see next section for further details.)

There is one further important observation to be made regarding Padé ap-

TABLE 2

Nonnegative, real roots in the numerator and denominator of the $[U/V]$ Padé approximants for the sample $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$, $\boldsymbol{c} = \left(\begin{smallmatrix} 10 & 7 \\ 2 & 1 \end{smallmatrix}\right)$.

$[U/V]$	Roots of numerator		Roots of denominator	
$[0/0]$				
$[0/1]$	0.000			
$[1/1]$	4.871			
$[1/2]$	0.000			
$[2/2]$				
$[2/3]$	0.000		0.912	
$[3/3]$				
$[3/4]$	0.000			
$[4/4]$				
$[4/5]$	0.000			
$[5/5]$				
$[5/6]$	0.000	8474538.140	8474538.140	
$[6/6]$				
$[6/7]$	0.000	306.846	306.846	
$[7/7]$				
$[7/8]$	0.000			
$[8/8]$	82.033		82.032	
$[8/9]$	0.000	77.366	0.284	77.364
$[9/9]$	82.121	4412.751	82.120	4412.751
$[9/10]$	0.000			
$[10/10]$	5.543		4.252	

proximants. There is nothing to prevent the polynomial in the denominator from having positive real roots, thus creating singularities. This is indeed observed once in Figure 2(b); $q_{\text{Padé}}^{[2/3]}$ exhibits a singularity at $\rho = 0.9$. To examine this behavior further, in Table 2 we tabulate the nonnegative real roots of the numerator and denominator of each approximant in the staircase sequence up to $[10/10]$. We see some interesting patterns: (i) The total number of nonnegative real roots does not seem to grow with M. (ii) Roots in the denominator are almost invariably accompanied by a nearby root in the numerator, and their proximity is usually extremely close, with agreement to several decimal places. (iii) Pairs of roots appear transiently; the roots of one approximant in the sequence provide almost no information regarding the next. Such pairs of roots are known as *defects*, and are an inevitable feature of Padé approximants (Baker and Graves-Morris, 1996, p48). Provided we anticipate them, defects need not act as a serious obstacle. Because of the proximity of the two roots in a pair, they almost behave like removable singularities. In particular, Padé approximants are still well-behaved outside a reasonably narrow interval around the defect, and can still approximate the likelihood accurately. In light of these observations, henceforth we use

the following simple heuristic for dealing with defects.

DEFINITION 6.1 (Defect heuristic). *Suppose we are interested in approximating the likelihood curve at a particular value ρ_0 and have the resources to compute up to M in the partial sum (4.1). Then proceed as follows.*

1. *Initialize with $M'' = M$.*
2. *Construct the $[U/V]$ Padé approximant in the staircase sequence, where $U + V = M''$.*
3. *If it exhibits a root in the interval $(\rho_0 - \epsilon, \rho_0 + \epsilon) \cap [0, \infty)$, either in its numerator or denominator, then decrement M'' by one and go to step 2, otherwise use this approximant.*

Choice of threshold ϵ involves a trade-off between the disruption caused by the nearby defect and the loss incurred by reverting to the previous Padé approximant. Throughout the following section, we employ this heuristic with $\epsilon = 25$, which seemed to work well.

6.3. *Rate of convergence: empirical study.* To investigate to what extent the observations of Section 6.2 hold across all samples, we performed the following empirical study. Following Jenkins and Song (2009), we focused on samples of the form $(\mathbf{0}, \mathbf{0}, \mathbf{c})$ for which all alleles are observed at both loci, and measured the accuracy of the partial sum (4.1) by the *unsigned relative error*

$$e_{\mathrm{PS}}^{(M)}(\mathbf{0}, \mathbf{0}, \mathbf{c}) = \left| \frac{q_{\mathrm{PS}}^{(M)}(\mathbf{0}, \mathbf{0}, \mathbf{c}) - q(\mathbf{0}, \mathbf{0}, \mathbf{c})}{q(\mathbf{0}, \mathbf{0}, \mathbf{c})} \right| \times 100\%.$$

An analogous definition can be made for $e_{\mathrm{Padé}}^{(M)}$, the unsigned relative error of the staircase Padé approximants $q_{\mathrm{Padé}}^{[U/V]}$, where $U = \lfloor M/2 \rfloor$ and $V = \lceil M/2 \rceil$. When Approximation 6.1 is additionally used, the respective unsigned errors are denoted $\mathring{e}_{\mathrm{PS}}^{(M)}$ and $\mathring{e}_{\mathrm{Padé}}^{(M)}$. These quantities are implicit functions of the parameters and of the sample configuration. In our study, we focused on sufficiently small sample sizes so that the true sampling distribution $q(\mathbf{a}, \mathbf{b}, \mathbf{c})$ could be computed. Specifically, we computed $q(\mathbf{0}, \mathbf{0}, \mathbf{c})$ for *all* samples of a given size c using the method described in Jenkins and Song (2009). By weighting samples according to their sampling probability, we may compute the distributions of $e_{\mathrm{PS}}^{(M)}$ and $e_{\mathrm{Padé}}^{(M)}$, and similarly the distributions of $\mathring{e}_{\mathrm{PS}}^{(M)}$ and $\mathring{e}_{\mathrm{Padé}}^{(M)}$. Table 3 summarizes the cumulative distributions of $\mathring{e}_{\mathrm{PS}}^{(M)}$ and $\mathring{e}_{\mathrm{Padé}}^{(M)}$ for $\rho = 50$, across all samples of size $c = 20$ that are dimorphic at both loci. The corresponding table for $e_{\mathrm{PS}}^{(M)}$ and $e_{\mathrm{Padé}}^{(M)}$ was essentially identical (not

TABLE 3

Cumulative distribution $\Phi(x) = \mathbb{P}(\mathring{e}^{(M)} < x\%)$ (where $\mathring{e}^{(M)}$ denotes either $\mathring{e}_{\text{PS}}^{(M)}$ or $\mathring{e}_{\text{Padé}}^{(M)}$) of the unsigned relative error of the partial sum $q_{\text{PS}}^{(M)}$ and the corresponding Padé approximants, for all samples of size 20 dimorphic at both loci. Here, $\rho = 50$, and Approximation 6.1 is used.

M	Type of sum	$\Phi(1)$	$\Phi(5)$	$\Phi(10)$	$\Phi(25)$	$\Phi(50)$	$\Phi(100)$
0	PS	0.49[†]	0.56[†]	0.63	0.81	0.98	1.00
	Padé	0.49	0.56	0.63	0.81	0.98	1.00
1	PS	0.51	0.74	0.87	0.99	1.00	1.00
	Padé	0.59	0.77	0.84	0.91	0.98	0.99
2	PS	0.59	0.87	0.92	1.00	1.00	1.00
	Padé	0.77	0.97	0.98	0.99	1.00	1.00
3	PS	0.57	0.86	0.97	1.00	1.00	1.00
	Padé	0.91	0.96	0.98	0.99	1.00	1.00
4	PS	0.45	0.62	0.87	0.98	0.98	1.00
	Padé	0.95	0.99	1.00	1.00	1.00	1.00
5	PS	0.30	0.50	0.61	0.76	0.79	0.97
	Padé	0.98	1.00	1.00	1.00	1.00	1.00
10	PS	0.00	0.02	0.03	0.07	0.08	0.11
	Padé	1.00	1.00	1.00	1.00	1.00	1.00
	OTR	0.25	0.36	0.65	0.89	0.90	1.00

† These two values were misquoted in the text of Jenkins and Song (2009, p. 1093); this table corrects them.

shown), with agreement usually to two decimal places. This confirms that utilizing Approximation 6.1 is justified.

Table 3 illustrates that the observations of Section 6.2 hold much more generally, as described below:

1. The error $\mathring{e}_{\text{PS}}^{(M)}$ for partial sums is not a monotonically decreasing function of M; i.e., the accuracy of $\mathring{q}_{\text{PS}}^{(M)}$ improves as one adds more terms up to a certain point, before quickly becoming very inaccurate.

2. Empirically, the actual optimal truncation point for the parameter settings we considered is at $M' = 2$ or $M' = 3$, which perform comparably. Moreover, both provide consistently higher accuracy than employing the OTR, which is a serious issue when we wish to use this rule without external information about which truncation point really is the most accurate.

3. Overall, using Padé approximants is much more reliable. Note that the accuracy of Padé approximants continues to improve as we incorporate more terms. For a sample drawn at random, the probability that its Padé approximant is within 1% of the true sampling distribution is 1.00, compared to 0.59 or 0.57 for truncating the partial sums

Table 4

Effect of ρ on the cumulative distribution $\Phi(x) = \mathbb{P}(\mathring{e}_{Pad\acute{e}}^{(M)} < x\%)$ of the unsigned relative error of the Padé approximants, for all samples of size 20 dimorphic at both loci.

M	ρ	$\Phi(1)$	$\Phi(5)$	$\Phi(10)$	$\Phi(25)$	$\Phi(50)$	$\Phi(100)$
0	25	0.39	0.52	0.58	0.69	0.84	1.00
	50	0.49	0.56	0.63	0.81	0.98	1.00
	100	0.50	0.61	0.72	0.97	1.00	1.00
	200	0.54	0.72	0.95	0.99	1.00	1.00
1	25	0.51	0.62	0.75	0.85	0.94	0.96
	50	0.59	0.77	0.84	0.91	0.98	0.99
	100	0.74	0.91	0.95	0.98	0.99	1.00
	200	0.90	0.98	0.99	1.00	1.00	1.00
2	25	0.59	0.82	0.91	0.94	0.95	0.97
	50	0.77	0.97	0.98	0.99	1.00	1.00
	100	0.95	1.00	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	1.00
3	25	0.64	0.91	0.95	0.96	0.98	1.00
	50	0.91	0.96	0.98	0.99	1.00	1.00
	100	0.99	1.00	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	1.00
4	25	0.83	0.96	0.99	0.99	1.00	1.00
	50	0.95	0.99	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	1.00
5	25	0.82	0.94	0.98	0.99	1.00	1.00
	50	0.98	1.00	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	1.00
10	25	0.97	0.99	0.99	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	1.00

respectively at $M' = 2$ or $M' = 3$, and only 0.25 for using the OTR.

For the remainder of this section, we focus on the accuracy of the staircase Padé approximants. It was shown in Jenkins and Song (2009) that the accuracy of the partial sum $q_{\mathrm{PS}}^{(2)}$ increases with ρ, but (perhaps surprisingly) decreases with increasing sample size. In Table 4 and Table 5, we address the same issue for the Padé approximant. Table 4 confirms that accuracy increases with ρ, as one might expect. Furthermore, it is also the case that substantial accuracy is achievable even for moderate values of ρ (say, $\rho = 25$), provided that sufficiently many terms are utilized in the construction of the Padé approximant. For example, when $M = 5$ the probability that the Padé approximant of a sample drawn at random is within 5% of the truth is 0.94.

TABLE 5

Effect of sample size c on the cumulative distribution $\Phi(x) = \mathbb{P}(\mathring{e}_{Pad\acute{e}}^{(M)} < x\%)$ of the unsigned relative error of the Padé approximants, for all samples of size c dimorphic at both loci. Here $\rho = 50$.

M	c	$\Phi(1)$	$\Phi(5)$	$\Phi(10)$	$\Phi(25)$	$\Phi(50)$	$\Phi(100)$
0	10	0.58	0.67	0.72	0.96	1.00	1.00
	20	0.49	0.56	0.63	0.81	0.98	1.00
	30	0.44	0.50	0.58	0.72	0.91	1.00
1	10	0.72	0.90	0.94	0.97	1.00	1.00
	20	0.59	0.77	0.84	0.91	0.98	0.99
	30	0.53	0.69	0.76	0.85	0.95	0.97
2	10	0.94	1.00	1.00	1.00	1.00	1.00
	20	0.77	0.97	0.98	0.99	1.00	1.00
	30	0.62	0.89	0.95	0.98	0.99	0.99
3	10	0.98	1.00	1.00	1.00	1.00	1.00
	20	0.91	0.96	0.98	0.99	1.00	1.00
	30	0.81	0.92	0.94	0.98	0.99	1.00
4	10	1.00	1.00	1.00	1.00	1.00	1.00
	20	0.95	0.99	1.00	1.00	1.00	1.00
	30	0.90	0.99	0.99	1.00	1.00	1.00
5	10	1.00	1.00	1.00	1.00	1.00	1.00
	20	0.98	1.00	1.00	1.00	1.00	1.00
	30	0.86	0.99	0.99	1.00	1.00	1.00
10	10	1.00	1.00	1.00	1.00	1.00	1.00
	20	1.00	1.00	1.00	1.00	1.00	1.00
	30	0.99	1.00	1.00	1.00	1.00	1.00

Also very encouraging is the pattern shown in Table 5. Provided that sufficiently many terms are used, the accuracy of the Padé approximant is only slightly affected by looking at larger sample sizes. For example, when $M = 5$ the probability that the Padé approximant of a randomly drawn sample of size 30 is within 5% of the truth is 0.99, compared with 1.00 for a sample of size 20. This loss in accuracy is much less severe than the corresponding loss in accuracy of the partial sums; for $q_{\mathrm{PS}}^{(M)}$, the highest accuracy is achieved for $M = 2$, in which case which the corresponding loss in accuracy is from 0.87 when $c = 20$, to 0.70 when $c = 30$.

7. Proofs. In this section, we provide proofs of the results presented earlier.

7.1. *Proof of Theorem 3.1.* For an infinitesimal generator $\mathscr{A}$ of a diffusion process on the state space Ω and a twice continuously differentiable function $h : \Omega \to \mathbb{R}$ with compact support, it is well known that

$$\mathbb{E}[\mathscr{A}h(\boldsymbol{X})] = 0,$$

where expectation is with respect to the stationary distribution of $\boldsymbol{X}$. Apply this result to the generator $\widetilde{\mathscr{L}}$ shown in (3.3) and monomial $G^{(m)}(\widetilde{\boldsymbol{x}}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ shown in (3.5). This provides a linear equation relating the expectations $\mathbb{E}[G^{(m+1)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}', \boldsymbol{b}', \boldsymbol{r}')]$, $\mathbb{E}[G^{(m)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}', \boldsymbol{b}', \boldsymbol{r}')]$, $\mathbb{E}[G^{(m-1)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}', \boldsymbol{b}', \boldsymbol{r}')]$, and $\mathbb{E}[G^{(m-2)}(\widetilde{\boldsymbol{X}}; \boldsymbol{a}', \boldsymbol{b}', \boldsymbol{r}')]$, each appearing with various different arguments $(\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{r}')$ depending on $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$; we omit the simple but algebraically lengthy details. Now, for these four choices of m we substitute the proposed expansion (3.6). If $m > 0$, then compare the coefficients of $\rho^{1-\frac{u}{2}}$ in the resulting expression; if $m = 0$, then compare the coefficients of $\rho^{-\frac{u}{2}}$. We then obtain the following:

(i) If $m = 0$ and $u = 0$, then the resulting expression is

$$[a(a - 1 + \theta_A) + b(b - 1 + \theta_B)]g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) =$$
$$\sum_i a_i(a_i - 1)g_0^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b}, \boldsymbol{0}) + \sum_j b_j(b_j - 1)g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{0})$$
$$+ \theta_A \sum_{i,k} a_i P_{ki}^A g_0^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0})$$
$$(7.1) \qquad + \theta_B \sum_{j,l} b_j P_{lj}^B g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{0}),$$

with boundary conditions given by (3.9). This is the sum of two copies of a familiar recursion (Griffiths and Tavaré, 1994) for the sampling distribution of a single locus, one for locus A and one for locus B. In our notation the solution is therefore
$$g_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) = q^A(\boldsymbol{a})q^B(\boldsymbol{b}).$$

(ii) If $m > 0$, then the resulting expression is

$$mg_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) =$$
$$\sum_{i,j} \Big\{ r_{ij}(r_{ij} - 1)g_u^{(m-2)}(\boldsymbol{a} + \boldsymbol{e}_i, \boldsymbol{b} + \boldsymbol{e}_j, \boldsymbol{r} - 2\boldsymbol{e}_{ij})$$
$$- \sum_l r_{ij}(r_{il} - \delta_{jl})g_u^{(m-2)}(\boldsymbol{a} + \boldsymbol{e}_i, \boldsymbol{b} + \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{r} - \boldsymbol{e}_{ij} - \boldsymbol{e}_{il})$$
$$- \sum_k r_{ij}(r_{kj} - \delta_{ik})g_u^{(m-2)}(\boldsymbol{a} + \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b} + \boldsymbol{e}_j, \boldsymbol{r} - \boldsymbol{e}_{ij} - \boldsymbol{e}_{kj})$$
$$+ \sum_{k,l} r_{ij}(r_{kl} - \delta_{ik}\delta_{jl})g_u^{(m-2)}(\boldsymbol{a} + \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b} + \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{r} - \boldsymbol{e}_{ij} - \boldsymbol{e}_{kl})$$
$$+ r_{ij}(r_{ij} - 1)g_{u-1}^{(m-1)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{ij})$$

$$- 2r_{ij}(r_{i\cdot} - 1)g_{u-1}^{(m-1)}(\boldsymbol{a}, \boldsymbol{b} + \boldsymbol{e}_j, \boldsymbol{r} - \boldsymbol{e}_{ij})$$

$$- 2r_{ij}(r_{\cdot j} - 1)g_{u-1}^{(m-1)}(\boldsymbol{a} + \boldsymbol{e}_i, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{ij})$$

$$+ 2\sum_{k,l} r_{kj}(r_{il} - \delta_{ik}\delta_{jl})g_{u-1}^{(m-1)}(\boldsymbol{a} + \boldsymbol{e}_k, \boldsymbol{b} + \boldsymbol{e}_l, \boldsymbol{r} - \boldsymbol{e}_{kj} - \boldsymbol{e}_{il} + \boldsymbol{e}_{ij})$$

$$+ 2(m-1)r_{ij}g_{u-1}^{(m-1)}(\boldsymbol{a} + \boldsymbol{e}_i, \boldsymbol{b} + \boldsymbol{e}_j, \boldsymbol{r} - \boldsymbol{e}_{ij}) \Big\}$$

$$+ \sum_i a_i(a_i + 2r_{i\cdot} - 1)g_{u-2}^{(m)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b}, \boldsymbol{r})$$

$$+ \sum_j b_j(b_j + 2r_{\cdot j} - 1)g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{r})$$

$$- 2\sum_{i,j}\sum_k a_i r_{kj}g_{u-2}^{(m)}(\boldsymbol{a} - \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{kj} + \boldsymbol{e}_{ij})$$

$$- 2\sum_{i,j}\sum_l b_j r_{il}g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{r} - \boldsymbol{e}_{il} + \boldsymbol{e}_{ij})$$

$$+ \theta_A \sum_{i,k} P_{ki}^A \Big[a_i g_{u-2}^{(m)}(\boldsymbol{a} - \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{r})$$

$$+ \sum_j r_{ij}g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{kj}) \Big]$$

$$+ \theta_B \sum_{j,l} P_{lj}^B \Big[b_j g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{r})$$

$$+ \sum_i r_{ij}g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{il}) \Big]$$

$$- [(a + m)(a + m + \theta_A - 1) + (b + m)(b + m + \theta_B - 1)$$

$$- m(m - 3)]g_{u-2}^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}),$$

$$(7.2) \quad + 2\sum_{i,j} a_i b_j g_{u-3}^{(m+1)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{r} + \boldsymbol{e}_{ij}).$$

Equation (7.2) relates $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ to the known expressions $g_u^{(m-2)}$, $g_{u-1}^{(m-1)}$, $g_{u-2}^{(m)}$, and $g_{u-3}^{(m+1)}$, as claimed.

(iii) If $m = 0$ and $u \geq 1$, then the resulting expression is

$$[a(a + \theta_A - 1) + b(b + \theta_A - 1)]g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) =$$

$$\sum_i a_i(a_i - 1)g_u^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b}, \boldsymbol{0}) + \sum_j b_j(b_j - 1)g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{0})$$

$$+ \theta_A \sum_{i,k} a_i P_{ki}^A g_u^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0})$$

$$+ \theta_B \sum_{j,l} b_j P_{lj}^B g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{0})$$

$$(7.3) \qquad + 2 \sum_{i,j} a_i b_j g_{u-1}^{(1)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{e}_{ij}),$$

with boundary conditions (3.9). Hence, this provides a recursion relation for $g_u^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ when $g_{u-1}^{(1)}$ is known.   □

7.2. *Proof of Corollary 3.1.*   For the base case $\ell = 1$, note that if $m = 1$ and $u = 0$ then (7.2) simplifies to

$$g_0^{(1)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = 0,$$

and hence if $m = 0$ and $u = 1$ then (7.3) simplifies to

$$[a(a + \theta_A - 1) + b(b + \theta_A - 1)]g_1^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) =$$
$$\sum_i a_i(a_i - 1)g_1^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i, \boldsymbol{b}, \boldsymbol{0}) + \sum_j b_j(b_j - 1)g_1^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{0})$$
$$+ \theta_A \sum_{i,k} a_i P_{ki}^A g_1^{(0)}(\boldsymbol{a} - \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0})$$
$$(7.4) \qquad + \theta_B \sum_{j,l} b_j P_{lj}^B g_1^{(0)}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{e}_j + \boldsymbol{e}_l, \boldsymbol{0}),$$

with boundary conditions (3.9). This has solution

$$g_1^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) = 0,$$

which is unique (since $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is unique). This completes $\ell = 1$. Now suppose inductively that $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = 0$ for every m, u such that $m + u = \ell - 2$ where ℓ is a fixed odd number greater than 1. Then for m, u such that $m + u = \ell$, (7.2) becomes

$$g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}) = 0,$$

as required.   □

7.3. *Proof of Lemma 4.1.*   In what follows, define the *length* of a sample configuration $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ to be $a + b + 2c$. Under a neutral, finite-alleles model, the probability of a sample with length δ satisfies a closed system of equations [e.g., see equation (5) of Jenkins and Song (2009)] which can be expressed in matrix form:

$$\boldsymbol{M}\boldsymbol{q} = \boldsymbol{v},$$

where $\boldsymbol{q}$ is a vector composed of the probabilities of samples of length less than or equal to δ, $\boldsymbol{v}$ is a constant vector of the same dimension as $\boldsymbol{q}$, and $\boldsymbol{M}$ is an invertible matrix (since the solution to this equation is unique). The entries of $\boldsymbol{M}$ and $\boldsymbol{v}$ are rational functions of ρ, and hence $\boldsymbol{q} = \boldsymbol{M}^{-1}\boldsymbol{v}$ is a vector each of whose entries is a rational function of ρ.

Let U_0 denote the degree of the numerator, and V_0 the degree of the denominator. If $U_0 > V_0$, then $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ becomes unbounded as $\rho \to \infty$, while if $V_0 > U_0$ then $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \to 0$ as $\rho \to \infty$. But we know that $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is a probability, and hence bounded. Moreover, it has support over all samples of a fixed size, since we assume that $\boldsymbol{P}^A$ and $\boldsymbol{P}^B$ are irreducible. Thus, to ensure $\lim_{\rho \to \infty} q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in (0, 1)$ we must have $U_0 = V_0$. By a similar argument as $\rho \to 0$, we must have that the coefficients of ρ^0 are nonzero both in the numerator and in the denominator. We can therefore divide the numerator and denominator by ρ^{U_0} to obtain a rational function of $1/\rho$ whose degree in the numerator and denominator are both U_0 ($= V_0$). $\square$

7.4. *Proof of Theorem 4.1.* This is an application of Theorem 1.4.4 of Baker and Graves-Morris (1996), which we spell out for completeness. By Lemma 4.1, $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is a rational function of $1/\rho$ and is analytic at $\rho = \infty$ with Taylor series (2.6). Denote the degree of its numerator and denominator by U_0. Then, $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ has $U_0 + U_0 + 1$ independent coefficients determined by the first $U_0 + U_0 + 1$ terms of its Taylor series expansion. Thus, provided $U \geq U_0$ and $V \geq U_0$, by the definition of the $[U/V]$ Padé approximant it must coincide uniquely with $q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. $\square$

7.5. *Proof of Theorem 5.1.* This is simply an application of Theorem 3.1 applied to the generator for the diffusion under selection, $\widetilde{\mathscr{L}}_{\mathrm{s}}$, rather than $\widetilde{\mathscr{L}}$, and so we just summarize the procedure.

The change in generator results in slight modifications to the relationships between the $g_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$, in order to obtain the relationships between the $h_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ for each m and u:

1. $h_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ satisfies (7.1) (replacing each $g_0^{(0)}$ with $h_0^{(0)}$), but with extra terms

$$+ \sum_{i,k} a_i \left[\sigma_{ik} h_0^{(0)}(\boldsymbol{a} + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0}) - \sum_{k'} \sigma_{kk'}^A h_0^{(0)}(\boldsymbol{a} + \boldsymbol{e}_k + \boldsymbol{e}_{k'}, \boldsymbol{b}, \boldsymbol{0}) \right],$$

on the right-hand side. The solution is

$$h_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0}) = q_{\mathrm{s}}^A(\boldsymbol{a}) q^B(\boldsymbol{b}).$$

2. For $m > 0$, $h_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ satisfies (7.2) but with extra terms

$$(7.5) \quad + \sum_{i,k} (a_i + r_{i\cdot}) \sigma_{ik}^A h_{u-2}^{(m)}(\boldsymbol{a} + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{r})$$

$$- (a + m) \sum_{k,k'} \sigma_{kk'}^A h_{u-2}^{(m)}(\boldsymbol{a} + \boldsymbol{e}_k + \boldsymbol{e}_{k'}, \boldsymbol{b}, \boldsymbol{r})$$

$$- \sum_{i,j,k,k'} r_{ij} \sigma_{kk'}^A h_{u-2}^{(m)}(\boldsymbol{a} + \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{r} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{k'j})$$

on the right-hand side.

3. For $m = 0$ and $u \geq 1$, $h_0^{(0)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{0})$ satisfies (7.2) but with extra terms

$$+ \sum_{i,k} \left[a_i \sigma_{ik}^A h_u^{(0)}(\boldsymbol{a} + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0}) - a \sigma_{ik}^A h_u^{(0)}(\boldsymbol{a} + \boldsymbol{e}_i + \boldsymbol{e}_k, \boldsymbol{b}, \boldsymbol{0}) \right]$$

$$+ \sum_{i,j,k} b_j \sigma_{ik}^A h_{u-1}^{(1)}(\boldsymbol{a} + \boldsymbol{e}_k, \boldsymbol{b} - \boldsymbol{e}_j, \boldsymbol{e}_{ij}),$$

on the right-hand side.

Using these equations to evaluate $h_u^{(m)}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r})$ on levels $\ell = 0, 1, \ldots, 4$ provides expressions for the nonneutral versions of $q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, and $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. Those for $q_0(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ and $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ in terms of the relevant one-locus sampling distributions are unchanged, while the new generator makes some minor modifications to the expression for $q_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$. The analytic part of this term, $\sigma_s(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, is easily calculated from $h_0^{(4)}$, $h_1^{(3)}$, $h_2^{(2)}$, and $h_3^{(1)}$, while the recursive part, $q_{2,s}(\boldsymbol{a} + \boldsymbol{c}_A, \boldsymbol{b} + \boldsymbol{c}_B, \boldsymbol{0})$ follows from $h_4^{(0)}$. $\square$

References.

Baker, G. A. and Graves-Morris, P. *Padé Approximants.* Cambridge University Press, 2nd edition, 1996.

Dingle, R. B. *Asymptotic expansions: Their derivation and interpretation.* Academic Press, New York, 1973.

Donnelly, P., Nordborg, M., and Joyce, P. (2001). Likelihoods and simulation methods for a class of nonneutral population genetics models. *Genetics*, **159**, 853–867.

Ethier, S. N. (1979). A limit theorem for two-locus diffusion models in population genetics. *Journal of Applied Probability*, **16**, 402–408.

Ethier, S. N. and Griffiths, R. C. (1990). On the two-locus sampling distribution. *Journal of Mathematical Biology*, **29**, 131–159.

Ethier, S. N. and Nagylaki, T. (1989). Diffusion approximations of the two-locus Wright-Fisher model. *Journal of Mathematical Biology*, **27**, 17–28.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Fearnhead, P. (2003). Haplotypes: the joint distribution of alleles at linked loci. *Journal of Applied Probability*, **40**, 505–512.

Golding, G. B. (1984). The sampling distribution of linkage disequilibrium. *Genetics*, **108**, 257–274.

Griffiths, R. C. and Tavaré, S. (1994). Simulating probability distributions in the coalescent. *Theoretical Population Biology*, **46**, 131–159.

Grote, M. and Speed, T. (2002). Approximate Ewens formulae for symmetric overdominance selection. *Annals of Applied Probability*, **12**, 637–663.

Handa, K. (2005). Sampling formulae for symmetric selection. *Electronic Communications in Probability*, **10**, 223–234.

Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics*, **159**, 1805–1817.

Huillet, T. (2007). Ewens sampling formulae with and without selection. *Journal of Computational and Applied Mathematics*, **206**, 755–773.

Jenkins, P. A. and Song, Y. S. (2009). Closed-form two-locus sampling distributions: accuracy and universality. *Genetics*, **183**, 1087–1103.

Jenkins, P. A. and Song, Y. S. (2010). An asymptotic sampling formula for the coalescent with recombination. *Annals of Applied Probability*, **20**, 1005–1028. (Technical Report 775, Department of Statistics, University of California, Berkeley, 2009).

McVean, G., Awadalla, P., and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231–1241.

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.

Ohta, T. and Kimura, M. (1969a). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutations. *Genetics*, **63**, 229–238.

Ohta, T. and Kimura, M. (1969b). Linkage disequilibrium due to random genetic drift. *Genetical Research*, **13**, 47–55.

Song, Y. S. and Song, J. S. (2007). Analytic computation of the expectation of the linkage disequilibrium coefficient r^2. *Theoretical Population Biology*, **71**, 49–60.

Wright, S. Adaptation and selection. In Jepson, G. L., Mayr, E., and Simpson, G. G., editors, *Genetics, Paleontology and Evolution*, pages 365–389. Princeton University Press, 1949.

P. A. Jenkins
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720
USA
E-mail: pauljenk@eecs.berkeley.edu

Y. S. Song
Department of Statistics and
    Computer Science Division
University of California, Berkeley
Berkeley, CA 94720
USA
E-mail: yss@stat.berkeley.edu