IMPORTANCE SAMPLING AND THE TWO-LOCUS MODEL WITH SUBDIVIDED POPULATION STRUCTURE

ROBERT C. GRIFFITHS * AND

PAUL A. JENKINS,* *University of Oxford*

YUN S. SONG,** *University of California, Berkeley*

Abstract

The diffusion-generator approximation technique developed by De Iorio and Griffiths (2004a) is a very useful method of constructing importance-sampling proposal distributions. Being based on general mathematical principles, the method can be applied to various models in population genetics. In this paper we extend the technique to the neutral coalescent model with recombination, thus obtaining novel sampling distributions for the two-locus model. We consider the case with subdivided population structure, as well as the classic case with only a single population. In the latter case we also consider the importance-sampling proposal distributions suggested by Fearnhead and Donnelly (2001), and show that their two-locus distributions generally differ from ours. In the case of the infinitely-many-alleles model, our approximate sampling distributions are shown to be generally closer to the true distributions than are Fearnhead and Donnelly's.

Keywords: Coalescent process; recombination; diffusion process; importance sampling; migration; subdivided population

2000 Mathematics Subject Classification: Primary 60G40
Secondary 93E25; 92D15

1. Introduction

Estimating model parameters and making ancestral inference are an important part of molecular population genetics. At the core of these studies is the problem of computing the likelihood of the type configuration of sample sequences. In the context of the coalescent model and its various extensions, closed-form formulae are generally not known for such likelihoods and, therefore, several computationally intensive statistical methods have been proposed for approximating them. Most of these statistical approaches fall into one of two categories: one based on Markov chain Monte Carlo methods—for examples, see Kuhner *et al.* (1995), (2000), Wilson and Balding (1998), and Beaumont (1999)—and the other based on importance-sampling (IS) methods, some notable examples being Griffiths and Tavaré (1994a), (1994b), (1994c), Griffiths and Marjoram (1996), Stephens and Donnelly (2000), and Fearnhead and Donnelly (2001).

On the importance-sampling side, new impetus was given when Stephens and Donnelly (2000) constructed a very efficient IS scheme for the neutral coalescent model for a single population. Recently, De Iorio and Griffiths (2004a) developed a general method of constructing

IS proposal distributions from a diffusion-process generator and showed that their proposal distributions coincide with that of Stephens and Donnelly (2000) in the case of the neutral coalescent model for a single population. A particularly appealing property of the technique developed by De Iorio and Griffiths is that the construction of the proposal distributions in that approach is based on general mathematical principles. The technique is systematic and can be applied to various settings. For instance, De Iorio and Griffiths (2004b) applied their technique to the neutral coalescent model with subdivided population structure, obtaining significant improvement over previous IS schemes.

The goal of the present paper is to extend the method of De Iorio and Griffiths (2004a), (2004b) to the neutral coalescent model with recombination. We consider the case with subdivided population structure, as well as that with only a single population. We focus on the two-locus model in this paper and defer addressing the general case to a later paper; for now, we just mention that much of what we discuss here can be carried over to multilocus models as well. Throughout this paper, two specific models—namely, diallelic models and parent-independent mutation (PIM) models—are examined in detail, thus illustrating how our method works. For these models, we obtain explicit formulae for conditional sampling distributions, which can be used to devise an IS scheme.

Of all hitherto suggested IS schemes for the coalescent model with recombination in the case of a single population, that proposed by Fearnhead and Donnelly (2001) seems most efficient. In a recent study of the fine-scale variation of recombination rates in the human genome (see McVean *et al.* (2004), Myers *et al.* (2005), and Fearnhead and Smith (2005)), Fearnhead and Donnelly's IS scheme was employed to compute two-locus likelihoods, which were then combined using Hudson's (2001) composite likelihood idea. In this paper we construct novel conditional sampling distributions and compare them with that used in Fearnhead and Donnelly's IS scheme. In the case of the two-locus model we show that our sampling distributions are generally different from that of Fearnhead and Donnelly's. Furthermore, for the infinitely-many-alleles model, in which case we can numerically compute the true sampling distributions for a small sample size, we show that our sampling distributions are generally closer to the true distributions than are Fearnhead and Donnelly's. Note that IS for the neutral coalescent model with both recombination and subdivided population structure has not been studied before; our sampling distribution for that case is therefore the first of its kind.

The organization of this paper is as follows. In Section 2 we review the one-locus case studied by De Iorio and Griffiths (2004a) and describe their general diffusion-generator approximation technique. Our two-locus sampling distributions for a single population are discussed in Section 3, whereas Fearnhead and Donnelly's (2001) corresponding distributions are examined in Section 4. In Section 5 the aforementioned comparison of the approximate sampling distributions with the true distributions is carried out in the case of the infinitely-many-alleles model. The two-locus model with subdivided population structure is discussed in Section 6. In Section 7 we conclude with a brief discussion on future directions.

2. A brief review of the one-locus case in a single population

We first consider the one-locus case in a single population. In addition to serving as a simple example that clearly illustrates the general idea behind our approach, the one-locus case will resurface in an important way when we discuss the two-locus case.

2.1. Diffusion approximation

We denote the type space for alleles by $E = \{1, 2, \ldots, d\}$ and the population allele frequencies by $X = (X_i)_{i \in E}$. The generator for the diffusion process of allele frequencies in the space $\Delta := \{x = (x_i)_{i \in E} \mid x_i \geq 0 \text{ for all } i \in E \text{ and } \sum_{i \in E} x_i = 1\}$ is given by

$$\mathcal{L} = \frac{1}{2} \sum_{i,j \in E} x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i \in E} \sum_{j \in E} (-x_i \alpha_{ij} + x_j \alpha_{ji}) \frac{\partial}{\partial x_i},$$

where α_{ij} are mutation parameters. If mutation events occur according to a Poisson process with rate $\theta/2$ and type changes are governed by a Markov chain with transition matrix $P = (P_{ij})$, then $\alpha_{ij} = (\theta/2) P_{ij}$ and $\sum_{j \in E} \alpha_{ij} = \theta/2$, in which case the diffusion process generator becomes

$$\mathcal{L} = \frac{1}{2} \sum_{i,j \in E} x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\theta}{2} \sum_{i \in E} \sum_{j \in E} x_j (P_{ji} - \delta_{ji}) \frac{\partial}{\partial x_i}.$$

Note that this generator can be written as

$$\mathcal{L} = \sum_{i \in E} L_i \frac{\partial}{\partial x_i}, \quad \text{where } L_i = \frac{1}{2} \sum_{j \in E} x_i (\delta_{ij} - x_j) \frac{\partial}{\partial x_j} + \frac{\theta}{2} \sum_{j \in E} x_j (P_{ji} - \delta_{ji}).$$

With E denoting the expectation with respect to the stationary distribution of the diffusion process, the sampling distribution of an unordered type configuration $n = (n_1, n_2, \ldots, n_d)$ is given by $p(n) = \binom{n}{n} E(\prod_{k \in E} X_k^{n_k})$, where $n = \sum_{k=1}^{d} n_k$ and $\binom{n}{n}$ denotes the multinomial coefficient $\binom{n}{n_1, \ldots, n_d}$. An important fact is that this sampling distribution satisfies the exchangeability condition,

$$\pi(i \mid n - e_j) p(n - e_j) = \frac{n_i + 1 - \delta_{ij}}{n} p(n - e_j + e_i), \tag{1}$$

where e_j denotes the unit vector with a 1 in the jth component and $\pi(i \mid n)$ denotes the conditional sampling probability of an additionally sampled allele being of type i, given that the current unordered sample configuration is n. Conditional probabilities are normalized so that $\sum_{i \in E} \pi(i \mid n) = 1$. Another important point to note is that

$$E\left(\mathcal{L} \prod_{k \in E} X_k^{n_k}\right) = E\left(\sum_{i \in E} L_i \frac{\partial}{\partial x_i} \prod_{k \in E} X_k^{n_k}\right) = 0, \tag{2}$$

which follows from the fact that $E(\mathcal{L} f(X)) = 0$ for any bounded continuous function f with well-defined second derivatives. The key to the technique developed by De Iorio and Griffiths (2004a) is to assume that there exists a distribution with expectation operator $\hat{E}$ such that the vanishing of (2) holds componentwise; that is, for all $i \in E$,

$$\hat{E}\left(L_i \frac{\partial}{\partial x_i} \prod_{k \in E} X_k^{n_k}\right) = 0. \tag{3}$$

Furthermore, assuming that the exchangeability condition shown in (1) holds for the sampling probabilities $\hat{p}(n) = \binom{n}{n} \hat{E}(\prod_{k \in E} X_k^{n_k})$ and the corresponding $\hat{\pi}(i \mid n)$, a system of equations satisfied by $\hat{\pi}(i \mid n)$ can be found (see De Iorio and Griffiths (2004a)). A key observation is that these conditional sampling probabilities can be used to construct efficient IS proposal distributions.

TABLE 1: Proposal distribution and importance weights for a neutral coalescent model of a single locus.

H_{k-1}	$\hat{p}(H_{k-1} \mid H_k)$	IS weight
$\boldsymbol{n} - \boldsymbol{e}_j$	$\dfrac{n_j - 1}{n + \theta - 1} \dfrac{n_j}{n} \dfrac{1}{\hat{\pi}(j \mid \boldsymbol{n} - \boldsymbol{e}_j)}$	$\dfrac{n}{n_j} \hat{\pi}(j \mid \boldsymbol{n} - \boldsymbol{e}_j)$
$\boldsymbol{n} + \boldsymbol{e}_i - \boldsymbol{e}_j$	$\dfrac{\theta P_{ij}}{n + \theta - 1} \dfrac{n_j}{n} \dfrac{\hat{\pi}(i \mid \boldsymbol{n} - \boldsymbol{e}_j)}{\hat{\pi}(j \mid \boldsymbol{n} - \boldsymbol{e}_j)}$	$\dfrac{n_i + 1 - \delta_{ij}}{n_j} \dfrac{\hat{\pi}(j \mid \boldsymbol{n} - \boldsymbol{e}_j)}{\hat{\pi}(i \mid \boldsymbol{n} - \boldsymbol{e}_j)}$

2.2. Using $\hat{\pi}(i \mid n)$ for importance sampling

The likelihood of a configuration can be calculated by sequential IS on coalescent histories which begin at a configuration H_0 of sample genes, and move through states $H_{-1}, \ldots, H_{-m}$ back in time. Changes of state occur when a coalescence, mutation, migration (in the case of a subdivided population structure), or recombination (in the case of more than one locus) takes place. The most recent common ancestor (MRCA) is reached at H_{-m}.

The forward transition probabilities $p(H_k \mid H_{k-1})$ from the MRCA to the sample are known from the coalescent process, whilst the reverse transition probabilities $p(H_{k-1} \mid H_k)$ are unknown and replaced by an IS proposal $\hat{p}(H_{k-1} \mid H_k)$. The IS weight in a transition from H_k to H_{k-1} is then $p(H_k \mid H_{k-1})/\hat{p}(H_{k-1} \mid H_k)$. Using Bayes' rule and (1) applied to $\hat{p}(\boldsymbol{n})$ and $\hat{\pi}(i \mid \boldsymbol{n})$, the proposal distribution and IS weights can be expressed in terms of $\hat{\pi}$. The simplest one-locus case is illustrated in Table 1. We briefly mention that approaches other than IS can be made to exploit the sampling distribution $\hat{\pi}(i \mid \boldsymbol{n})$. For example, Li and Stephens (2003) constructed an efficient way to estimate the likelihood by introducing the product of approximate conditionals (PAC), defined as $\hat{p}(\boldsymbol{n}) = \hat{\pi}(i_1)\hat{\pi}(i_2 \mid i_1) \cdots \hat{\pi}(i_n \mid i_1, \ldots, i_{n-1})$, where $(i_1, \ldots, i_n)$ denotes a random permutation of the ordered configuration $(a_1, \ldots, a_n)$ of genes corresponding to the unordered configuration $\boldsymbol{n}$. For the stepwise mutation model, Cornuet and Beaumont (2007) provided a comparison of a PAC scheme with an IS approach, when $\hat{\pi}(i \mid \boldsymbol{n})$ is derived using the method of the preceding subsection.

2.3. A general solution to the sampling distribution $\hat{\pi}(i \mid n)$

The system of equations for $\hat{\pi}(i \mid \boldsymbol{n})$ that we can obtain using (1) and (3) is

$$(n + \theta)\hat{\pi}(i \mid \boldsymbol{n}) = n_i + \theta \sum_{k \in E} \hat{\pi}(k \mid \boldsymbol{n}) P_{ki}. \tag{4}$$

(See De Iorio and Griffiths (2004a) for the details of the computation.) This system of equations can easily be solved using matrix inversion. More precisely, $\hat{\pi}(i \mid \boldsymbol{n})$ is the ith component of the row vector $c\boldsymbol{n}(\boldsymbol{I} - c\theta\boldsymbol{P})^{-1}$, where $c = 1/(n + \theta)$ and $\boldsymbol{I}$ is the $d \times d$ identity matrix. As discussed in De Iorio and Griffiths (2004a), this solution is exactly what Stephens and Donnelly (2000) also obtained using a different approach. Furthermore, note that, as $n \to \infty$ and $n_i/n \to x_i$, with x_i being the population frequency of the allele type i, $\hat{\pi}(i \mid \boldsymbol{n}) \to x_i$.

The sampling distributions $\hat{\pi}(i \mid \boldsymbol{n})$ depend on θ and $\boldsymbol{P}$ only through the rate matrix $\theta(\boldsymbol{P} - \boldsymbol{I})$. The proposal distribution and IS weights in Table 1 can be modified to depend only on the rate matrix by considering the proposal distribution conditional on state changes by mutation $j \to i$, where $i \neq j$.

2.4. $\hat{\pi}(i \mid n)$ for a diallelic locus

When there are only two alleles at each locus, i.e. $E = \{1, 2\}$, the expression $c\, \boldsymbol{n}(\boldsymbol{I} - c\,\theta\,\boldsymbol{P})^{-1}$ described in the previous subsection takes on a simple form. More explicitly, the conditional sampling distributions are given by

$$\hat{\pi}(1 \mid \boldsymbol{n}) = \frac{n_1 + \theta P_{21}}{n + \theta(P_{12} + P_{21})} \quad \text{and} \quad \hat{\pi}(2 \mid \boldsymbol{n}) = \frac{n_2 + \theta P_{12}}{n + \theta(P_{12} + P_{21})}.$$

2.5. $\hat{\pi}(i \mid n)$ for PIM models

For PIM models, the transition probability satisfies $P_{ki} = P_i$ and, therefore, it follows from $\sum_{k \in E} \hat{\pi}(k \mid \boldsymbol{n}) = 1$ that $\theta \sum_{k \in E} \hat{\pi}(k \mid \boldsymbol{n}) P_{ki} = \theta P_i$. Hence, the sampling distribution $\hat{\pi}(i \mid \boldsymbol{n})$ for $i \in E$ has the following simple form for PIM models:

$$\hat{\pi}(i \mid \boldsymbol{n}) = \frac{n_i + \theta P_i}{n + \theta}. \tag{5}$$

The two-allele and PIM models are seen to be equivalent by choosing θ and $\boldsymbol{P}$ such that

$$\theta_{\text{PIM}} = \theta_{\text{two-allele}}(P_{12} + P_{21}),$$

$$P_1 = \frac{P_{21}}{P_{12} + P_{21}},$$

$$P_2 = \frac{P_{12}}{P_{12} + P_{21}}.$$

3. The two-locus model in a single population

In this section we apply the diffusion approximation technique to the two-locus model in a single population. As we elaborate presently, there exists an intricate link between the one-locus and the two-locus models. This property allows us to obtain closed-form formulae for sampling distributions for certain models.

3.1. Notation

We first define some useful notation to be used in the remaining part of this paper.

1. The first locus is denoted by A and the second locus by B.

2. Let E_A and E_B denote the allele type spaces for the first and the second loci, respectively.

3. We use '·' to denote that an index has been summed over. For example, $n_{i\cdot} = \sum_{j \in E_B} n_{ij}$ and $n_{\cdot j} = \sum_{i \in E_A} n_{ij}$.

4. Given a rank-2 tensor $\boldsymbol{n} = (n_{ij})_{(i,j) \in E_A \times E_B}$, we define two vectors $\boldsymbol{n}_A = (n_{i\cdot})_{i \in E_A}$ and $\boldsymbol{n}_B = (n_{\cdot j})_{j \in E_B}$ by summing over one of the indices.

5. A scalar n is defined as $n = \sum_{(i,j) \in E_A \times E_B} n_{ij}$.

6. We use $\boldsymbol{e}_i$ to denote the unit vector whose ith component is 1 while all other components are 0. Similarly, $\boldsymbol{e}_{ij}$ denotes a rank-2 tensor whose (i, j) component is 1 while all other components are 0.

3.2. The neutral two-locus diffusion model

The state space of the neutral two-locus diffusion model is

$$\Delta = \left\{ \boldsymbol{x} = (x_{ij})_{(i,j) \in E_A \times E_B} \;\middle|\; x_{ij} \geq 0 \text{ for all } (i, j) \in E_A \times E_B \text{ and } \sum_{(i,j) \in E_A \times E_B} x_{ij} = 1 \right\},$$

and the generator for the diffusion process is

$$\mathcal{L} = \sum_{(i,j) \in E_A \times E_B} L_{ij} \frac{\partial}{\partial x_{ij}},$$

where

$$L_{ij} = \frac{1}{2} \sum_{(k,l) \in E_A \times E_B} x_{ij} (\delta_{ik} \delta_{jl} - x_{kl}) \frac{\partial}{\partial x_{kl}} + b_{ij}(\boldsymbol{x}) + \frac{1}{2} \rho (x_{i.} x_{.j} - x_{ij}).$$

In the infinitesimal mean part of the generator, ρ is the population-scaled recombination rate and $b_{ij}(\boldsymbol{x})$ is

$$b_{ij}(\boldsymbol{x}) = \frac{\theta_A}{2} \sum_{k \in E_A} x_{kj} (P_{ki}^A - \delta_{ki}) + \frac{\theta_B}{2} \sum_{l \in E_B} x_{il} (P_{lj}^B - \delta_{lj}),$$

where θ_α and P_{ij}^α are the population-scaled mutation rate and entries of the transition matrix for locus $\alpha \in \{A, B\}$. For clarity of discussion, we keep the parameters of the two loci distinguished.

Consider the multinomial probability

$$Q(\boldsymbol{x}, \boldsymbol{n}) = \left(\frac{n!}{\prod_{(i,j) \in E_A \times E_B} n_{ij}!} \right) \prod_{(k,l) \in E_A \times E_B} (x_{kl})^{n_{kl}}. \tag{6}$$

The sampling distribution $p(\boldsymbol{n})$ of the unordered sample configuration $\boldsymbol{n}$ is defined as $p(\boldsymbol{n}) = \mathrm{E}(Q(\boldsymbol{X}, \boldsymbol{n}))$, where the expectation operator E is defined with respect to the stationary distribution of the diffusion process. Let $\pi((i, j) \mid \boldsymbol{n})$ denote the conditional probability that the $(n+1)$th sampled allele is of type $(i, j) \in E_A \times E_B$, given that the first n alleles have multiplicity configuration $\boldsymbol{n}$. These probabilities are normalized so that $\sum_{(i,j) \in E_A \times E_B} \pi((i, j) \mid \boldsymbol{n}) = 1$. The sampling distributions π and p satisfy the exchangeability condition

$$\pi((i, j) \mid \boldsymbol{n}) p(\boldsymbol{n}) = \mathrm{E}(X_{ij} Q(\boldsymbol{X}, \boldsymbol{n})) = \frac{n_{ij} + 1}{n + 1} p(\boldsymbol{n} + \boldsymbol{e}_{ij}), \tag{7}$$

which implies the following collection of conditions:

$$\frac{p(\boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{kl})}{p(\boldsymbol{n} - \boldsymbol{e}_{ij})} = \frac{n}{n_{kl} + 1 - \delta_{ik} \delta_{jl}} \pi((k, l) \mid \boldsymbol{n} - \boldsymbol{e}_{ij}), \tag{8}$$

$$\frac{p(\boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{il} + \boldsymbol{e}_{kj})}{p(\boldsymbol{n} - \boldsymbol{e}_{ij})} = \frac{n_{ij}}{n_{ij} + \delta_{ik} \delta_{jl}} \frac{n(n+1)}{(n_{kj} + 1 - \delta_{ik})(n_{il} + 1 - \delta_{jl})}$$
$$\times \pi(\{(i, l), (k, j)\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}), \tag{9}$$

where $\pi(\{(i, l), (k, j)\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}) = \mathrm{E}(X_{il} X_{kj} Q(\boldsymbol{X}, \boldsymbol{n} - \boldsymbol{e}_{ij}))/p(\boldsymbol{n} - \boldsymbol{e}_{ij})$. The sampling distribution π satisfies the symmetry identity

$$\pi((k, j) \mid \boldsymbol{n} + \boldsymbol{e}_{il}) \, \pi((i, l) \mid \boldsymbol{n}) = \pi((i, l) \mid \boldsymbol{n} + \boldsymbol{e}_{kj}) \, \pi((k, j) \mid \boldsymbol{n}), \tag{10}$$

which implies that

$$\pi(\{(i, l), (k, j)\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}) = \pi((k, j) \mid \boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{il})\pi((i, l) \mid \boldsymbol{n} - \boldsymbol{e}_{ij})$$
$$= \pi((i, l) \mid \boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{kj})\pi((k, j) \mid \boldsymbol{n} - \boldsymbol{e}_{ij}).$$

3.3. The main recursion for $\hat{\pi}((i, j) \mid \boldsymbol{n})$

Adopting the idea of De Iorio and Griffiths (2004a), we now assume that there exists a distribution with expectation operator $\hat{E}$ such that, for all $(i, j) \in E_A \times E_B$,

$$\hat{E}\left(L_{ij} \frac{\partial}{\partial x_{ij}} \prod_{(k,l) \in E_A \times E_B} X_{kl}^{n_{kl}} \right) = 0. \tag{11}$$

Proposition 1. *Let $\hat{p}(\boldsymbol{n}) = \hat{E}(Q(X, \boldsymbol{n}))$, where $Q(X, \boldsymbol{n})$ is the multinomial expression shown in (6), be an approximate sampling distribution, and let $\hat{\pi}$ be the corresponding approximate conditional sampling distribution that satisfies the exchangeability conditions (7)–(9). Then,*

$$(n + \rho + \theta_A + \theta_B)\hat{\pi}((i, j) \mid \boldsymbol{n}) = n_{ij} + \theta_A \sum_{k \in E_A} \hat{\pi}((k, j) \mid \boldsymbol{n}) P_{ki}^A + \theta_B \sum_{l \in E_B} \hat{\pi}((i, l) \mid \boldsymbol{n}) P_{lj}^B$$
$$+ \rho\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}), \tag{12}$$

where $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) = \hat{E}(X_{i.} X_{.j} Q(X, \boldsymbol{n}))/\hat{p}(\boldsymbol{n})$.

Proof. It is straightforward to show that the componentwise vanishing property (11) implies the following relation for $\hat{p}$:

$$n_{ij}((n - 1) + \rho + \theta_A + \theta_B)\hat{p}(\boldsymbol{n})$$
$$= n(n_{ij} - 1)\hat{p}(\boldsymbol{n} - \boldsymbol{e}_{ij})$$
$$+ \theta_A \sum_{k \in E_A} P_{ki}^A (n_{kj} + 1 - \delta_{ik})\hat{p}(\boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{kj})$$
$$+ \theta_B \sum_{l \in E_B} P_{lj}^B (n_{il} + 1 - \delta_{jl})\hat{p}(\boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{il})$$
$$+ \frac{\rho}{n + 1} \sum_{(k,l) \in E_A \times E_B} \left(\hat{p}(\boldsymbol{n} - \boldsymbol{e}_{ij} + \boldsymbol{e}_{il} + \boldsymbol{e}_{kj}) \right.$$
$$\left. \times \frac{n_{ij} + \delta_{ik}\delta_{jl}}{n_{ij}} (n_{il} + 1 - \delta_{jl})(n_{kj} + 1 - \delta_{ik}) \right). \tag{13}$$

This can also be obtained by assuming in the ancestral recombination graph that the next event (coalescence, mutation, or recombination) back in time has probability n_{ij}/n of occurring to a gene of type (i, j). Using the exchangeability conditions shown in (7)–(9), the above equation for $\hat{p}$ can be written in terms of $\hat{\pi}$, as shown in (12), after setting $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{e}_{ij}$.

Note that the approximate conditional distribution $\hat{\pi}$ may not satisfy the symmetry identity (10) satisfied by the true distribution π. Therefore, the formula we obtain for $\hat{\pi}((i, j) \mid \boldsymbol{n})$ may depend on how $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n})$ is treated. Motivated by the symmetry identity (10), we

use the following symmetric form when evaluating $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n})$:

$$\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n})$$
$$= \frac{1}{2} \sum_{k \in E_A} \sum_{l \in E_B} (\hat{\pi}((i, l) \mid \boldsymbol{n}) \, \hat{\pi}((k, j) \mid \boldsymbol{n} + \boldsymbol{e}_{il}) + \hat{\pi}((i, l) \mid \boldsymbol{n} + \boldsymbol{e}_{kj}) \, \hat{\pi}((k, j) \mid \boldsymbol{n})).$$

(14)

The main difficulty in solving (12) comes from this part. At first sight, it is not obvious how these terms quadratic in $\hat{\pi}((a, b) \mid \boldsymbol{n})$ should be handled. It turns out that the system of equations shown in (12) possesses hidden structures that prove very useful. We turn to this property next.

3.4. Marginal distributions

The key observation that allows us to solve for $\hat{\pi}((i, j) \mid \boldsymbol{n})$ in (12) is as follows. If we sum over the index j in (12) then the terms that contain the mutation parameter θ_B cancel out nicely between the left- and right-hand sides. The same holds true for the terms that contain ρ. Hence, what remains is the following simple system of equations:

$$(n + \theta_A)\hat{\pi}((i, \cdot) \mid \boldsymbol{n}) = n_{i\cdot} + \theta_A \sum_{k \in E_A} \hat{\pi}((k, \cdot) \mid \boldsymbol{n}) \, P_{ki}^A.$$

(15)

Note that this system of equations is a marginal system depending only on $\boldsymbol{n}_A$. In fact, it is precisely what the one-locus conditional distribution for the first locus satisfies; summing over the index for the second locus has reduced the two-locus equations (12) to the one-locus equations (4) for the first locus. There exists a unique solution to (15), namely $\hat{\pi}((i, \cdot) \mid \boldsymbol{n}) = \hat{\pi}_A(i \mid \boldsymbol{n}_A)$, the latter being the one-locus distribution for locus A.

In a similar vein, if we sum over the index i in (12) then we obtain

$$(n + \theta_B)\hat{\pi}((\cdot, j) \mid \boldsymbol{n}) = n_{\cdot j} + \theta_B \sum_{l \in E_B} \hat{\pi}((\cdot, l) \mid \boldsymbol{n}) P_{lj}^B,$$

(16)

which is a marginal system depending only on $\boldsymbol{n}_B$. The unique solution to (16) is $\hat{\pi}((\cdot, j) \mid \boldsymbol{n}) = \hat{\pi}_B(j \mid \boldsymbol{n}_B)$, the latter being the one-locus distribution for locus B.

3.5. $\hat{\pi}((i, j) \mid n)$ for diallelic loci

In the case of diallelic loci, i.e. $E_A = E_B = \{1, 2\}$, it becomes particularly clear how the observation made in the previous subsection can be utilized to obtain $\hat{\pi}((i, j) \mid \boldsymbol{n})$. In what follows, we define $\bar{i} = 1$ if $i = 2$ and $\bar{i} = 2$ if $i = 1$.

Proposition 2. *Assume the symmetric form (14) for $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n})$. Then, for diallelic loci, a solution to (12) is given by*

$$\hat{\pi}((i, j) \mid \boldsymbol{n})$$
$$= \frac{1}{\mathcal{N}} \bigg(n_{ij} + \theta_A P_{\bar{i}i}^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_{\bar{j}j}^B \hat{\pi}_A(i \mid \boldsymbol{n}_A)$$
$$+ \frac{\rho}{2} \bigg(\frac{n + \theta_A(P_{12}^A + P_{21}^A)}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)} + \frac{n + \theta_B(P_{12}^B + P_{21}^B)}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)} \bigg) \hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B) \bigg),$$

(17)

where the normalization constant $\mathcal{N}$ is defined as

$$\mathcal{N} = n + \theta_A(P_{12}^A + P_{21}^A) + \theta_B(P_{12}^B + P_{21}^B)$$
$$+ \frac{\rho}{2}\left(\frac{n + \theta_A(P_{12}^A + P_{21}^A)}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)} + \frac{n + \theta_B(P_{12}^B + P_{21}^B)}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)}\right).$$

Proof. As in the one-locus case (cf. Section 2.4), (15) and (16) for two-locus marginal distributions admit simple solutions when there are only two possible alleles at each locus. More precisely, we have

$$\hat{\pi}((i,\cdot) \mid \boldsymbol{n}) = \hat{\pi}_A(i \mid \boldsymbol{n}_A) = \frac{n_{i\cdot} + \theta_A P_{\bar{i}i}^A}{n + \theta_A(P_{12}^A + P_{21}^A)}, \tag{18}$$

$$\hat{\pi}((\cdot,j) \mid \boldsymbol{n}) = \hat{\pi}_B(j \mid \boldsymbol{n}_B) = \frac{n_{\cdot j} + \theta_B P_{\bar{j}j}^B}{n + \theta_B(P_{12}^B + P_{21}^B)}. \tag{19}$$

Hence, it follows that

$$\hat{\pi}((i,\cdot) \mid \boldsymbol{n} + \boldsymbol{e}_{kj}) = \frac{n_{i\cdot} + \delta_{ik} + \theta_A P_{\bar{i}i}^A}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)} = \frac{(n + \theta_A(P_{12}^A + P_{21}^A))\hat{\pi}_A(i \mid \boldsymbol{n}_A) + \delta_{ik}}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)},$$

$$\hat{\pi}((\cdot,j) \mid \boldsymbol{n} + \boldsymbol{e}_{il}) = \frac{n_{\cdot j} + \delta_{jl} + \theta_B P_{\bar{j}j}^B}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)} = \frac{(n + \theta_B(P_{12}^B + P_{21}^B))\hat{\pi}_B(j \mid \boldsymbol{n}_B) + \delta_{jl}}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)},$$

and, therefore, $\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \boldsymbol{n})$ can be written as

$$\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \boldsymbol{n})$$
$$= \frac{1}{2}\left(\left(\frac{1}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)} + \frac{1}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)}\right)\hat{\pi}((i,j) \mid \boldsymbol{n})\right.$$
$$\left. + \left(\frac{n + \theta_A(P_{12}^A + P_{21}^A)}{n + 1 + \theta_A(P_{12}^A + P_{21}^A)} + \frac{n + \theta_B(P_{12}^B + P_{21}^B)}{n + 1 + \theta_B(P_{12}^B + P_{21}^B)}\right)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B)\right),$$

where we have used the symmetric form (14). Using this, along with (18) and (19), we can easily solve for $\hat{\pi}((i,j) \mid \boldsymbol{n})$ in (12). After some simple algebra, we obtain (17). $\quad\square$

Note that everything on the right-hand side of (17) is completely known and that the solution has a very nice form. In particular, the contribution of recombination to $\hat{\pi}((i,j) \mid \boldsymbol{n})$ is proportional to a simple product of independent one-locus distributions at the two loci. Similar to the case of a single locus, as $n \to \infty$ and $n_{ij}/n \to x_{ij}$, with x_{ij} being the population frequency of the allele type (i,j), $\hat{\pi}((i,j) \mid \boldsymbol{n}) \to x_{ij}$.

Note that the approximate conditional sampling distribution $\hat{\pi}((i,j) \mid \boldsymbol{n})$ obtained above satisfies $\sum_{(i,j)\in E_A \times E_B} \hat{\pi}((i,j) \mid \boldsymbol{n}) = 1$. Also, for $\rho = \infty$, it satisfies the symmetry condition (10). For $0 < \rho < \infty$, the symmetry condition is not satisfied in general.

3.6. $\hat{\pi}((i,j) \mid \boldsymbol{n})$ for PIM models

We first need to obtain a system of equations relating $\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \boldsymbol{n})$ to $\hat{\pi}((i,j) \mid \boldsymbol{n})$. In (15) set $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{e}_{lj}$, multiply by $\hat{p}(\boldsymbol{n} + \boldsymbol{e}_{lj})(\prod_{(r,s)\in E\times E}(n_{rs} + \delta_{rl}\delta_{sj})!)/(n+1)!$, and then sum over the index l. The resulting equations are

$$(n + \theta_A + 1)\hat{E}(X_{i\cdot}X_{\cdot j}\boldsymbol{X}^{\boldsymbol{n}}) = n_{i\cdot}\hat{E}(X_{\cdot j}\boldsymbol{X}^{\boldsymbol{n}}) + \hat{E}(X_{ij}\boldsymbol{X}^{\boldsymbol{n}}) + \theta_A\sum_{k\in E_A}\hat{E}(X_{k\cdot}X_{\cdot j}\boldsymbol{X}^{\boldsymbol{n}})P_{ki}^A,$$

where X^n denotes $\prod_{(a,b) \in E_A \times E_B} X_{ab}^{n_{ab}}$. Then, multiplying by $n!/(\hat{p}(\boldsymbol{n}) \prod_{(r,s) \in E_A \times E_B} n_{rs}!)$ gives the recursion

$$(n + \theta_A + 1)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) = n_{i\cdot}\hat{\pi}_B(j \mid \boldsymbol{n}_B) + \hat{\pi}((i, j) \mid \boldsymbol{n})$$
$$+ \theta_A \sum_{k \in E_A} \hat{\pi}(\{(k, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) P_{ki}^A. \qquad (20)$$

In the PIM model $P_{ki}^A = P_i^A$ for all $k, i \in E_A$. Hence, (15) reduces to

$$(n + \theta_A)\hat{\pi}_A(i \mid \boldsymbol{n}_A) = n_{i\cdot} + \theta_A P_i^A, \qquad (21)$$

and (20) reduces to

$$(n + \theta_A + 1)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) = (n_{i\cdot} + \theta_A P_i^A)\hat{\pi}_B(j \mid \boldsymbol{n}_B) + \hat{\pi}((i, j) \mid \boldsymbol{n})$$
$$= (n + \theta_A)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B) + \hat{\pi}((i, j) \mid \boldsymbol{n}).$$

In a similar vein we can show that

$$(n + \theta_B)\hat{\pi}_B(j \mid \boldsymbol{n}_B) = n_{\cdot j} + \theta_B P_i^B \qquad (22)$$

and

$$(n + \theta_B + 1)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) = (n + \theta_B)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B) + \hat{\pi}((i, j) \mid \boldsymbol{n}).$$

Note that (21) and (22) are analogues of the one-locus distribution (5) for loci A and B, respectively. Symmetrizing with respect to the two loci thus gives

$$\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \boldsymbol{n}) = \frac{1}{2}\left(\frac{n + \theta_A}{n + 1 + \theta_A} + \frac{n + \theta_B}{n + 1 + \theta_B}\right)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B)$$
$$+ \frac{1}{2}\left(\frac{1}{n + 1 + \theta_A} + \frac{1}{n + 1 + \theta_B}\right)\hat{\pi}((i, j) \mid \boldsymbol{n}). \qquad (23)$$

This leads to the following result.

Proposition 3. *For the PIM model, a solution to (12) is given by*

$$\hat{\pi}((i, j) \mid \boldsymbol{n}) = \frac{1}{\mathcal{N}'}\left(n_{ij} + \theta_A P_i^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_j^B \hat{\pi}_A(i \mid \boldsymbol{n}_A)\right.$$
$$\left. + \frac{1}{2}\rho\left(\frac{n + \theta_A}{n + 1 + \theta_A} + \frac{n + \theta_B}{n + 1 + \theta_B}\right)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B)\right), \qquad (24)$$

where

$$\mathcal{N}' = n + \theta_A + \theta_B + \frac{1}{2}\rho\left(\frac{n + \theta_A}{n + 1 + \theta_A} + \frac{n + \theta_B}{n + 1 + \theta_B}\right).$$

Proof. This follows from substituting (23) into (12) and then solving for $\hat{\pi}((i, j) \mid \boldsymbol{n})$.

Similar to the case of diallelic loci, as $n \to \infty$ and $n_{ij}/n \to x_{ij}$, $\hat{\pi}((i, j) \mid \boldsymbol{n}) \to x_{ij}$.

3.7. Using $\hat{\pi}((i, j) \mid n)$ for IS

In a similar manner to the one-locus case, backward transition probabilities can be expressed in terms of the sampling distribution $\hat{\pi}((i, j) \mid n)$. However, it is clear from (13) that each recombination event increments the size of the sample, resulting in an extremely inefficient IS scheme that needlessly simulates complete ancestral recombination graphs (ARGs). This inefficiency can be circumvented by noting that the histories of genes which are not ancestral to the sample cannot affect the sample configuration, and so there is no gain in tracing them back in time. We propose a modified IS scheme which simulates coalescent histories only of genes carrying ancestral material, rather than complete ARGs. The sampling distribution $\hat{\pi}((i, j) \mid n)$ is modified accordingly, as we describe below, and forward transition probabilities are obtained from the corresponding two-locus recursion. This recursion is related to those studied in Golding (1984) and Ethier and Griffiths (1990), which differ from this approach in that they do not assign types to nonancestral loci.

Denote a gene of type $(i, j) \in E_A \times E_B$ which is ancestral at locus A only, at locus B only, and at both loci as $(i, j)^A$, $(i, j)^B$, and $(i, j)^C$, respectively, with corresponding multiplicities n_{ij}^A, n_{ij}^B, and n_{ij}^C. The state space for genes in this system can then be denoted by $(i, j, \gamma) \in E_A \times E_B \times \Gamma$, where $\Gamma = \{A, B, C\}$ and γ indicates at which loci the gene is ancestral (A only, B only, or both). Define $\boldsymbol{n}_\gamma = (n_{ij}^\gamma)_{(i,j) \in E_A \times E_B}$ and $n^\gamma = \sum_{(i,j) \in E_A \times E_B} n_{ij}^\gamma$ for $\gamma \in \Gamma$, so that $\boldsymbol{n} = (\boldsymbol{n}_A, \boldsymbol{n}_B, \boldsymbol{n}_C)$ and $n = n^A + n^B + n^C$. In the reduced scheme only genes ancestral at both loci, $\gamma = C$, can undergo recombination as we trace back in time. When sampling types $(i, j)^A$, $(i, j)^B$, and $(i, j)^C$, we use

$$\hat{\pi}((i, j)^A \mid n) = \hat{\pi}((i, j)^B \mid n) = \hat{\pi}((i, j)^C \mid n) = \hat{\pi}((i, j) \mid n).$$

An equation for $p(\boldsymbol{n})$ under this reduced scheme is

$$
\begin{aligned}
D_0 p(\boldsymbol{n}) = n \sum_{(i,j) \in E_A \times E_B} & \left(\sum_{\gamma \in \Gamma} (n_{ij}^\gamma - 1) p(\boldsymbol{n} - \boldsymbol{e}_{ij}^\gamma) + 2n_{ij}^C (p(\boldsymbol{n} - \boldsymbol{e}_{ij}^A) + p(\boldsymbol{n} - \boldsymbol{e}_{ij}^B)) \right. \\
& + 2(n_{ij}^C + 1) p(\boldsymbol{n} - \boldsymbol{e}_{ij}^A - \boldsymbol{e}_{ij}^B + \boldsymbol{e}_{ij}^C) \\
& + \theta_A \sum_{k \in E_A} \sum_{\gamma \in \Gamma} P_{ki}^A \frac{n_{kj}^\gamma + 1 - \delta_{ik}}{n} p(\boldsymbol{n} - \boldsymbol{e}_{ij}^\gamma + \boldsymbol{e}_{kj}^\gamma) \\
& + \theta_B \sum_{l \in E_B} \sum_{\gamma \in \Gamma} P_{lj}^B \frac{n_{il}^\gamma + 1 - \delta_{jl}}{n} p(\boldsymbol{n} - \boldsymbol{e}_{ij}^\gamma + \boldsymbol{e}_{il}^\gamma) \\
& \left. + \rho \sum_{(k,l) \in E_A \times E_B} \frac{(n_{il}^A + 1)(n_{kj}^B + 1)}{n(n+1)} p(\boldsymbol{n} + \boldsymbol{e}_{il}^A + \boldsymbol{e}_{kj}^B - \boldsymbol{e}_{ij}^C) \right),
\end{aligned}
$$

where $D_0 = n(n - 1) + n\theta_A + n\theta_B + \rho n^C$; $\boldsymbol{n} - \boldsymbol{e}_{ij}^A$ denotes $(\boldsymbol{n}_A - \boldsymbol{e}_{ij}, \boldsymbol{n}_B, \boldsymbol{n}_C)$, and so on.

Let $H_0, H_{-1}, \ldots, H_{-m}$ denote a sequence of states backwards in time, with the state H_0 denoting the input data. Recall that the proposal density for sequential IS is $\hat{p}(H_{k-1} \mid H_k) = p(H_k \mid H_{k-1}) \hat{p}(H_{k-1}) / \hat{p}(H_k)$, with $p(H_k \mid H_{k-1})$ being the forward transition probability, and the associated importance weight is $\hat{p}(H_k) / \hat{p}(H_{k-1})$. The forward transition probabilities and the ratio $\hat{p}(H_{k-1}) / \hat{p}(H_k)$ for the scheme developed above are shown in Table 2, whence the backward transition probabilities and IS weights can be derived. For sampling distributions

TABLE 2: Forward transition probabilities $\hat{p}(H_k \mid H_{k-1})$ and the ratio $\hat{p}(H_{k-1})/\hat{p}(H_k)$. The constant D_0 is defined as $D_0 = n(n-1) + n\theta_A + n\theta_B + n^C\rho$, and the multiplicity of H_k is $\boldsymbol{n} = (\boldsymbol{n}^A, \boldsymbol{n}^B, \boldsymbol{n}^C)$.

H_{k-1}	$p(H_k \mid H_{k-1})$	$\hat{p}(H_{k-1})/\hat{p}(H_k)$
	Coalescence	
$\boldsymbol{n} - \boldsymbol{e}_{ij}^A$	$\dfrac{n(2n_{ij}^C + n_{ij}^A - 1)}{D_0}$	$\dfrac{n_{ij}^A}{n\hat{\pi}((i,j)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^B$	$\dfrac{n(2n_{ij}^C + n_{ij}^B - 1)}{D_0}$	$\dfrac{n_{ij}^B}{n\hat{\pi}((i,j)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^B)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^C$	$\dfrac{n(n_{ij}^C - 1)}{D_0}$	$\dfrac{n_{ij}^C}{n\hat{\pi}((i,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^A - \boldsymbol{e}_{ij}^B + \boldsymbol{e}_{ij}^C$	$\dfrac{2n(n_{ij}^C + 1)}{D_0}$	$\dfrac{n_{ij}^A n_{ij}^B \hat{\pi}((i,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A - \boldsymbol{e}_{ij}^B)}{n(n_{ij}^C + 1)\hat{\pi}(\{(i,j)^A, (i,j)^B\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A - \boldsymbol{e}_{ij}^B)}$
	Mutation	
$\boldsymbol{n} - \boldsymbol{e}_{ij}^A + \boldsymbol{e}_{kj}^A$	$\dfrac{\theta_A P_{ki}^A (n_{kj}^A + 1 - \delta_{ik})}{D_0}$	$\dfrac{n_{ij}^A}{n_{kj}^A + 1 - \delta_{ik}} \dfrac{\hat{\pi}((k,j)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A)}{\hat{\pi}((i,j)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^B + \boldsymbol{e}_{kj}^B$	$\dfrac{\theta_A P_{ki}^A (n_{kj}^B + 1 - \delta_{ik})}{D_0}$	$\dfrac{n_{ij}^B}{n_{kj}^B + 1 - \delta_{ik}} \dfrac{\hat{\pi}((k,j)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^B)}{\hat{\pi}((i,j)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^B)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^C + \boldsymbol{e}_{kj}^C$	$\dfrac{\theta_A P_{ki}^A (n_{kj}^C + 1 - \delta_{ik})}{D_0}$	$\dfrac{n_{ij}^C}{n_{kj}^C + 1 - \delta_{ik}} \dfrac{\hat{\pi}((k,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}{\hat{\pi}((i,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^A + \boldsymbol{e}_{il}^A$	$\dfrac{\theta_B P_{lj}^B (n_{il}^A + 1 - \delta_{jl})}{D_0}$	$\dfrac{n_{ij}^A}{n_{il}^A + 1 - \delta_{jl}} \dfrac{\hat{\pi}((i,l)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A)}{\hat{\pi}((i,j)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^A)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^B + \boldsymbol{e}_{il}^B$	$\dfrac{\theta_B P_{lj}^B (n_{il}^B + 1 - \delta_{jl})}{D_0}$	$\dfrac{n_{ij}^B}{n_{il}^B + 1 - \delta_{jl}} \dfrac{\hat{\pi}((i,l)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^B)}{\hat{\pi}((i,j)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^B)}$
$\boldsymbol{n} - \boldsymbol{e}_{ij}^C + \boldsymbol{e}_{il}^C$	$\dfrac{\theta_B P_{lj}^B (n_{il}^C + 1 - \delta_{jl})}{D_0}$	$\dfrac{n_{ij}^C}{n_{il}^C + 1 - \delta_{jl}} \dfrac{\hat{\pi}((i,l)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}{\hat{\pi}((i,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}$
	Recombination	
$\boldsymbol{n} - \boldsymbol{e}_{ij}^C + \boldsymbol{e}_{il}^A + \boldsymbol{e}_{kj}^B$	$\dfrac{\rho(n_{il}^A + 1)(n_{kj}^B + 1)}{(n+1)D_0}$	$\dfrac{n_{ij}^C(n+1)\hat{\pi}(\{(i,l)^A, (k,j)^B\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}{(n_{il}^A + 1)(n_{kj}^B + 1)\hat{\pi}((i,j)^C \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)}$

involving the sampling of two additional alleles, we propose to use the following symmetrized definition:

$$\hat{\pi}(\{(i,l)^A, (k,j)^B\} \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C) = \tfrac{1}{2}(\hat{\pi}((i,l)^A \mid \boldsymbol{n} + \boldsymbol{e}_{kj}^B - \boldsymbol{e}_{ij}^C)\hat{\pi}((k,j)^B \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)$$
$$+ \hat{\pi}((k,j)^B \mid \boldsymbol{n} + \boldsymbol{e}_{il}^A - \boldsymbol{e}_{ij}^C)\,\hat{\pi}((i,l)^A \mid \boldsymbol{n} - \boldsymbol{e}_{ij}^C)).$$

4. Fearnhead and Donnelly's (2001) sampling distributions

In this section we describe the sampling distributions suggested by Fearnhead and Donnelly (2001). It was shown in De Iorio and Griffiths (2004a) that, in the absence of recombination, the diffusion approximation technique leads to the same sampling distributions as that proposed by Stephens and Donnelly (2000). It would therefore be of interest to study whether the diffusion approximation technique can produce novel sampling distributions when recombination is taken into account. For certain cases of the two-locus model, we obtain closed-form formulae for Fearnhead and Donnelly's sampling distributions and show that they differ from our sampling distributions. How these two different approximate distributions compare with the true distribution will be considered in Section 5.

4.1. Approximation for $\rho = 0$

Let H denote a multiset of n alleles. In the absence of recombination, Fearnhead and Donnelly's (2001) conditional distribution of the $(n + 1)$th sampled allele being of type ψ, given the type configuration H of the first n samples, is defined as

$$\hat{\pi}_{\text{FD}}(\psi \mid H) = \sum_{\phi \in H} \sum_{m=0}^{\infty} \frac{1}{n} \left(\frac{\theta}{n + \theta} \right)^m \left(\frac{n}{n + \theta} \right) [P^m]_{\phi \psi}, \tag{25}$$

where θ denotes the population-scaled mutation rate for the region being considered and P is the transition matrix. This distribution was first suggested by Stephens and Donnelly (2000). It corresponds to summing over all possible ways of choosing an allele from H at random with probability $1/n$, and then mutating it m number of times according to the mutation transition matrix P, with m being geometrically distributed with parameter $\theta/(n + \theta)$. If there are ℓ loci so that $\phi = (\phi_1, \ldots, \phi_\ell)$ and $\psi = (\psi_1, \ldots, \psi_\ell)$, then $[P^m]_{\phi \psi}$ in (25) needs to be replaced by

$$\sum_{\substack{m_1, \ldots, m_\ell \in \mathbb{Z}_{\geq 0}, \\ m_1 + \cdots + m_\ell = m}} \binom{m}{m_1, \ldots, m_\ell} \prod_{\alpha=1}^{\ell} \frac{1}{\ell^{m_\alpha}} [P^{m_\alpha}]_{\phi_\alpha \psi_\alpha}.$$

The multinomial coefficient $\binom{m}{m_1, \ldots, m_\ell}$ corresponds to the number of ways of arranging m_i, $i = 1, \ldots, \ell$, mutations at locus i, into a sequence of length $m = m_1 + \cdots + m_\ell$. (In this counting, mutations are regarded as being labeled by loci.) Dividing the multinomial by $\ell^{m_1 + \cdots + m_\ell}$ gives the probability of having m_1 mutations at locus 1, m_2 mutations at locus 2, and so on. Hence, for the ℓ-locus case, (25) can be written as

$$\hat{\pi}_{\text{FD}}(\psi \mid H) = \sum_{\phi \in H} \sum_{m_1=0}^{\infty} \cdots \sum_{m_\ell=0}^{\infty} \frac{1}{n + \theta} \binom{m}{m_1, \ldots, m_\ell} \prod_{\alpha=1}^{\ell} \frac{1}{\ell^{m_\alpha}} \left(\frac{\theta}{n + \theta} \right)^{m_\alpha} [P^{m_\alpha}]_{\phi_\alpha \psi_\alpha}.$$

Note that, since

$$\frac{1}{n} \int_0^\infty e^{-t} \prod_{\alpha=1}^{\ell} \left(e^{-t\theta/n\ell} \frac{(t\theta/n\ell)^{m_\alpha}}{m_\alpha!} [P^{m_\alpha}]_{\phi_\alpha \psi_\alpha} \right) dt = \frac{m!}{n + \theta} \prod_{\alpha=1}^{\ell} \left(\frac{\theta}{n + \theta} \right)^{m_\alpha} \frac{[P^{m_\alpha}]_{\phi_\alpha \psi_\alpha}}{\ell^{m_\alpha} m_\alpha!}$$

and $\exp(\lambda(A - I)) = \sum_{j=0}^{\infty} e^{-\lambda}(\lambda^j/j!)A^j$ for A a general $k \times k$ matrix and I a $k \times k$ identity matrix, we obtain

$$\hat{\pi}_{\text{FD}}(\psi \mid H) = \sum_{\phi \in H} \frac{1}{n} \int_0^\infty e^{-t} \prod_{\alpha=1}^{\ell} \left[M_\alpha \left(\frac{t}{n} \right) \right]_{\phi_\alpha \psi_\alpha} dt,$$

where $M_\alpha(t/n) = \exp((t\theta/n\ell)(P - I))$. More generally, if each locus α has its own mutation parameter θ_α and transition matrix P^α, then

$$M_\alpha\left(\frac{t}{n}\right) = \exp\left(\frac{t}{n}\theta_\alpha(P^\alpha - I)\right)$$

should be used.

4.2. Approximation for $\rho \neq 0$

To compare $\hat{\pi}_{\mathrm{FD}}(\psi \mid H)$ with our results, we assume that every allele in H contains ancestral material at every locus. Then, for $\rho \neq 0$, Fearnhead and Donnelly's sampling distribution is given by $\hat{\pi}_{\mathrm{FD}}(\psi \mid H) = p_\ell[\psi]$, where $p_\ell[\psi]$ is determined recursively using

$$p_\alpha[\psi] = \sum_{\phi \in H} \frac{1}{n} \int_0^\infty e^{-t} p_\alpha\left[\psi \mid \phi, \frac{t}{n}\right] dt \tag{26}$$

and

$$p_\alpha\left[\psi \mid \phi, \frac{t}{n}\right] = \left((1 - q_{\alpha-1})p_{\alpha-1}\left[\psi \mid \phi, \frac{t}{n}\right] + q_{\alpha-1}p_{\alpha-1}[\psi]\right)\left[M_\alpha\left(\frac{t}{n}\right)\right]_{\phi_\alpha\psi_\alpha}, \tag{27}$$

with the initial condition

$$p_1\left[\psi \mid \phi, \frac{t}{n}\right] = \left[M_1\left(\frac{t}{n}\right)\right]_{\phi_1\psi_1}. \tag{28}$$

Here, $q_{\alpha-1} := \rho_{\alpha-1}/(n + \rho_{\alpha-1})$, where $\rho_{\alpha-1}$ is the recombination rate between the $(\alpha-1)$th and the αth loci. The first factor $(1 - q_{\alpha-1})p_{\alpha-1}[\psi \mid \phi, t/n][M_\alpha(t/n)]_{\phi_\alpha\psi_\alpha}$ in (27) corresponds to there being no recombination between the $(\alpha - 1)$th and the αth loci, whereas the second factor $q_{\alpha-1}p_{\alpha-1}[\psi][M_\alpha(t/n)]_{\phi_\alpha\psi_\alpha}$ gives the contribution from having a recombination event between the two loci.

4.3. $\hat{\pi}_{\mathrm{FD}}((i, j) \mid n)$ for a diallelic two-locus model

We now focus on the two-locus model with two allele types at each locus. Let n denote the multiplicity configuration of H. As in Section 3, the type spaces for the first and the second loci are denoted by E_A and E_B, respectively, and the recombination rate between the two loci is denoted by ρ. For $(i, j) \in E_A \times E_B = \{1, 2\} \times \{1, 2\}$, we wish to compute

$$\hat{\pi}_{\mathrm{FD}}((i, j) \mid n) = p_2[(i, j)] = \sum_{(k,l) \in E_A \times E_B} \frac{n_{kl}}{n} \int_0^\infty e^{-t} p_2\left[(i, j) \mid (k, l), \frac{t}{n}\right] dt. \tag{29}$$

Here, $p_2[(i, j) \mid (k, l), t/n] = ((1-q_1)[M_1(t/n)]_{ki} + q_1 p_1[(i, j)])[M_2(t/n)]_{lj}$, where $q_1 = \rho/(n + \rho)$. In our computation we keep the mutation parameters for the two loci distinguished. Therefore, for locus α, $M_\alpha(t/n) = \exp((t/n)\theta_\alpha(P^\alpha - I))$, where

$$P^\alpha - I = \begin{pmatrix} -P_{12}^\alpha & P_{12}^\alpha \\ P_{21}^\alpha & -P_{21}^\alpha \end{pmatrix}.$$

From the initial condition (28), we find that

$$
\begin{aligned}
p_1\,[(i, j)] &= \sum_{(k,l)\in E_A \times E_B} \frac{n_{kl}}{n} \int_0^{\infty} e^{-t} \left[M_1\left(\frac{t}{n}\right) \right]_{ki} dt \\
&= \sum_{k\in E_A} \frac{n_{k\cdot}}{n} \int_0^{\infty} e^{-t} \left[M_1\left(\frac{t}{n}\right) \right]_{ki} dt \\
&= \frac{n_{i\cdot} + \theta_A P_{ii}^A}{n + \theta_A (P_{12}^A + P_{21}^A)},
\end{aligned}
$$

which is none other than $\hat{\pi}_A(i \mid \boldsymbol{n}_A)$ (cf. (18)). Furthermore, since

$$
\begin{aligned}
\sum_{(k,l)\in E_A \times E_B} \frac{n_{kl}}{n} \int_0^{\infty} e^{-t} \left[M_2\left(\frac{t}{n}\right) \right]_{lj} dt &= \sum_{l\in E_B} \frac{n_{\cdot l}}{n} \int_0^{\infty} e^{-t} \left[M_2\left(\frac{t}{n}\right) \right]_{lj} dt \\
&= \frac{n_{\cdot j} + \theta_B P_{jj}^B}{n + \theta_B (P_{12}^B + P_{21}^B)},
\end{aligned}
$$

which is equal to $\hat{\pi}_B(j \mid \boldsymbol{n}_B)$ (cf. (19)), Fearnhead and Donnelly's (2001) two-locus distribution in (29) can be written as

$$
\begin{aligned}
\hat{\pi}_{\mathrm{FD}}((i, j) \mid \boldsymbol{n}) &= \frac{n}{n + \rho} \sum_{(k,l)\in E_A \times E_B} \frac{n_{kl}}{n} \int_0^{\infty} e^{-t} \left[M_1\left(\frac{t}{n}\right) \right]_{ki} \left[M_2\left(\frac{t}{n}\right) \right]_{lj} dt \\
&\quad + \frac{\rho}{n + \rho} \hat{\pi}_A(i \mid \boldsymbol{n}_A) \hat{\pi}_B(j \mid \boldsymbol{n}_B).
\end{aligned}
$$

After performing the integral in the above expression, we obtain

$$
\begin{aligned}
&\hat{\pi}_{\mathrm{FD}}((i, j) \mid \boldsymbol{n}) \\
&= \frac{1}{\mathcal{N}_{\mathrm{FD}}} \Bigg(n_{ij} + \theta_A P_{ii}^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_{jj}^B \hat{\pi}_A(i \mid \boldsymbol{n}_A) \\
&\quad + \rho \left(\frac{n + \theta_A (P_{12}^A + P_{21}^A) + \theta_B (P_{12}^B + P_{21}^B)}{n} \right) \hat{\pi}_A(i \mid \boldsymbol{n}_A) \hat{\pi}_B(j \mid \boldsymbol{n}_B) \Bigg), \quad (30)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{N}_{\mathrm{FD}} &= n + \theta_A (P_{12}^A + P_{21}^A) + \theta_B (P_{12}^B + P_{21}^B) \\
&\quad + \rho \left(\frac{n + \theta_A (P_{12}^A + P_{21}^A) + \theta_B (P_{12}^B + P_{21}^B)}{n} \right).
\end{aligned}
$$

That $\hat{\pi}_{\mathrm{FD}}((i, j) \mid \boldsymbol{n})$ may take on this concise form is not obvious a priori. Simply applying the prescription shown in (26) and the equations thereafter generates complicated expressions, which are difficult to interpret at first sight. It is only after gathering many terms appropriately that we are able to write the distribution in such a simple form. In contrast, recall that, because in our approach we directly utilize the link between the one-locus and the two-locus diffusion models, our sampling distribution (17) could be obtained without intensive computation.

Note that the distribution in (30) is, in general, different from our sampling distribution (17). For $\rho = 0$ or $\rho = \infty$, however, the two distributions are equal. More exactly, we have

$$\lim_{\rho \to 0} \hat{\pi}((i, j) \mid \boldsymbol{n}) = \lim_{\rho \to 0} \hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n}) = \frac{n_{ij} + \theta_A P_{\bar{i}i}^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_{\bar{j}j}^B \hat{\pi}_A(i \mid \boldsymbol{n}_A)}{n + \theta_A(P_{12}^A + P_{21}^A) + \theta_B(P_{12}^B + P_{21}^B)}$$

and

$$\lim_{\rho \to \infty} \hat{\pi}((i, j) \mid \boldsymbol{n}) = \lim_{\rho \to \infty} \hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n}) = \hat{\pi}_A(i \mid \boldsymbol{n}_A) \hat{\pi}_B(j \mid \boldsymbol{n}_B).$$

As expected, the two loci become independent as ρ approaches ∞.

4.4. $\hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n})$ for two-locus PIM models

For PIM models, the transition matrix $\boldsymbol{P}^\alpha$ for locus $\alpha \in \{A, B\}$ satisfies $P_{ki}^\alpha = P_i^\alpha$ for all $i, k \in E_\alpha$. This implies that

$$\boldsymbol{M}_\alpha\left(\frac{t}{n}\right) = \exp\left(\frac{t}{n}\theta_\alpha(\boldsymbol{P}^\alpha - \boldsymbol{I})\right)$$

reduces to the following:

$$\left[\boldsymbol{M}_\alpha\left(\frac{t}{n}\right)\right]_{ki} = \begin{cases} \exp\left(-\frac{t}{n}\theta_\alpha\right)\left(\left(\exp\left(\frac{t}{n}\theta_\alpha\right) - 1\right)P_i^\alpha + 1\right) & \text{if } k = i, \\[2ex] \exp\left(-\frac{t}{n}\theta_\alpha\right)\left(\exp\left(\frac{t}{n}\theta_\alpha\right) - 1\right)P_i^\alpha & \text{if } k \neq i. \end{cases}$$

We omit the details of the computation of $\hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n})$, as they are similar to that discussed in the previous subsection. After carrying out the steps outlined in Section 4.2 and then gathering terms appropriately, we can show that

$$\hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n}) = \frac{1}{\mathcal{N}'_{\text{FD}}}\left(n_{ij} + \theta_A P_i^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_j^B \hat{\pi}_A(i \mid \boldsymbol{n}_A) \right.$$
$$\left. + \rho\left(\frac{n + \theta_A + \theta_B}{n}\right)\hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B)\right), \qquad (31)$$

where

$$\mathcal{N}'_{\text{FD}} = n + \theta_A + \theta_B + \rho\left(\frac{n + \theta_A + \theta_B}{n}\right).$$

Here, $\hat{\pi}_A(i \mid \boldsymbol{n}_A)$ and $\hat{\pi}_B(j \mid \boldsymbol{n}_B)$ are as shown in (21) and (22), respectively. As in the case of diallelic loci, Fearnhead and Donnelly's distributions, (31), for PIM models are generally different from our distributions, (24). However, for $\rho = 0$ or $\rho = \infty$, it is easy to see that

$$\lim_{\rho \to 0} \hat{\pi}((i, j) \mid \boldsymbol{n}) = \lim_{\rho \to 0} \hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n}) = \frac{n_{ij} + \theta_A P_i^A \hat{\pi}_B(j \mid \boldsymbol{n}_B) + \theta_B P_j^B \hat{\pi}_A(i \mid \boldsymbol{n}_A)}{n + \theta_A + \theta_B}$$

and

$$\lim_{\rho \to \infty} \hat{\pi}((i, j) \mid \boldsymbol{n}) = \lim_{\rho \to \infty} \hat{\pi}_{\text{FD}}((i, j) \mid \boldsymbol{n}) = \hat{\pi}_A(i \mid \boldsymbol{n}_A)\hat{\pi}_B(j \mid \boldsymbol{n}_B).$$

5. The two-locus infinitely-many-alleles model

As discussed in Section 4, our sampling distributions generally differ from that of Fearnhead and Donnelly's (2001). Unfortunately, it is in general difficult to know how these approximate distributions compare with the true distributions, the main reason being that the true distributions are difficult to obtain. In the case of the infinitely-many-alleles model, however, there exists a system of recursion relations satisfied by the true distribution (see Golding (1984) and Ethier and Griffiths (1990)), and it is possible to obtain exact solutions to the system when the sample size is less than 40 or so.

5.1. Distribution of unordered, unlabeled sample configurations

Let $\mathcal{A}_n$ denote the ordered configuration of n sequentially sampled alleles. Then, the probability distribution of an unordered, unlabeled sample configuration for the infinitely-many-alleles model is

$$p^*(\boldsymbol{n}) = \frac{n!}{\prod_{(i,j) \in E_A \times E_B} n_{ij}!} \frac{1}{\sigma(\boldsymbol{n})} p(\mathcal{A}_n),$$

where the symmetry factor $\sigma(\boldsymbol{n})$ is defined as $\sigma(\boldsymbol{n}) = |\{g \in \mathbb{Z}_2 \times \mathbb{Z}_2 \mid g(\boldsymbol{n}) = \boldsymbol{n}\}|$. The first $\mathbb{Z}_2$ acts on the first index of n_{ij}, whereas the second $\mathbb{Z}_2$ acts on the second index.

The true conditional distributions $\pi^*((i, j) \mid \boldsymbol{n})$ for an unordered, unlabeled sample configuration are defined as

$$\pi^*((i, j) \mid \boldsymbol{n}) = \frac{n_{ij} + 1}{n + 1} \frac{p^*(\boldsymbol{n} + \boldsymbol{e}_{ij})}{p^*(\boldsymbol{n})}.$$

One way to obtain the sampling distribution for the infinitely-many-alleles model is to take a limit in the PIM model. Let $p_{d,d'}(\boldsymbol{n})$ denote the probability distribution of a sample configuration $\boldsymbol{n}$ in the two-locus PIM model with allele type spaces $E_A = \{1, 2, \ldots, d\}$ and $E_B = \{1, 2, \ldots, d'\}$. Suppose that the sample contains k_A and k_B distinct allele types for loci A and B, respectively. Then, the probability distribution of an unordered, unlabeled sample configuration in the infinitely-many-alleles model is given by

$$p^*(\boldsymbol{n}) = \frac{1}{\sigma(\boldsymbol{n})} \lim_{d \to \infty} \lim_{d' \to \infty} d_{[k_A]} d'_{[k_B]} p_{d,d'}(\boldsymbol{n}),$$

where $d_{[k]} := d(d-1) \cdots (d-k+1)$. The corresponding conditional distributions are

$\pi^*((i, j) \mid \boldsymbol{n})$

$$= \frac{\sigma(\boldsymbol{n})}{\sigma(\boldsymbol{n} + \boldsymbol{e}_{ij})} \times \begin{cases} \lim_{d \to \infty} \lim_{d' \to \infty} \pi_{d,d'}((i, j) \mid \boldsymbol{n}) & \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} \neq 0, \\ \lim_{d \to \infty} \lim_{d' \to \infty} (d - k_A) \pi_{d,d'}((i, j) \mid \boldsymbol{n}) & \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} \neq 0, \\ \lim_{d \to \infty} \lim_{d' \to \infty} (d' - k_B) \pi_{d,d'}((i, j) \mid \boldsymbol{n}) & \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} = 0, \\ \lim_{d \to \infty} \lim_{d' \to \infty} (d - k_A)(d' - k_B) \pi_{d,d'}((i, j) \mid \boldsymbol{n}) & \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} = 0. \end{cases}$$

In what follows, we assume that at most two allele types are observed at each locus. Therefore, given a configuration $\boldsymbol{n} = (n_{11}, n_{12}, n_{21}, n_{22})$, the symmetry factor $\sigma(\boldsymbol{n})$ is as follows.

$$\sigma(\boldsymbol{n}) = \begin{cases} 4 & \text{if } n_{11} = n_{12} = n_{21} = n_{22}, \\ 2 & \text{if } n_{11} = n_{12}, n_{21} = n_{22}, \text{ and } n_{11} \neq n_{21}, \\ 2 & \text{if } n_{11} = n_{21}, n_{12} = n_{22}, \text{ and } n_{11} \neq n_{12}, \\ 1 & \text{otherwise.} \end{cases}$$

5.2. Our approximation

Following the above prescription, we use our approximate sampling distribution (24) for the PIM model to obtain the following approximate sampling distributions for the infinitely-many-alleles model:

$$
\hat{\pi}^*((i, j) \mid \boldsymbol{n})
$$

$$
= \frac{\sigma(\boldsymbol{n})}{\sigma(\boldsymbol{n} + e_{ij})}
$$

$$
\times \frac{1}{n + \theta_A + \theta_B + (\rho/2)((n + \theta_A)/(n + 1 + \theta_A) + (n + \theta_B)/(n + 1 + \theta_B))}
$$

$$
\times
\begin{cases}
n_{ij} + \dfrac{\rho}{2}\left(\dfrac{n_{i\cdot}}{n + 1 + \theta_A} \dfrac{n_{\cdot j}}{n + \theta_B} + \dfrac{n_{i\cdot}}{n + \theta_A} \dfrac{n_{\cdot j}}{n + 1 + \theta_B} \right) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} \neq 0, \\[2mm]
\theta_A \dfrac{n_{\cdot j}}{n + \theta_B} + \dfrac{\rho}{2}\left(\dfrac{\theta_A}{n + 1 + \theta_A} \dfrac{n_{\cdot j}}{n + \theta_B} + \dfrac{\theta_A}{n + \theta_A} \dfrac{n_{\cdot j}}{n + 1 + \theta_B} \right) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} \neq 0, \\[2mm]
\theta_B \dfrac{n_{i\cdot}}{n + \theta_A} + \dfrac{\rho}{2}\left(\dfrac{n_{i\cdot}}{n + 1 + \theta_A} \dfrac{\theta_B}{n + \theta_B} + \dfrac{n_{i\cdot}}{n + \theta_A} \dfrac{\theta_B}{n + 1 + \theta_B} \right) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} = 0, \\[2mm]
\theta_A \theta_B \left(\dfrac{1}{n + \theta_A} + \dfrac{1}{n + \theta_B} \right) + \dfrac{\rho}{2}\left(\dfrac{\theta_A}{n + 1 + \theta_A} \dfrac{\theta_B}{n + \theta_B} + \dfrac{\theta_A}{n + \theta_A} \dfrac{\theta_B}{n + 1 + \theta_B} \right) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} = 0.
\end{cases}
$$

5.3. Fearnhead and Donnelly's (2001) approximation

Using Fearnhead and Donnelly's approximate distribution, (31), for the PIM model leads the following approximate distributions for the infinitely-many-alleles model:

$$
\hat{\pi}^*_{\mathrm{FD}}((i, j) \mid \boldsymbol{n})
$$

$$
= \frac{\sigma(\boldsymbol{n})}{\sigma(\boldsymbol{n} + e_{ij})} \frac{1}{n + \theta_A + \theta_B + \rho((n + \theta_A + \theta_B)/n)}
$$

$$
\times
\begin{cases}
n_{ij} + \rho \dfrac{n + \theta_A + \theta_B}{n} \dfrac{n_{i\cdot}}{n + \theta_A} \dfrac{n_{\cdot j}}{n + \theta_B} & \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} \neq 0, \\[3mm]
\theta_A \dfrac{n_{\cdot j}}{n + \theta_B} + \rho \dfrac{n + \theta_A + \theta_B}{n} \dfrac{\theta_A}{n + \theta_A} \dfrac{n_{\cdot j}}{n + \theta_B} \\
\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} \neq 0, \\[3mm]
\theta_B \dfrac{n_{i\cdot}}{n + \theta_A} + \rho \dfrac{n + \theta_A + \theta_B}{n} \dfrac{n_{i\cdot}}{n + \theta_A} \dfrac{\theta_B}{n + \theta_B} \\
\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} \neq 0 \text{ and } n_{\cdot j} = 0, \\[3mm]
\theta_A \theta_B \left(\dfrac{1}{n + \theta_A} + \dfrac{1}{n + \theta_B} \right) + \rho \dfrac{n + \theta_A + \theta_B}{n} \dfrac{\theta_A}{n + \theta_A} \dfrac{\theta_B}{n + \theta_B} \\
\qquad\qquad\qquad\qquad\qquad \text{if } n_{i\cdot} = 0 \text{ and } n_{\cdot j} = 0.
\end{cases}
$$

5.4. Comparison with the true distribution

For $\rho = 0$, our approximate distributions agree exactly with that of Fearnhead and Donnelly's (2001) for all values of n_i. and $n_{\cdot j}$. For $\rho = \infty$, both

$$\hat{\pi}^*((i, j) \mid \boldsymbol{n}) \quad \text{and} \quad \hat{\pi}^*_{\text{FD}}((i, j) \mid \boldsymbol{n})$$

are equal to the true distribution

$$\lim_{\rho \to \infty} \pi^*((i, j) \mid \boldsymbol{n}) = \frac{\sigma(\boldsymbol{n})}{\sigma(\boldsymbol{n} + e_{ij})} \frac{1}{(n + \theta_A)(n + \theta_B)}$$
$$\times \begin{cases} n_i. n_{\cdot j} & \text{if } n_i. \neq 0 \text{ and } n_{\cdot j} \neq 0, \\ \theta_A n_{\cdot j} & \text{if } n_i. = 0 \text{ and } n_{\cdot j} \neq 0, \\ n_i. \theta_B & \text{if } n_i. \neq 0 \text{ and } n_{\cdot j} = 0, \\ \theta_A \theta_B & \text{if } n_i. = 0 \text{ and } n_{\cdot j} = 0, \end{cases}$$

which follows from Theorem 2.3 of Ethier and Griffiths (1990).

For comparing the two approximate distributions with the true distribution, we used Richard Hudson's C program (available at http://home.uchicago.edu/~rhudson1/) for solving Golding's recursion (see Golding (1984)) numerically. For ease of comparison, we used $\theta_A = \theta_B = \theta/2$. In what follows, we measure the deviation of an approximate distribution $\hat{\pi}^*_{\text{approx}}((i, j) \mid \boldsymbol{n})$ (either $\hat{\pi}^*((i, j) \mid \boldsymbol{n})$ or $\hat{\pi}^*_{\text{FD}}((i, j) \mid \boldsymbol{n})$) from the true distribution $\pi^*((i, j) \mid \boldsymbol{n})$ by

$$\frac{\hat{\pi}^*_{\text{approx}}((i, j) \mid \boldsymbol{n}) - \pi^*((i, j) \mid \boldsymbol{n})}{\pi^*((i, j) \mid \boldsymbol{n})} \times 100\%.$$

Shown in Figure 1(a) are the deviations of $\hat{\pi}^*((1, 1) \mid \boldsymbol{n})$ and $\hat{\pi}^*_{\text{FD}}((1, 1) \mid \boldsymbol{n})$ from the true distribution $\pi^*((1, 1) \mid \boldsymbol{n})$ for $\boldsymbol{n} = (4, 3, 2, 3)$. The case for $\boldsymbol{n} = (5, 4, 4, 5)$ is illustrated in Figure 1(b). Note that the deviations for $\boldsymbol{n} = (5, 4, 4, 5)$ are generally less than those for $\boldsymbol{n} = (4, 3, 2, 3)$. Furthermore, both cases indicate that, in general, our distribution $\hat{\pi}^*((1, 1) \mid \boldsymbol{n})$ is a more accurate approximation of the true distribution than is Fearnhead and Donnelly's $\hat{\pi}^*_{\text{FD}}((1, 1) \mid \boldsymbol{n})$.

In contrast to the above case where $n_i. \neq 0$ and $n_{\cdot j} \neq 0$, for $n_i. = 0$ or $n_{\cdot j} = 0$, both $\hat{\pi}^*((i, j) \mid \boldsymbol{n})$ and $\hat{\pi}^*_{\text{FD}}((i, j) \mid \boldsymbol{n})$ diverge more and more from the true distribution as n increases. This puzzling behavior is illustrated in Figure 1(c)–(d) for $(i, j) = (1, 1)$, with $n_1. = 0$ for $n_{\cdot 1} \neq 0$. In Figure 1(c) $\boldsymbol{n} = (0, 0, 1, 0)$, whereas $\boldsymbol{n} = (0, 0, 2, 1)$ in Figure 1(d). In general, it still seems true that our approximate distribution $\hat{\pi}^*((i, j) \mid \boldsymbol{n})$ is closer to the true distribution than is $\hat{\pi}^*_{\text{FD}}((i, j) \mid \boldsymbol{n})$.

6. The two-locus model with subdivided population structure

De Iorio and Griffiths (2004a) applied their diffusion approximation technique to the one-locus model with subdivided population structure and showed that, in the case of the infinitely-many-sites model, it leads to a significant improvement over previous studies (see Bahlo and Griffiths (2000)). In this section we apply the diffusion approximation technique to the two-locus model with subdivided population structure, thus obtaining novel results. Much of what we discussed in Section 3 carries over nicely to the present case. This example well illustrates the wide applicability of the diffusion approximation technique.
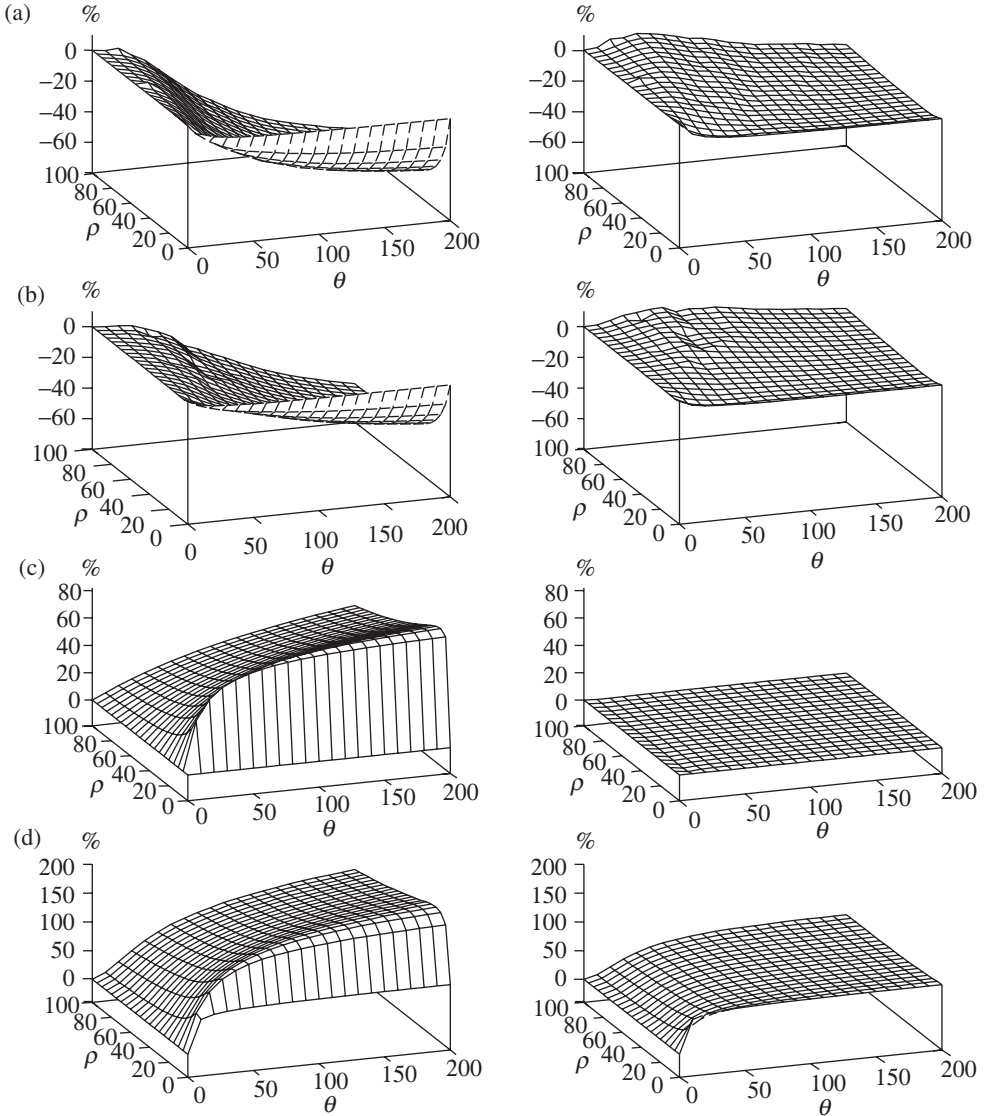
FIGURE 1: *Left column*: Deviation of Fearnhead and Donnelly's distribution $\hat{\pi}^*_{\text{FD}}((1, 1) \mid \boldsymbol{n})$ from the true distribution. *Right column*: Deviation of our distribution $\hat{\pi}^*((1, 1) \mid \boldsymbol{n})$ from the true distribution. For all figures, $\theta_A = \theta_B = \theta/2$. (a) $\boldsymbol{n} = (4, 3, 2, 3)$. (b) $\boldsymbol{n} = (5, 4, 4, 5)$. (c) $\boldsymbol{n} = (0, 0, 1, 0)$. (d) $\boldsymbol{n} = (0, 0, 2, 1)$.

6.1. Diffusion approximation

Subpopulations are labeled by $\Gamma = \{1, \ldots, g\}$. The total population size is $N = \sum_{\alpha \in \Gamma} N_\alpha$, where N_α denotes the size of subpopulation α. We define $q_\alpha = N_\alpha/N$ for all $\alpha \in \Gamma$. For $\alpha, \beta \in \Gamma$, where $\alpha \neq \beta$, the scaled backward migration rates are defined as $m_{\alpha\beta} = 2Nc_{\alpha\beta}$, where $c_{\alpha\beta}$ denotes the probability that the parent of an individual in subpopulation α is

from subpopulation β one generation back in time. The overall scaled migration rate for subpopulation α is $m_\alpha = \sum_{\beta \neq \alpha} m_{\alpha\beta}$. See De Iorio and Griffiths (2004b) for a detailed discussion on the coalescent model with migration.

Similar notation as in Section 3 is used in this section. The population gene frequencies are denoted by $X = (X_{\alpha ij})_{(\alpha,i,j) \in \Gamma \times E_A \times E_B}$. These frequencies are normalized so that, for each subpopulation $\alpha \in \Gamma$, $\sum_{(i,j) \in E_A \times E_B} X_{\alpha ij} = 1$. The diffusion process generator for the two-locus model with migration is

$$\mathcal{L} = \sum_{\alpha \in \Gamma} \sum_{(i,j) \in E_A \times E_B} L_{\alpha ij} \frac{\partial}{\partial x_{\alpha ij}},$$

where

$$L_{\alpha ij} = \frac{1}{2} \sum_{(k,l) \in E_A \times E_B} x_{\alpha ij} (\delta_{ik} \delta_{jl} - x_{\alpha kl}) q_\alpha^{-1} \frac{\partial}{\partial x_{\alpha kl}} + \frac{\theta_A}{2} \sum_{k \in E_A} x_{\alpha kj} (P_{ki}^A - \delta_{ki})$$
$$+ \frac{\theta_B}{2} \sum_{l \in E_B} x_{\alpha il} (P_{lj}^B - \delta_{lj}) + \frac{\rho}{2} (x_{\alpha i \cdot} x_{\alpha \cdot j} - x_{\alpha ij}) - \frac{1}{2} m_\alpha x_{\alpha ij} + \frac{1}{2} \sum_{\beta \neq \alpha} m_{\alpha\beta} x_{\beta ij}.$$

In what follows, define $\boldsymbol{n} := (n_{\alpha ij})_{(\alpha,i,j) \in \Gamma \times E_A \times E_B}$, $\boldsymbol{n}_\alpha := (n_{\alpha ij})_{(i,j) \in E_A \times E_B}$, and

$$Q_S(\boldsymbol{x}, \boldsymbol{n}) := \prod_{\alpha \in \Gamma} \binom{n_\alpha}{\boldsymbol{n}_\alpha} \prod_{(i,j) \in E_A \times E_B} x_{\alpha ij}^{n_{\alpha ij}}.$$

Proposition 4. *Assume that there exists a probability distribution with expectation operator $\hat{E}$ such that $\hat{E}(L_{\alpha ij} Q_S(X, \boldsymbol{n})) = 0$ for all $(\alpha, i, j) \in \Gamma \times E_A \times E_B$, and define $\hat{p}(\boldsymbol{n}) := \hat{E}(Q_S(X, \boldsymbol{n}))$ and $\hat{\pi}((i,j) \mid \alpha, \boldsymbol{n}) = \hat{E}(X_{\alpha ij} Q_S(X, \boldsymbol{n}))\hat{p}(\boldsymbol{n})$. Furthermore, assume that exchangeability conditions analogous to (7)–(9) hold. Then,*

$$(n_\alpha q_\alpha^{-1} + \rho + m_\alpha + \theta_A + \theta_B)\hat{\pi}((i,j) \mid \alpha, \boldsymbol{n})$$
$$= n_{\alpha ij} q_\alpha^{-1} + \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((i,j) \mid \beta, \boldsymbol{n}) + \theta_A \sum_{k \in E_A} \hat{\pi}((k,j) \mid \alpha, \boldsymbol{n}) P_{ki}^A$$
$$+ \theta_B \sum_{l \in E_B} \hat{\pi}((i,l) \mid \alpha, \boldsymbol{n}) P_{lj}^B + \rho \hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \alpha, \boldsymbol{n}), \qquad (32)$$

where $\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \alpha, \boldsymbol{n}) = \hat{E}(X_{\alpha i \cdot} X_{\alpha \cdot j} Q_S(X, \boldsymbol{n}))/\hat{p}(\boldsymbol{n})$. (Normalization is such that $\sum_{(i,j) \in E_A \times E_B} \hat{\pi}((i,j) \mid \alpha, \boldsymbol{n}) = 1$ for each $\alpha \in \Gamma$.)

Proof. This follows from a similar set of steps as in the proof of Proposition 1.

As in the case of a single population, the main difficulty in solving equation (32) for the conditional sampling distribution $\hat{\pi}((i,j) \mid \alpha, \boldsymbol{n})$ comes from the recombination term $\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \alpha, \boldsymbol{n})$. As before, our approach is to first obtain a system of equations relating $\hat{\pi}(\{(i,\cdot),(\cdot,j)\} \mid \alpha, \boldsymbol{n})$ to $\hat{\pi}((i,j) \mid \alpha, \boldsymbol{n})$; this can be done using techniques similar to the ones already discussed in Section 3.

6.2. Marginal distributions

Summing over the index j in (32) yields

$$(n_\alpha q_\alpha^{-1} + m_\alpha + \theta_A)\hat{\pi}((i, \cdot) \mid \alpha, \boldsymbol{n}) = n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A \sum_{k \in E_A} \hat{\pi}((k, \cdot) \mid \alpha, \boldsymbol{n}) P_{ki}^A$$
$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((i, \cdot) \mid \beta, \boldsymbol{n}), \qquad (33)$$

whereas summing over the index i yields

$$(n_\alpha q_\alpha^{-1} + m_\alpha + \theta_B)\hat{\pi}((\cdot, j) \mid \alpha, \boldsymbol{n}) = n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B \sum_{l \in E_B} \hat{\pi}((\cdot, l) \mid \alpha, \boldsymbol{n}) P_{lj}^B$$
$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((\cdot, j) \mid \beta, \boldsymbol{n}). \qquad (34)$$

Note that these are marginal systems depending only on $\boldsymbol{n}_A = (n_{\gamma a \cdot})_{(\gamma, a) \in \Gamma \times E_A}$ and $\boldsymbol{n}_B = (n_{\gamma \cdot b})_{(\gamma, b) \in \Gamma \times E_B}$, respectively. The one-locus distribution discussed in De Iorio and Griffiths (2004b) satisfies an equation exactly like (33) and (34). There exists a unique solution to such an equation; in general, it can be obtained via matrix inversion. (See De Iorio and Griffiths (2004b) for details.) We therefore conclude that $\hat{\pi}((i, \cdot) \mid \alpha, \boldsymbol{n}) = \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)$ and $\hat{\pi}((\cdot, j) \mid \alpha, \boldsymbol{n}) = \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B)$, where $\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)$ and $\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B)$ are one-locus distributions for the A and B loci, respectively.

6.3. $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$ for diallelic loci

We first consider the case of diallelic loci; that is, $E_A = E_B = \{1, 2\}$. Recall that we define $\bar{i} = 1$ if $i = 2$ and $\bar{i} = 2$ if $i = 1$.

6.3.1. *An arbitrary number of subpopulations.* To be concise, we define

$$\mathcal{F}_\alpha^A(k) = k q_\alpha^{-1} + m_\alpha + \theta_A(P_{12}^A + P_{21}^A) \quad \text{and} \quad \mathcal{F}_\alpha^B(k) = k q_\alpha^{-1} + m_\alpha + \theta_B(P_{12}^B + P_{21}^B).$$

Then, (33) and (34) can be respectively written as

$$\mathcal{F}_\alpha^A(n_\alpha)\hat{\pi}((i, \cdot) \mid \alpha, \boldsymbol{n}) = n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A P_{\bar{i}i}^A + \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((i, \cdot) \mid \beta, \boldsymbol{n}),$$

$$\mathcal{F}_\alpha^B(n_\alpha)\hat{\pi}((\cdot, j) \mid \alpha, \boldsymbol{n}) = n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_{\bar{j}j}^B + \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((\cdot, j) \mid \beta, \boldsymbol{n}).$$

These systems of equations can be solved as follows. Let $\boldsymbol{v}_i^A = (v_{i\alpha}^A)_{\alpha \in \Gamma}$ and $\boldsymbol{v}_j^B = (v_{j\alpha}^B)_{\alpha \in \Gamma}$ denote the g-dimensional column vectors with entries

$$v_{i\alpha}^A = \frac{n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A P_{\bar{i}i}^A}{\mathcal{F}_\alpha^A(n_\alpha)} \quad \text{and} \quad v_{j\alpha}^B = \frac{n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_{\bar{j}j}^B}{\mathcal{F}_\alpha^B(n_\alpha)}, \qquad (35)$$

respectively. Also, let $\boldsymbol{M}^A = (M_{\alpha\beta}^A)_{(\alpha, \beta) \in \Gamma \times \Gamma}$ be the $g \times g$ matrix with entries

$$M_{\alpha\beta}^A = \begin{cases} \dfrac{m_{\alpha\beta}}{\mathcal{F}_\alpha^A(n_\alpha)} & \text{if } \alpha \neq \beta, \\ 0 & \text{otherwise.} \end{cases} \qquad (36)$$

The matrix $\boldsymbol{M}^B$ is similarly defined, with $\mathcal{F}_\alpha^A(n_\alpha)$ replaced with $\mathcal{F}_\alpha^B(n_\alpha)$. Then, $\hat{\pi}((i, \cdot) \mid \alpha, \boldsymbol{n})$ is the αth component of the g-dimensional vector $(\boldsymbol{I} - \boldsymbol{M}^A)^{-1}\boldsymbol{v}_i^A$, where $\boldsymbol{I}$ is the $g \times g$ identity matrix. Likewise, $\hat{\pi}((\cdot, j) \mid \alpha, \boldsymbol{n})$ is the αth component of the g-dimensional vector $(\boldsymbol{I} - \boldsymbol{M}^B)^{-1}\boldsymbol{v}_j^B$.

Now, from (35) and (36), we see that

$$\hat{\pi}((i, \cdot) \mid \alpha, \boldsymbol{n} + \boldsymbol{e}_{\alpha kj}) = \mathcal{C}_\alpha^A \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) + \mathcal{D}_\alpha^A \delta_{ik}, \tag{37}$$

where $\mathcal{C}_\alpha^A$ and $\mathcal{D}_\alpha^A$ are some constants that depend on $\theta_A \boldsymbol{P}^A$, n_α, q_α, and $m_{\alpha\beta}$, where $\alpha, \beta \in \Gamma$. Similarly,

$$\hat{\pi}((\cdot, j) \mid \alpha, \boldsymbol{n} + \boldsymbol{e}_{\alpha il}) = \mathcal{C}_\alpha^B \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) + \mathcal{D}_\alpha^B \delta_{jl}. \tag{38}$$

Using these facts, (32) can be written as

$$\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n}) = \lambda_\alpha \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}((i, j) \mid \beta, \boldsymbol{n}) + v_{ij\alpha},$$

where $\lambda_\alpha = 2\{\mathcal{F}_\alpha^A(n_\alpha) + \mathcal{F}_\alpha^B(n_\alpha) + \rho((1 - \mathcal{D}_\alpha^A) + (1 - \mathcal{D}_\alpha^B))\}^{-1}$ and

$$v_{ij\alpha} := \lambda_\alpha \left(n_{\alpha ij} q_\alpha^{-1} + \theta_A P_{ii}^A \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) + \theta_B P_{jj}^B \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) \right.$$
$$\left. + \frac{\rho}{2}(\mathcal{C}_\alpha^A + \mathcal{C}_\alpha^B)\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) \right).$$

Here, we have used a symmetrized form of $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$ similar to that shown in (14). Define $\boldsymbol{v}_{ij} := (v_{ij\alpha})_{\alpha \in \Gamma}$, a g-dimensional column vector. Note that we have already discussed how everything that appears in $\boldsymbol{v}_{ij}$ can be computed. Hence, defining $\boldsymbol{R} = (R_{\alpha\beta})_{(\alpha, \beta) \in \Gamma \times \Gamma}$ as the $g \times g$ matrix with entries

$$R_{\alpha\beta} = \begin{cases} \lambda_\alpha m_{\alpha\beta} & \text{if } \alpha \neq \beta, \\ 0 & \text{otherwise}, \end{cases}$$

we can obtain $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$ as the αth component of $(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{v}_{ij}$.

6.3.2. *Two subpopulations.* When there are only two subpopulations, labeled by $\Gamma = \{1, 2\}$, the method described above reduces to solving a system of two independent equations in two variables, and we can avoid doing matrix inversion. For instance, the A locus marginal distributions satisfy $\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) = M_{\alpha\beta}^A \hat{\pi}_A(i \mid \beta, \boldsymbol{n}_A) + v_{i\alpha}^A$, where constants $v_{i\alpha}^A$ and $M_{\alpha\beta}^A$ are as defined in (35) and (36), respectively. Here, $\alpha, \beta \in \Gamma$ with $\alpha \neq \beta$. We can easily solve the above system of equations to obtain

$$\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) = \frac{M_{\alpha\beta}^A v_{i\beta}^A + v_{i\alpha}^A}{1 - M_{\alpha\beta}^A M_{\beta\alpha}^A}.$$

Solutions to the B locus marginal distributions $\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B)$ can be obtained in a similar way. In terms of the given parameters, these marginal distributions can be written as

$$\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) = \left(\frac{n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A P_{ii}^A}{\mathcal{F}_\alpha^A(n_\alpha)} + \frac{m_{\alpha\beta}(n_{\beta i \cdot} q_\beta^{-1} + \theta_A P_{ii}^A)}{\mathcal{F}_\alpha^A(n_\alpha)\,\mathcal{F}_\beta^A(n_\beta)} \right) \left(1 - \frac{m_{\alpha\beta} m_{\beta\alpha}}{\mathcal{F}_\alpha^A(n_\alpha)\,\mathcal{F}_\beta^A(n_\beta)} \right)^{-1},$$

$$\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) = \left(\frac{n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_{jj}^B}{\mathcal{F}_\alpha^B(n_\alpha)} + \frac{m_{\alpha\beta}(n_{\beta \cdot j} q_\beta^{-1} + \theta_B P_{jj}^B)}{\mathcal{F}_\alpha^B(n_\alpha)\,\mathcal{F}_\beta^B(n_\beta)} \right) \left(1 - \frac{m_{\alpha\beta} m_{\beta\alpha}}{\mathcal{F}_\alpha^B(n_\alpha)\,\mathcal{F}_\beta^B(n_\beta)} \right)^{-1},$$

from which we see that the constants $\mathcal{C}_\alpha^A$ and $\mathcal{D}_\alpha^A$ in (37) are given by

$$\mathcal{C}_\alpha^A = \frac{\mathcal{F}_\alpha^A(n_\alpha)\,\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^A(n_\alpha + 1)\,\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \quad \text{and} \quad \mathcal{D}_\alpha^A = \frac{q_\alpha^{-1}\mathcal{F}_\beta^A(n_\beta)}{\mathcal{F}_\alpha^A(n_\alpha)\,\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}.$$

The constants $\mathcal{C}_\alpha^B$ and $\mathcal{D}_\alpha^B$ in (38) are similarly defined. As discussed in Section 6.3.1, these constants, along with the marginal distributions found above, completely determine $R_{\alpha\beta}$ and $v_{ij\alpha}$ in the system of equations $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n}) = R_{\alpha\beta}\hat{\pi}((i, j) \mid \beta, \boldsymbol{n}) + v_{ij\alpha}$, where $\alpha, \beta \in \Gamma$ with $\alpha \neq \beta$. The two-locus sampling distributions $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$ are thus given by

$$\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n}) = \frac{R_{\alpha\beta}v_{ij\beta} + v_{ij\alpha}}{1 - R_{\alpha\beta}R_{\beta\alpha}}.$$

6.4. $\hat{\pi}((i, j) \mid \alpha, n)$ for PIM models

To obtain a system of equations relating $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$ and $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$ for PIM models, we also need

$$\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\alpha, \beta\}, \boldsymbol{n}) = \frac{\hat{\mathrm{E}}(X_{\alpha i\cdot}X_{\beta\cdot j}Q_S(\boldsymbol{X}, \boldsymbol{n}))}{\hat{p}(\boldsymbol{n})},$$

the probability that the marginal types of two genes from subpopulations α and β are i and j at the A and B loci, respectively. In (33) set $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{e}_{\alpha l j}$ and sum over l, after multiplying by $\hat{p}(\boldsymbol{n} + \boldsymbol{e}_{\alpha l j})(\prod_{(\sigma, r, s) \in \Gamma \times E_A \times E_B}(n_{\sigma r s} + \delta_{\sigma\alpha}\delta_{rl}\delta_{sj})!)/(n + 1)!$. The resulting equations after multiplying again by $n!/(\hat{p}(\boldsymbol{n})\prod_{(\sigma, r, s) \in \Gamma \times E_A \times E_B} n_{\sigma r s}!)$ are

$$((n_\alpha + 1)q_\alpha^{-1} + m_\alpha + \theta_A)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$$
$$= q_\alpha^{-1}(n_{\alpha i\cdot}\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) + \hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})) + \theta_A \sum_{k \in E} P_{ki}^A \hat{\pi}(\{(k, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$$
$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta}\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}). \tag{39}$$

In the above computation if we instead set $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{e}_{\gamma l j}$, where $\gamma \neq \alpha$, then we obtain

$$(n_\alpha q_\alpha^{-1} + m_\alpha + \theta_A)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\alpha, \gamma\}, \boldsymbol{n})$$
$$= n_{\alpha i\cdot}q_\alpha^{-1}\hat{\pi}_B(j \mid \gamma, \boldsymbol{n}_B) + \theta_A \sum_{k \in E} P_{ki}^A \hat{\pi}(\{(k, \cdot), (\cdot, j)\} \mid \{\alpha, \gamma\}, \boldsymbol{n})$$
$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta}\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \gamma\}, \boldsymbol{n}). \tag{40}$$

In a similar vein, the following systems of equations can be found using (34):

$$((n_\alpha + 1)q_\alpha^{-1} + m_\alpha + \theta_B)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$$
$$= q_\alpha^{-1}(n_{\alpha\cdot j}\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) + \hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})) + \theta_B \sum_{l \in E} P_{lj}^B \hat{\pi}(\{(i, \cdot), (\cdot, l)\} \mid \alpha, \boldsymbol{n})$$
$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta}\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}), \tag{41}$$

and, for $\gamma \neq \alpha$,

$$(n_\alpha q_\alpha^{-1} + m_\alpha + \theta_B)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\alpha, \gamma\}, \boldsymbol{n})$$

$$= n_{\alpha \cdot j} q_\alpha^{-1} \hat{\pi}_A(i \mid \gamma, \boldsymbol{n}_A) + \theta_B \sum_{l \in E} P_{lj}^B \hat{\pi}(\{(i, \cdot), (\cdot, l)\} \mid \{\alpha, \gamma\}, \boldsymbol{n})$$

$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \gamma\}, \boldsymbol{n}). \tag{42}$$

6.4.1. *An arbitrary number of subpopulations.* Henceforward, we let $\mathcal{F}_\alpha^A(k) = kq_\alpha^{-1} + m_\alpha + \theta_A$ and $\mathcal{F}_\alpha^B(k) = kq_\alpha^{-1} + m_\alpha + \theta_B$. For PIM models, (33) and (34) reduce to

$$\mathcal{F}_\alpha^A(n_\alpha) \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) = n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A P_i^A + \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}_A(i \mid \beta, \boldsymbol{n}_A), \tag{43}$$

$$\mathcal{F}_\alpha^B(n_\alpha) \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) = n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_j^B + \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}_B(j \mid \beta, \boldsymbol{n}_B). \tag{44}$$

These one-locus equations, also considered in De Iorio and Griffiths (2004b), can be solved using matrix inversion. Moreover, (39) and (41) become

$$\mathcal{F}_\alpha^A(n_\alpha + 1)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$$

$$= (n_{\alpha i \cdot} q_\alpha^{-1} + \theta_A P_i^A)\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) + q_\alpha^{-1}\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$$

$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}), \tag{45}$$

$$\mathcal{F}_\alpha^B(n_\alpha + 1)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$$

$$= (n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_j^B)\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) + q_\alpha^{-1}\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$$

$$+ \sum_{\beta \neq \alpha} m_{\alpha\beta} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}), \tag{46}$$

while (40) and (42) imply, for $\beta \neq \alpha$, the following equations:

$$\mathcal{F}_\beta^A(n_\beta)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}) = (n_{\beta i \cdot} q_\beta^{-1} + \theta_A P_i^A)\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B)$$

$$+ \sum_{\gamma \neq \beta} m_{\beta\gamma} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\gamma, \alpha\}, \boldsymbol{n}), \tag{47}$$

$$\mathcal{F}_\beta^B(n_\beta)\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\beta, \alpha\}, \boldsymbol{n}) = (n_{\beta \cdot j} q_\beta^{-1} + \theta_B P_j^B)\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)$$

$$+ \sum_{\gamma \neq \beta} m_{\beta\gamma} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\gamma, \alpha\}, \boldsymbol{n}). \tag{48}$$

Now, (45)–(48) can be used to express $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$ in terms of $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$, $\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)$, $\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B)$, and known parameters. That relation can then be used in (32) to solve for $\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$. A concrete example is discussed in the next section.

6.4.2. *Two subpopulations.* Let $\alpha, \beta \in \Gamma$ such that $\alpha \neq \beta$. For $\Gamma = \{1, 2\}$, it is easy to solve the systems of equations shown in (43) and (44). The solutions are

$$\hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) = \left(\frac{n_{\alpha i} \cdot q_\alpha^{-1} + \theta_A P_i^A}{\mathcal{F}_\alpha^A(n_\alpha)} + \frac{m_{\alpha\beta}(n_{\beta i} \cdot q_\beta^{-1} + \theta_A P_i^A)}{\mathcal{F}_\alpha^A(n_\alpha)\mathcal{F}_\beta^A(n_\beta)} \right) \left(1 - \frac{m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^A(n_\alpha)\mathcal{F}_\beta^A(n_\beta)} \right)^{-1},$$

$$\hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) = \left(\frac{n_{\alpha \cdot j} q_\alpha^{-1} + \theta_B P_j^B}{\mathcal{F}_\alpha^B(n_\alpha)} + \frac{m_{\alpha\beta}(n_{\beta \cdot j} q_\beta^{-1} + \theta_B P_j^B)}{\mathcal{F}_\alpha^B(n_\alpha)\mathcal{F}_\beta^B(n_\beta)} \right) \left(1 - \frac{m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^B(n_\alpha)\mathcal{F}_\beta^B(n_\beta)} \right)^{-1}.$$

In (47) and (48), note that the sum $\sum_{\gamma \neq \beta} m_{\beta\gamma} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \{\gamma, \alpha\}, \boldsymbol{n})$ reduces to the single term $m_{\beta\alpha} \hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$. It is then easy to see that (45) and (47) imply that

$$\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n}) = \frac{q_\alpha^{-1} \mathcal{F}_\beta^A(n_\beta)}{\mathcal{F}_\alpha^A(n_\alpha + 1)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$$

$$+ \frac{\mathcal{F}_\alpha^A(n_\alpha)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^A(n_\alpha + 1)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B),$$

whereas (46) and (48) imply that

$$\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n}) = \frac{q_\alpha^{-1} \mathcal{F}_\beta^B(n_\beta)}{\mathcal{F}_\alpha^B(n_\alpha + 1) \mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \hat{\pi}((i, j) \mid \alpha, \boldsymbol{n})$$

$$+ \frac{\mathcal{F}_\alpha^B(n_\alpha) \mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^B(n_\alpha + 1) \mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B).$$

Symmetrizing $\hat{\pi}(\{(i, \cdot), (\cdot, j)\} \mid \alpha, \boldsymbol{n})$ with respect to the A and B loci using the above results, (32) can now be written as

$$\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n}) = \xi_\alpha m_{\alpha\beta} \hat{\pi}((i, j) \mid \beta, \boldsymbol{n}) + w_{ij\alpha}, \tag{49}$$

where

$$w_{ij\alpha} = \xi_\alpha \Bigg(n_{\alpha ij} + \theta_A P_i^A \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) + \theta_B P_j^B \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A)$$

$$+ \frac{\rho}{2} \left(\frac{\mathcal{F}_\alpha^A(n_\alpha)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^A(n_\alpha + 1)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} + \frac{\mathcal{F}_\alpha^B(n_\alpha)\mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}}{\mathcal{F}_\alpha^B(n_\alpha + 1)\mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \right)$$

$$\times \hat{\pi}_A(i \mid \alpha, \boldsymbol{n}_A) \hat{\pi}_B(j \mid \alpha, \boldsymbol{n}_B) \Bigg)$$

and

$$\xi_\alpha^{-1} = \frac{1}{2} \Bigg(\mathcal{F}_\alpha^A(n_\alpha) + \mathcal{F}_\alpha^B(n_\alpha) + \rho \left(1 - \frac{q_\alpha^{-1} \mathcal{F}_\beta^A(n_\beta)}{\mathcal{F}_\alpha^A(n_\alpha + 1)\mathcal{F}_\beta^A(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \right)$$

$$+ \rho \left(1 - \frac{q_\alpha^{-1} \mathcal{F}_\beta^B(n_\beta)}{\mathcal{F}_\alpha^B(n_\alpha + 1)\mathcal{F}_\beta^B(n_\beta) - m_{\alpha\beta}m_{\beta\alpha}} \right) \Bigg).$$

Finally, we can now solve (49) to obtain

$$\hat{\pi}((i, j) \mid \alpha, \boldsymbol{n}) = \frac{\xi_\alpha m_{\alpha\beta} w_{ij\beta} + w_{ij\alpha}}{1 - \xi_\alpha \xi_\beta m_{\alpha\beta} m_{\beta\alpha}}.$$

7. Discussion

The efficiency of an IS method hinges heavily on how close the adopted proposal distribution is to the true distribution. The key insight provided by Stephens and Donnelly (2000) is that, in sequential IS schemes that arise in population genetics, proposal distributions can be written in terms of one-dimensional conditional sampling distributions whose optimality properties can be characterized. We can therefore translate the problem of constructing good IS proposal distributions into that of constructing good one-dimensional conditional sampling distributions. So far, most approaches to the latter problem have relied to a large degree on one's intuition and experience. In contrast, the diffusion-generator approximation method developed by De Iorio and Griffiths (2004a), (2004b) is a systematic approach that provides a general mathematical prescription for constructing good conditional sampling distributions.

In the present paper we have extended the diffusion-generator approximation technique of De Iorio and Griffiths (2004a), (2004b) to the neutral coalescent model with recombination, obtaining explicit sampling formulae for diallelic and PIM models. We have addressed the case with subdivided population structure, as well as the classic case with only a single population. Although we have focused on the two-locus model in this paper, we believe that much of our results can be generalized to multilocus models. In particular, the three-locus case can be solved exactly using the technique we have developed here. Cases with more than three loci seem more difficult, but our preliminary study looks promising. Our findings on multilocus models will be reported in a later paper.

We plan to use our proposal distributions in actual sequential IS schemes and compare their performance with using other proposal distributions. In the case of a single population we have shown that our conditional sampling distributions generally differ from that suggested by Fearnhead and Donnelly (2001). Furthermore, in the case of the infinitely-many-alleles model we have shown that our distributions are generally closer to the true distributions than are Fearnhead and Donnelly's. Given that the diffusion-generator approximation technique has been shown (see De Iorio and Griffiths (2004b)) to lead to significant improvements over previous IS methods, the theoretical work presented here may have much practical value. On a related note, it would be worthwhile and interesting to use our sampling distributions in the composite likelihood method (see Hudson (2001), McVean *et al.* (2002), (2004), and Myers *et al.* (2005)) and the PAC method (Li and Stephens (2003)).

Acknowledgements

References

BAHLO, M. AND GRIFFITHS, R. C. (2000). Inference from gene trees in a subdivided population. *Theoret. Pop. Biol.* **57,** 79–95.

BEAUMONT, M. (1999). Detecting population expansion and decline using microsatellites. *Genetics* **153,** 2013–2029.

CORNUET, J. M. AND BEAUMONT, M. A. (2007). A note on the accuracy of PAC-likelihood inference with microsatellite data. *Theoret. Pop. Biol.* **71,** 12–19.

DE IORIO, M. AND GRIFFITHS, R. C. (2004a). Importance sampling on coalescent histories. I. *Adv. Appl. Prob.* **36,** 417–433.

DE IORIO, M. AND GRIFFITHS, R. C. (2004b). Importance sampling on coalescent histories. II: subdivided population models. *Adv. Appl. Prob.* **36,** 434–454.

ETHIER, S. N. AND GRIFFITHS, R. C. (1990). On the two-locus sampling distribution. *J. Math. Biol.* **29,** 131–159.

FEARNHEAD, P. AND DONNELLY, P. (2001). Estimating recombination rates from population genetic data. *Genetics* **159,** 1299–1318.

FEARNHEAD, P. AND SMITH, N. G. C. (2005) A novel method with improved power to detect recombination hotspots from polymorphism data reveals multiple hotspots in human genes. *Amer. J. Human Genetics* **77,** 781–794.

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

GRIFFITHS, R. C. AND MARJORAM, P. (1996). Ancestral inference from samples of DNA sequences with recombination. *J. Comput. Biol.* **3,** 479–502.

GRIFFITHS, R. C. AND TAVARÉ, S. (1994a). Ancestral inference in population genetics. *Statist. Sci.* **9,** 307–319.

GRIFFITHS, R. C. AND TAVARÉ, S. (1994b). Sampling theory for neutral alleles in a varying environment. *Proc. R. Soc. London B* **344,** 403–410.

GRIFFITHS, R. C. AND TAVARÉ, S. (1994c). Simulating probability distributions in the coalescent. *Theoret. Pop. Biol.* **46,** 131–159.

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

KUHNER, M. K., YAMATO, J. AND FELSENSTEIN, J. (1995). Estimating effective population size and mutation rate from sequence data using Metropolis–Hastings sampling. *Genetics* **140,** 1421–1430.

KUHNER, M. K., YAMATO, J. AND FELSENSTEIN, J. (2000). Maximum likelihood estimation of recombination rates from population data. *Genetics* **156,** 1393–1401.

LI, N. AND STEPHENS, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165,** 2213–2233.

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. *et al.* (2004). The fine-scale structure of recombination rate variation in the human genome. *Science* **304,** 581–584.

MYERS, S. *et al.* (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310,** 321–324.

STEPHENS, M. AND DONNELLY, P. (2000). Inference in molecular population genetics. *J. R. Statist. Soc. Ser. B* **62,** 605–655.

WILSON, I. J. AND BALDING, D. J. (1998). Genealogical inference from microsatellite data. *Genetics* **150,** 499–510.