Mathematical Biology

An estimator for the recombination rate from a continuously observed diffusion of haplotype frequencies

Robert C. Griffiths[1] · Paul A. Jenkins[2,3,4]

Abstract

Recombination is a fundamental evolutionary force, but it is difficult to quantify because the effect of a recombination event on patterns of variation in a sample of genetic data can be hard to discern. Estimators for the recombination rate, which are usually based on the idea of integrating over the unobserved possible evolutionary histories of a sample, can therefore be noisy. Here we consider a related question: how would an estimator behave if the evolutionary history actually was observed? This would offer an upper bound on the performance of estimators used in practice. In this paper we derive an expression for the maximum likelihood estimator for the recombination rate based on a continuously observed, multi-locus, Wright–Fisher diffusion of haplotype frequencies, complementing existing work for an estimator of selection. We show that, contrary to selection, the estimator has unusual properties because the observed information matrix can explode in finite time whereupon the recombination parameter is learned without error. We also show that the recombination estimator is robust to the presence of selection in the sense that incorporating selection into the model leaves the estimator unchanged. We study the properties of the estimator by simulation and show that its distribution can be quite sensitive to the underlying mutation rates.

✉ Paul A. Jenkins
p.Jenkins@warwick.ac.uk

1 School of Mathematics, Monash University, 9 Rainforest Walk, Melbourne, VIC 3800, Australia

2 Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

3 Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

4 The Alan Turing Institute, British Library, London NW1 2DB, UK

 Springer

1 Introduction

Recombination is a fundamental evolutionary force which shuffles genetic variation along a chromosome and gives rise to new haplotypes not previously seen in a population. It is a major goal of population genetics to infer rates of recombination along the genome and to disentangle its effects from other evolutionary forces such as mutation, selection, migration, and genetic drift. However, the effects of recombination can be difficult to detect; generally the signal of recombination is weak and a single recombination event may leave no discernible trace in a sample of genetic data (Hayman et al. 2022). Typically one observes a sample from the state of the population only at the present day, while the evolutionary history of the population, which can be much more informative for recombination, is a latent, unobserved variable. A wide range of inferential methods tackle this problem by positing a generative reproductive model for the population and integrating over all possible evolutionary histories, or by approximating this idea. A popular model is the diffusion limit of a Cannings-type model for recombination, genetic drift, and mutation. Under this limit the evolution of haplotype frequencies follows the *Wright–Fisher diffusion with recombination* (Ohta and Kimura 1969a, b) while the genealogical history of a sample is known as the *ancestral recombination graph (ARG)* (Griffiths and Marjoram 1997). Reconstruction of ARGs is a major current endeavour (see Peñalba and Wolf 2020, for recent review), and with the very large samples available in recent datasets it becomes ever more necessary to introduce computational and/or model heuristics.

In this paper we address a related question: in the idealised situation in which one observes the entire evolutionary history of a population, as defined via the trajectory of haplotype frequencies in the diffusion limit, can we define an estimator for the recombination rate based on this observation and derive its properties? Although observing the entire sample path of a diffusion is unrealistic in practice, we may regard the corresponding estimator as setting an upper bound on the information about recombination available to us. We note that statistical inference from a continuously observed diffusion is by now a standard problem; see Kutoyants (2004) for textbook treatment for scalar diffusions (though regularity conditions imposed throughout that work preclude most of it applying to the Wright–Fisher diffusion even in one dimension). Further, advances in sequencing technologies are leading to growing availability of genetic data sampled from a population across different times, sometimes over very long timescales, and providing great potential for improved statistical inference (Dehasque et al. 2020); such datasets can be considered as discrete, noisy versions of the idealised setting studied in this paper.

A motivation for this work is Watterson (1979) who derived the maximum likelihood estimator $\hat{s}$ for natural selection from an observation of the trajectory of a Wright–Fisher diffusion (here a diallelic, one-locus model comprising only selection and genetic drift). He found the complete distribution of the estimator. It is worth noting that in this model $\hat{s}$ does not enjoy the usual desirable asymptotic properties such as consistency, since one of the alleles will almost surely go extinct in finite time and thus the total information available about the parameter up to time T remains finite as $T \to \infty$. If we introduce bidirectional recurrent mutation to the model then it becomes ergodic, and Sant et al. (2022) have recently shown that in this situation

the estimator enjoys the properties of consistency (uniformly over compact subsets of the parameter space) as well as asymptotic normality and asymptotic efficiency. We will see that, with or without mutation, the estimator for recombination behaves very differently to that of selection because the 'information' (defined formally below) can become infinite in finite time. Essentially, in a model of selection the signal-to-noise ratio for the selection parameter remains finite on hitting a boundary of the simplex of possible frequencies, while for the recombination parameter it may not. We will see that if the information becomes infinite then the maximum likelihood estimator (MLE) for recombination becomes exact, $\hat{\rho}_{\mathrm{MLE}} = \rho$.

The paper is structured as follows. In Sect. 2 we summarise likelihood theory for a continuously observed diffusion and specialise it to the infinitesimal variance of a Wright–Fisher diffusion. In Sect. 3 we derive the MLE for a general Wright–Fisher diffusion with arbitrary infinitesimal drift subject only to the constraint that the drift is linear in its unknown parameters. We then specialise this to the model of our primary interest, a multi-locus model with unknown recombination rate. Throughout we focus on the two-locus case which illustrates the main ideas without complicating the notation. Our main results are to derive an expression for the MLE and to show that if the information explodes then it is possible to learn the recombination parameter without error. Section 4 studies the impact of the presence of selection on this estimator, and in Sect. 5 we conduct a simulation study to investigate the empirical properties of the MLE. We discuss some potential directions for future work in Sect. 6.

2 Likelihood in diffusion paths

2.1 General case

We first give a summary of general likelihood-based inference for the parameters of a diffusion before specialising to the Wright–Fisher diffusion. Let $\{X(t) : t \geq 0\}$ be a d-dimensional diffusion process and suppose its path $\{X(t) : t \in [0, T]\}$ is observed up to time T. The generator of the diffusion has a form

$$\mathcal{L} = \frac{1}{2} \sum_{i,j=1}^{d} V_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} \mu_i(x; \varphi) \frac{\partial}{\partial x_i}, \tag{1}$$

where the model has r parameters $\varphi = (\varphi_1, \ldots, \varphi_r)^\top$ in a parameter space Θ. We assume the drift $\mu = (\mu_1 \ldots, \mu_d)^\top$ can be written in the form

$$\mu(x; \varphi) = c(x) + a(x; \varphi), \tag{2}$$

with $a(x; \varphi_0) \equiv 0$ for a fixed reference parameter φ_0, and $c(\cdot)$ does not contain any parameters to be estimated. (For example, later we will estimate the rate of recombination in the presence of recurrent mutation with the latter having rates fixed and known. Then $a(\cdot; \varphi)$ will correspond to the contribution of recombination while $c(\cdot)$

will correspond to the contribution of mutation, containing known mutation parameters.)

We will denote the corresponding path measure on continuous functions from $[0, T]$ to $\mathbb{R}^d$ by $\mathbb{P}_\varphi^{(T)}$.

If the $d \times d$ matrix $V = (V_{ij})$ is non-singular for almost all $t \in [0, T]$ and, for each $\varphi \in \Theta$,

$$\mathbb{P}_\varphi^{(T)}(I_{ij} < \infty, \, i, j = 1, \ldots, r) = 1, \tag{3}$$

where $I_T = (I_{ij})$ is the $r \times r$ *observed information matrix*

$$I_T = \int_0^T Z(X(t); \varphi)^\top V^{-1}(X(t)) Z(X(t); \varphi) \, dt, \qquad Z_{ij}(x; \varphi) = \frac{\partial a_i(x; \varphi)}{\partial \varphi_j}, \tag{4}$$

then the likelihood for φ takes the form of a Radon–Nikodym derivative

$$L_T(\varphi) = \frac{d\mathbb{P}_\varphi^{(T)}}{d\mathbb{P}_{\varphi_0}^{(T)}}$$

given with respect to a dominating measure which here we have chosen to be the model parametrised by φ_0 so that $\mathbb{P}_{\varphi_0}^{(T)}$ is the distribution over paths with drift c. Under these conditions, the likelihood takes the form

$$L_T(\varphi) = \exp\left(\int_0^T a(X(t); \varphi)^\top V(X(t))^{-1} \, d\widetilde{X}(t) \right.$$
$$\left. -\frac{1}{2} \int_0^T a(X(t); \varphi)^\top V(X(t))^{-1} a(X(t); \varphi) \, dt \right), \tag{5}$$

where

$$\widetilde{X}(t) = X(t) - \int_0^t c(X(s)) \, ds. \tag{6}$$

The first integral in (5) is with respect to the path $\{\widetilde{X}(t) : t \in [0, T]\}$ and the second is with respect to t.

A heuristic way to understand Eq. (5) is as follows. Let $\Delta X(t) = X(t + \Delta t) - X(t)$. The distribution of $\Delta X(t)$ given $X(t) = x$ is taken as approximately normal with mean $\mu(x)\Delta t$ and covariance matrix $V(x)\Delta t$ as $\Delta t \to 0$. If $V(x)$ is non-singular, then the quadratic form in the exponent of the normal density of $\Delta X(t)$ is

$$\left[\Delta X(t) - \mu(x)\Delta t\right]^\top \left[\Delta t V(X)\right]^{-1} \left[\Delta X(t) - \mu(x)\Delta t\right] =$$
$$\mu(x)^\top V(x)^{-1} \mu(x)\Delta t - 2\mu(x)^\top V(x)^{-1} \Delta X(t) + \mathcal{O}((\Delta t)^2).$$

We are expressing this density with respect to another normal density with mean $c(x)\Delta t$ and covariance matrix $V(x)\Delta t$, and thus we subtract the corresponding

quadratic form

$$c(x)^\top V(x)^{-1} c(x) \Delta t - 2c(x)^\top V(x)^{-1} \Delta X(t) + \mathcal{O}((\Delta t)^2).$$

After some rearrangement, letting $\Delta t \to 0$, and integrating from 0 to T, we recover the quadratic form appearing in (5). Note that the likelihood ignores the terms $|V(x)|$ in the diffusion, since we assume that the likelihood is with respect to a parametric form only for μ. (Statistical inference for parameters of V would be trivial in this setting, since V is identifiable from the path via its quadratic variation.) See Basawa and Prakasa Rao (1980, Ch. 9) and Kloeden et al. (2003, Ch. 6) for further details on the general case.

2.2 Wright–Fisher diffusion

The family of Wright–Fisher diffusions has generator (1) with diffusion coefficient of the form

$$V_{ij}(x) = x_i(\delta_{ij} - x_j),$$

where δ_{ij} denotes the Kronecker delta (i.e. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$). The diffusion takes values in the simplex

$$\Delta_{d-1} := \left\{ x \in [0, 1]^d : \sum_{i=1}^d x_i = 1 \right\},$$

and the domain of $\mathcal{L}$ is $\mathcal{D}(\mathcal{L}) = C^2(\Delta_{d-1})$, twice continuously differentiable functions with domain Δ_{d-1}. For now we continue to leave the drift in the form (2) but otherwise unspecified.

The matrix $V(x)$ is singular since $\sum_{i=1}^d x_i = 1$. Our first task, then, is to modify the results from Sect. 2.1 to accommodate this issue. We achieve this by studying the first $d - 1$ coordinates of X, whose infinitesimal covariance matrix $V^*(x)$ is non-singular. Fortunately, its inverse $V^*(x)^{-1}$ takes on a particularly simple form, as we now show.

Theorem 1 *Assume (3) holds for a Wright–Fisher diffusion with drift coefficient $\mu(x; \varphi) = c(x) + a(x; \varphi)$ and diffusion coefficient $V = (V_{ij})$, $V_{ij}(x) = x_i(\delta_{ij} - x_j)$. Then the likelihood is*

$$L_T(\varphi) = \exp\left(\int_0^T \sum_{i=1}^d \frac{a_i(X(t); \varphi)}{X_i(t)} \, d\widetilde{X}_i(t) - \frac{1}{2} \int_0^T \sum_{i=1}^d \frac{a_i(X(t); \varphi)^2}{X_i(t)} \, dt \right), \quad (7)$$

with $\widetilde{X}$ given by (6).

Proof We consider the diffusion $(X_1(t), \ldots, X_{d-1}(t))$ with drift $\mu^*(x; \varphi) = (\mu_1(x; \varphi), \ldots, \mu_{d-1}(x; \varphi))^\top$ and non-singular $(d-1) \times (d-1)$ covariance matrix

$V^*(x)$. Define $X_d(t) = 1 - \sum_{i=1}^{d-1} X_i(t)$ and $\mu_d(x; \varphi) = -\sum_{i=1}^{d} \mu_i(x; \varphi)$. It follows from standard normal theory, for example Kendall et al. (1994, p520–521), that

$$[V^*(x)^{-1}]_{ij} = \left(x_d^{-1} + x_i^{-1}\delta_{ij}\right). \tag{8}$$

We know that $\sum_{i=1}^{d} a_i(x; \varphi) = 0$ and $\sum_{i=1}^{d} d\widetilde{X}_i(t) = 0$ (since both $\sum_{i=1}^{d} dX_i(t) = 0$ and $\sum_{i=1}^{d} c_i(X(t)) = 0$, the latter required for X to take values in Δ_{d-1} when $\varphi = \varphi_0$), so

$$
\begin{aligned}
a^*(X(t); \varphi)^\top V^*(X(t))^{-1} d\widetilde{X}(t) &= \sum_{i=1}^{d-1}\sum_{j=1}^{d-1} a_i(X(t); \varphi)(X_d(t)^{-1} + \delta_{ij}X_i(t)^{-1}) d\widetilde{X}_j(t) \\
&= \frac{a_d(X(t); \varphi)}{X_d(t)} d\widetilde{X}_d(t) + \sum_{i=1}^{d-1} \frac{a_i(X(t); \varphi)}{X_i(t)} d\widetilde{X}_i(t) \\
&= \sum_{i=1}^{d} \frac{a_i(X(t); \varphi)}{X_i(t)} d\widetilde{X}_i(t),
\end{aligned}
\tag{9}
$$

with a similar calculation for

$$
\begin{aligned}
a^*(X(t); \varphi)^\top V^*(X(t))^{-1} a^*(X(t); \varphi) &= \sum_{i=1}^{d-1}\sum_{j=1}^{d-1} a_i(X(t); \varphi)a_j(X(t); \varphi)V_{ij}^*(X(t))^{-1} \\
&= \sum_{i=1}^{d} \frac{a_i(X(t); \varphi)^2}{X_i(t)}.
\end{aligned}
\tag{10}
$$

Substituting (9) and (10) into (5) yields (7).   □

3 Theory for maximum likelihood estimators

3.1 General Wright–Fisher diffusion

Our next goal is to derive an MLE for the parameters φ of a Wright–Fisher diffusion. This is found by differentiating the log-likelihood with respect to the parameters. In all the examples we encounter, the drift is a linear function of the parameters so for the remainder of this article we assume $a(x; \varphi)$ to be of the form

$$a_i(x; \varphi) = \sum_{k=1}^{r} Z_{ik}(x)\varphi_k, \tag{11}$$

where $Z_{ik}(x) = \frac{\partial a_i(x;\varphi)}{\partial \varphi_k}$ does not depend on φ. To avoid issues of identifiability we suppose that the columns of $Z = (Z_{ij})$ are linearly independent functions. Then from

Theorem 1 the log-likelihood is a quadratic function

$$\log L_T(\varphi) = \sum_{k=1}^{r} \varphi_k \int_0^T \sum_{i=1}^{d} \frac{Z_{ik}(X(t))}{X_i(t)} \, d\widetilde{X}_i(t)$$

$$-\frac{1}{2} \sum_{k=1}^{r} \sum_{l=1}^{r} \varphi_k \varphi_l \int_0^T \sum_{i=1}^{d} \frac{Z_{ik}(X(t)) Z_{il}(X(t))}{X_i(t)} \, dt,$$

with a unique maximum, $\hat{\varphi}$, in $\mathbb{R}^r$, which is the solution of the set of equations for $k = 1, \ldots, r$:

$$0 = \int_0^T \sum_{i=1}^{d} \frac{Z_{ik}(X(t))}{X_i(t)} \, d\widetilde{X}_i(t) - \sum_{l=1}^{r} \varphi_l \int_0^T \sum_{i=1}^{d} \frac{Z_{ik}(X(t)) Z_{il}(X(t))}{X_i(t)} \, dt. \quad (12)$$

The set of Eq. (12) are familiar in regression theory. Now denote the $(r \times 1)$ vector $Y = (Y_k)$ with elements

$$Y_k = \int_0^T \sum_{i=1}^{d} \frac{Z_{ik}(X(t))}{X_i(t)} \, d\widetilde{X}_i(t),$$

and let $\Sigma(X(t)) = \text{diag}(X_i(t))$. Then the Eq. (12) can be written

$$\left[\int_0^T Z(X(t))^\top \Sigma^{-1}(X(t)) Z(X(t)) \, dt \right] \hat{\varphi} = Y.$$

Continuing to assume (3), the matrix on the left-hand side of the previous equation is non-singular (see Basawa and Prakasa Rao (1980, p223–224), Kloeden et al. (2003, p231)) and hence we arrive at the form

$$\hat{\varphi} = \left[\int_0^T Z(X(t))^\top \Sigma^{-1}(X(t)) Z(X(t)) \, dt \right]^{-1} Y. \quad (13)$$

Of course if $\Theta \subset \mathbb{R}^r$ then it is not guaranteed that $\hat{\varphi} \in \Theta$, and $\hat{\varphi}$ must be adjusted appropriately to ensure it is the MLE. An example of this adjustment is given later.

The observed information matrix (4) is a key quantity in telling us about how informative the data is for φ. For this model, the observed information matrix I_T has elements

$$I_{kl} = \int_0^T \sum_{i=1}^{d} \frac{1}{X_i(t)} Z_{ik}(X(t)) Z_{il}(X(t)) \, dt \quad (14)$$

using linearity of $a_i(x; \varphi)$; the information matrix does not depend on $\hat{\varphi}$. The expression (13) for $\hat{\varphi}$ can be written

$$\hat{\varphi} = I_T^{-1} Y.$$

3.1.1 Deterministic model

As a check on the expression for $\hat{\varphi}$, we can ask for the estimator we would obtain if the observed trajectory is that of the deterministic model

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \mu_i(x; \varphi), \qquad i = 1, \ldots, d. \tag{15}$$

Now from (2) and (6) we find

$$\mathrm{d}\widetilde{x} = \mathrm{d}x - c(x)\,\mathrm{d}t = (c(x) + a(x; \varphi))\,\mathrm{d}t - c(x)\,\mathrm{d}t = a(x; \varphi)\,\mathrm{d}t,$$

and so we can substitute this expression for $\mathrm{d}\widetilde{x}$ into the likelihood Eq. (12) to obtain that $\hat{\varphi}$ is a solution to

$$0 = \int_0^T \sum_{i=1}^d \frac{[a_i(x(t); \varphi) - a_i(x(t); \hat{\varphi})]}{x_i(t)} \frac{\partial a_i(x(t); \hat{\varphi})}{\partial \varphi_l}\,\mathrm{d}t. \tag{16}$$

Owing to the factor $a_i(x(t); \varphi) - a_i(x(t); \hat{\varphi})$, it is clear that a solution to the likelihood equation is given by $\hat{\varphi} = \varphi$. It is reassuring that the estimator is well behaved even in this crude level of approximation; the trajectory defined by (15) is not a realisation from the assumed model since it is a path of bounded variation.

3.2 Neutral two-locus model

We now turn to our main result, an expression for the MLE for the recombination parameter $\rho \in [0, \infty) =: \Theta$. Consider a neutral two-locus model in which there are K possible alleles at the first locus, locus A, and L possible alleles at the second, locus B. The haplotype of an individual is denoted $(i, j) \in \{1, \ldots, K\} \times \{1, \ldots, L\}$, and its frequency in the population is x_{ij}. Note that to reconcile this double-index notation with previous sections we must implicitly stack the KL possible haplotypes in some agreed order into a vector of length $d = KL$. We will switch between the two notations as required. To emphasise when haplotypes have been stacked we will use a bold index, so x_i denotes the frequency of haplotype i, $i = 1, \ldots, d$.

The model is completed by specifying the drift. Here it is of the form

$$a_{ij}(x; \rho) = \rho(x_{i\cdot} x_{\cdot j} - x_{ij}), \qquad i = 1, \ldots, K;\ j = 1, \ldots, L,$$

where $x_{i\cdot} := \sum_{l=1}^L x_{il}$ and $x_{\cdot j} := \sum_{k=1}^K x_{kj}$. Recombination occurs between the two loci at rate ρ; specifically this is a model of the homologous crossing-over that takes

place during meiosis. To simplify later results, we omit the conventional factor of $1/2$ in the recombination rate parameter.

In much of what follows the choice for $c(x)$ is immaterial, but for concreteness we will set

$$c_{ij}(x) = \frac{\theta_A}{2} \sum_{k=1}^{K} x_{kj}(P_{ki}^A - \delta_{ik}) + \frac{\theta_B}{2} \sum_{l=1}^{L} x_{il}(P_{lj}^B - \delta_{jl}).$$

Here mutation takes place at locus A and B at respective rates $\theta_A/2$ and $\theta_B/2$ on the timescale of the diffusion. When a mutation occurs, the change in allele is governed by the $K \times K$ and $L \times L$ mutation transition matrices P^A and P^B (i.e. if a mutation occurs at locus A on haplotype (k, j) then it mutates to haplotype (i, j) with probability P_{ki}^A, $i = 1, \ldots, K$; similarly for P^B). We allow $\theta_A, \theta_B \geq 0$, so the model may or may not be ergodic.

Note the separate roles for the two components of the drift: here it is only ρ to be estimated, with the other parameters appearing in $c(\cdot)$ considered known. The likelihood is expressed with respect to the parametrisation $\rho_0 = 0$, a model in which the two loci are completely linked but the mutation parameters are the same.

Using the results of Sect. 3.1, for this model the log-likelihood is

$$\log L_T(\rho) = \rho \int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \left(\frac{X_{i \cdot}(t) X_{\cdot j}(t)}{X_{ij}(t)} - 1 \right) d\widetilde{X}_{ij}(t)$$

$$- \frac{1}{2} \rho^2 \int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{(X_{ij}(t) - X_{i \cdot}(t) X_{\cdot j}(t))^2}{X_{ij}(t)} dt$$

$$= \rho \int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{X_{i \cdot}(t) X_{\cdot j}(t)}{X_{ij}(t)} d\widetilde{X}_{ij}(t)$$

$$- \frac{1}{2} \rho^2 \int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{(X_{ij}(t) - X_{i \cdot}(t) X_{\cdot j}(t))^2}{X_{ij}(t)} dt, \qquad (17)$$

where for the second equality we recall $\sum_{i=1}^{K} \sum_{j=1}^{L} d\widetilde{X}_{ij}(t) = 0$, with

$$\widetilde{X}_{ij}(t) = X_{ij}(t) - \int_0^t c_{ij}(X(s)) \, ds, \qquad i = 1, \ldots, K; \; j = 1, \ldots, L.$$

The estimator $\hat{\rho}$ is therefore

$$\hat{\rho} = \frac{\displaystyle\int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{X_{i \cdot}(t) X_{\cdot j}(t)}{X_{ij}(t)} d\widetilde{X}_{ij}(t)}{\displaystyle\int_0^T \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{(X_{ij}(t) - X_{i \cdot}(t) X_{\cdot j}(t))^2}{X_{ij}(t)} dt}, \qquad (18)$$

and the observed information is

$$I_T = \int_0^T \sum_{i=1}^K \sum_{j=1}^L \frac{(X_{ij}(t) - X_{i\cdot}(t)X_{\cdot j}(t))^2}{X_{ij}(t)} \, dt. \tag{19}$$

The denominator in $\hat{\rho}$ and the information can be simplified to

$$I_T = \int_0^T \sum_{i=1}^K \sum_{j=1}^L \frac{X_{i\cdot}(t)^2 X_{\cdot j}(t)^2}{X_{ij}(t)} \, dt - T.$$

It is worth remarking on the functional form (17) for $\log L_T(\rho)$. This is a polynomial in ρ and we can think of a trade-off between the order of the polynomial and the complexity of its coefficients. In this model we have a particularly simple quadratic polynomial in ρ, order only two, with the benefit of knowing that the function is convex with a unique finite maximum (since the coefficient of ρ^2 is negative). The price we pay is that the coefficients of the polynomial are highly cumbersome in the sense that they are given as integrals over the sample path of a diffusion. Contrast this with the dual coalescent model in which the likelihood for an observed sample path of an ARG would be a product of exponential waiting time densities times a product of rational functions for the transitions of the jump chain. With many possible jumps, these rational functions may be constructed from polynomials in ρ of very high order, though their coefficients are much simpler than the stochastic integrals encountered here. In a coalescent model, the shape of the likelihood curve as a function of ρ can be rather complicated, even exhibiting local minima when integrating over ARGs (Jenkins and Song 2009).

3.2.1 Deterministic model

Is $\hat{\rho}$ in (18) a reasonable estimate? Again we can check what happens when X solves a deterministic model. Setting $\theta_A = \theta_B = 0$ for the moment, the deterministic model is

$$\frac{dx_{ij}}{dt} = \rho(x_{i\cdot}x_{\cdot j} - x_{ij}), \qquad i = 1, \ldots, K; \; j = 1, \ldots, L. \tag{20}$$

and substituting dx_{ij} directly into (18) again shows that $\hat{\rho} = \rho$.

We can further describe the evolution of I_T. Note that in this deterministic model (summing (20) over j):

$$\frac{dx_{i\cdot}}{dt} = 0, \qquad i = 1, \ldots, K.$$

Therefore $x_{i\cdot}(t) = x_{i\cdot}(0)$ for all $t \geq 0$, and similarly for $x_{\cdot j}(t)$. The solution to (20) is then

$$x_{ij}(t) = x_{ij}(0)e^{-\rho t} + x_{i\cdot}(0)x_{\cdot j}(0)(1 - e^{-\rho t}). \tag{21}$$

We have that

$$\log\left(\frac{x_{ij}(T)}{x_{ij}(0)}\right) = \int_0^T \frac{\mathrm{d}x_{ij}}{x_{ij}} = \rho\left(\int_0^T \frac{x_{i.}x_{.j}}{x_{ij}}\,\mathrm{d}t - T\right),$$

so (provided $\rho > 0$):

$$I_T = \rho^{-1}\sum_{i=1}^K\sum_{j=1}^L x_{i.}(0)x_{.j}(0)\log\left(\frac{x_{ij}(T)}{x_{ij}(0)}\right). \qquad (22)$$

The limit information is therefore

$$\lim_{T\to\infty} I_T = \rho^{-1}\sum_{i=1}^K\sum_{j=1}^L x_{i.}(0)x_{.j}(0)\log\left(\frac{x_{i.}(0)x_{.j}(0)}{x_{ij}(0)}\right).$$

As far as the deterministic model goes, the information is in the transient phase until the frequencies come to equilibrium. The accumulated information I_T remains finite as $T \to \infty$. In a stochastic model on the other hand, we will see that the injection of noise allows $I_T \to \infty$ as $T \to \infty$. We note that one should regard this contrasting behaviour with caution: it does *not* mean that the estimator is consistent only in the stochastic setting. We have just seen that $\hat{\rho} = \rho$ in the deterministic setting, which is trivially consistent, and creates a paradox when we try to reconcile this fact with the asymptotic finiteness of I_T. The paradox is resolved by noting that the data-generating mechanism differs from the one assumed in designing the estimator. Had we assumed a deterministic model throughout our analysis then, since the parameter is simply a rate appearing in an observed ODE, the 'likelihood' would be a point mass on the true rate and the MLE would be equal to that true rate. The 'information' in this setting, being the curvature of the log-likelihood, is immediately infinite. The fact that $\hat{\rho} = \rho$ demonstrates that the estimator adapts automatically to a change in data-generating mechanism. The quantity I_T could be regarded not as the information under the true model but as a way of quantifying 'the informativeness of the deterministic trajectory under stochastic assumptions'. It is this quantity that remains finite as $T \to \infty$.

It is possible to repeat these calculations for a model with $\theta_A, \theta_B > 0$; that is, to solve the deterministic mutation-recombination equation. Again I_T converges to a finite limit; see Appendix A.

3.2.2 Stochastic differential equation interpretation

We can find an expression for the error associated with $\hat{\rho}$ by regarding $X(t)$ as the solution to a stochastic differential equation (SDE):

$$\mathrm{d}X(t) = [c(X(t)) + a(X(t); \varphi)]\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t), \qquad X(0) = x(0), \quad (23)$$

where W is a $(d-1)$-dimensional Brownian motion and $\sigma(x)$ is a (non-unique) $d \times (d-1)$ matrix satisfying $\sigma(x)\sigma(x)^\top = V(x)$.

There are various ways to define σ subject to this constraint. It is common to ask for σ to be lower triangular by applying a Cholesky decomposition to V. In the case of the covariance matrix of the Wright–Fisher diffusion, the Cholesky decomposition is given analytically by Sato (1976), though it is one that explodes at the boundaries.

Proposition 1 *The error associated with $\hat{\rho}$ is*

$$
\hat{\rho} - \rho = -\frac{\int_0^T \sum_{i=1}^d \frac{D_i(t)}{X_i(t)} \sum_{j=1}^{d-1} \sigma_{ij}(X(t)) \, dW_j(t)}{\int_0^T \sum_{i=1}^d \frac{D_i(t)^2}{X_i(t)} \, dt},
\tag{24}
$$

where $D_i(t) = X_{i_1 i_2}(t) - X_{i_1 \cdot}(t) X_{\cdot i_2}(t)$ is the coefficient of linkage disequilibrium for haplotype $i = (i_1, i_2)$.

Proof Rearranging (18) slightly and using $\sum_{i=1}^d d\widetilde{X}_i(t) = 0$, we have

$$
\hat{\rho} I_T = -\int_0^T \sum_{i=1}^d \frac{D_i(t)}{X_i(t)} \, d\widetilde{X}_i(t).
$$

Now substituting for $d\widetilde{X}_i(t) = dX_i(t) - c_i(X(t)) \, dt$ using (23),

$$
\hat{\rho} I_T = \rho I_T - \int_0^T \sum_{i=1}^d \frac{D_i(t)}{X_i(t)} \sum_{j=1}^{d-1} \sigma_{ij}(X(t)) \, dW_j(t),
$$

which leads to (24).   □

Thus the bias and mean squared error of $\hat{\rho}$ are given respectively by the expectation of the term on the right-hand side of (24) and the expectation of its square. Estimators of this form are not unbiased in general (Basawa and Prakasa Rao 1980, p 218).

3.2.3 Corrected MLE

There are two problems with the estimator $\hat{\rho}$ defined in (18). First, the parameter space is $\Theta = [0, \infty)$ but we cannot ensure $\hat{\rho} \geq 0$. (Although $\rho < 0$ is biologically unrealistic, mathematically it is nonetheless a valid model and a sample path may point to this region of the parameter space if $d\widetilde{X}_{ij}(t)$ is sufficiently negative.) This is easily corrected by applying a rectified linear unit, $\max\{0, \hat{\rho}\}$. The second issue is more serious: it is not guaranteed that (3) holds. In other words, we have not ruled out the possibility that I_T explodes in finite time. For observations for which $I_T < \infty$, we can still interpret (17) as a quasi-log-likelihood function (Kloeden et al. 2003, p231), but otherwise we must treat $L_T(\rho)$ as a *generalized* density valid only until the stopping time

$$
S := \inf \{t \in [0, \infty) : I_t = \infty\}.
\tag{25}
$$

See Liptser and Shiryaev (2001, Ch. 6) and Mijatović et al. (2012) for further discussion on this subtle point. Writing $\hat{\rho} = \hat{\rho}_T$ for the estimator in (18), we define the following corrected estimator:

$$\hat{\rho}_{\text{MLE}} := \mathbb{I}_{[0,S)}(T) \max \{0, \hat{\rho}_T\} + \mathbb{I}_{[S,\infty)}(T) \lim_{t \uparrow S} \hat{\rho}_t. \tag{26}$$

A similar issue arises in the estimation of the immigration rate of the continuous branching with immigration (CBI) diffusion, where a related correction is proposed (Overbeck 1998, in particular Theorem 2(iv)). The subscript in (26) rather suggestively posits this quantity as *the* MLE; this is proven shortly, in Corollary 1. Although taking $\lim_{t \uparrow S} \hat{\rho}_t$ in (26) might seem to be unstable, that this is the appropriate correction to our estimator is justified by the following theorem.

Theorem 2 *If $S \le T$ then $\hat{\rho}_{MLE} = \rho$ with probability 1.*

Proof From the definition (26) of $\hat{\rho}_{\text{MLE}}$ it suffices to show that $\hat{\rho} \to \rho$ as $T \uparrow S$. Let

$$N_T = -\int_0^{T \wedge S} \sum_{i=1}^{d} \frac{D_i(t)}{X_i(t)} \sum_{j=1}^{d-1} \sigma_{ij}(X(t)) \, dW_j(t).$$

This is a continuous, stopped martingale with $N_0 = 0$ and quadratic variation

$$\begin{aligned}
\langle N \rangle_T &= \left\langle -\int_0^{T \wedge S} \sum_{i=1}^{d} \frac{D_i(t)}{X_i(t)} \sum_{j=1}^{d-1} \sigma_{ij}(X(t)) \, dW_j(t) \right\rangle_T \\
&= \int_0^{T \wedge S} \sum_{i=1}^{d} \sum_{j=1}^{d-1} \frac{D_i(t)}{X_i(t)} \sigma_{ij}(X(t)) \sum_{k=1}^{d} \frac{D_k(t)}{X_k(t)} \sigma_{kj}(X(t)) \, dt \\
&= \int_0^{T \wedge S} \sum_{i=1}^{d} \sum_{k=1}^{d} \frac{D_i(t)}{X_i(t)} \frac{D_k(t)}{X_k(t)} V_{ik}(X(t)) \, dt \\
&= \int_0^{T \wedge S} \sum_{i=1}^{d} \frac{D_i(t)^2}{X_i(t)} \, dt \\
&= I_{T \wedge S}.
\end{aligned}$$

where the second equality uses that $\langle \cdot, \cdot \rangle$ is a bilinear form and $\langle dW_j, dW_l \rangle = \delta_{jl} \, dt$. Thus by the law of large numbers for local martingales (Revuz and Yor 1999, Ch. V.1, Exercise 1.16, p186),

$$\lim_{T \to \infty} \frac{N_T}{I_{T \wedge S}} = 0 \quad \text{with probability 1 on } \{I_\infty = \infty\}.$$

The limit as $T \uparrow S$ is the same. But $N_T / I_{T \wedge S}$ is precisely the error $\hat{\rho} - \rho$ given in Proposition 1, so $\hat{\rho} \to \rho$ as $T \uparrow S$ with probability 1.   □

Corollary 1 $\hat{\rho}_{MLE}$ is the MLE for ρ.

Proof This follows since we have separately verified that it is the MLE on $\{T < S\}$ and on $\{S \leq T\}$. In the latter case ρ is identifiable, so $L_T(\rho)$ is zero anywhere other than the true value. □

Corollary 2 The error associated with $\hat{\rho}_{MLE}$ is

$$\hat{\rho}_{MLE} - \rho = -\mathbb{I}_{[0,S)}(T) \times \begin{cases} \dfrac{\displaystyle\int_0^T \sum_{i=1}^d \frac{D_i(t)}{X_i(t)} \sum_{j=1}^{d-1} \sigma_{ij}(X(t)) \, \mathrm{d}W_j(t)}{\displaystyle\int_0^T \sum_{i=1}^d \frac{D_i(t)^2}{X_i(t)} \, \mathrm{d}t}, & \hat{\rho} \geq 0, \\[4ex] \rho, & \hat{\rho} < 0, \end{cases}$$

where we recall that $\hat{\rho}$ is the uncorrected estimator given in (18).

Proof This follows by combining Theorem 2 and Proposition 1. □

The relevance of Theorem 2 is: If the sample path is such that $I_T = \infty$, then we learn ρ without error. Inspecting the form of I_T in (19), we see that its integrand is locally integrable in the interior of Δ_{d-1}. Thus for $I_T = \infty$ it is necessary to have at least one haplotype frequency $X_{ij}(t) \to 0$ before time T. We should expect the same phenomenon when inferring the *mutation* parameters in a one-locus model, where hitting one of the boundaries is completely informative for one of the mutation parameters.

The next result shows that having $I_T = \infty$ is not a hypothetical concern, and the proof makes it clear that explosion of I_T is intimately related with hitting a boundary of Δ_{d-1}.

Theorem 3 Suppose that $\rho + \frac{\theta_A}{2} + \frac{\theta_B}{2} < \frac{1}{2}$, that mutation is parent-independent (i.e. P^A and P^B each have identical rows), and that $x(0)$ lies in the interior of Δ_{d-1}. Then $\mathbb{P}(I_T = \infty) > 0$.

Proof It is clear from the form of I_T in (19) that $\{I_T = \infty\}$ will occur if for some i, j,

(i) For some $\delta > 0$ and for all $t \in [0, T]$, $X(t)$ lies in $A_1 := \{x \in \Delta_{d-1} : (x_{ij} - x_i.x_{.j})^2 > \delta\}$;

(ii) $T_\varepsilon(X_{ij}) := \inf\{t \in [0, \infty) : X_{ij}(t) = \varepsilon\}$, the first hitting time of ε by X_{ij}, satisfies $T_0(X_{ij}) \in (0, T]$; and

(iii) The integral

$$\int_0^{T_\varepsilon(X_{ij})} \frac{1}{X_{ij}(t)} \, \mathrm{d}t$$

diverges as $\varepsilon \to 0$.

Condition (iii) extracts the explosion of I_T from a denominator of its integrand, while condition (i) controls the corresponding numerator. Condition (ii) ensures that such explosion takes place before time T.

To study the finiteness or otherwise of the integral in (iii), choose a decomposition $\sigma(x)\sigma(x)^\top = V(x)$ so that the component of the SDE (23) corresponding to $X_{ij}(t)$ has the form

$$dX_{ij}(t) = \mu_{ij}(X_{ij}(t))\,dt + \sqrt{X_{ij}(t)(1 - X_{ij}(t))}\,dW(t), \qquad X_{ij}(0) = x_{ij}(0),$$

for a scalar Brownian motion W, where

$$\mu_{ij}(X_{ij}(t)) = \rho[X_{i\cdot}(t)X_{\cdot j}(t) - X_{ij}(t)] + \frac{\theta_A}{2}[X_{\cdot j}(t)P_i^A - X_{ij}(t)]$$
$$+ \frac{\theta_B}{2}[X_{i\cdot}(t)P_j^B - X_{ij}(t)].$$

The idea is to show that this SDE behaves locally like a one-locus model of *mutation only*. More precisely we will compare X_{ij} to another diffusion which solves the SDE

$$dZ(t) = \frac{\vartheta}{2}[P - Z(t)]\,dt + \sqrt{Z(t)(1 - Z(t))}\,dW(t), \qquad Z(0) = x_{ij}(0),$$

for some $\vartheta \in [0, 1)$, $P \in (0, 1)$. Choose ϑ so that $\rho + \frac{\theta_A}{2} + \frac{\theta_B}{2} < \frac{\vartheta}{2}$ and choose P so that $x_{i\cdot}(0)x_{\cdot j}(0) < P$, $P_i^A < P$, and $P_j^B < P$. Then on the set $A_2 := \{x \in \Delta_{d-1} : x_{i\cdot}x_{\cdot j} < P\}$ it is straightforward to verify we have

$$\mu_{ij}(x) < \frac{\vartheta}{2}(P - x),$$

and thus by a standard comparison theorem (see Theorem 1.1 and Remark 1.1 in Ikeda and Watanabe, 1977), we can construct a probability space on which $Z(t) \geq X_{ij}(t)$ for all $t \in [0, T_{A_2^\complement})$, where $T_A := \inf\{t \in [0, \infty) : X(t) \in A\}$. (For the comparison theorem to hold there is a required growth condition on the diffusion coefficient. That this holds follows from the fact that $\sqrt{x(1-x)}$ is 1/2-Hölder continuous; see also Remark 3.9 on p298 of Ethier and Kurtz (1986).) Thus condition (iii) is implied by the a.s. divergence of

$$\int_0^{T_\varepsilon(Z)} \frac{1}{Z(t)}\,dt$$

as $\varepsilon \to 0$, which in turn follows from Lemma 4.4 of Barton et al. (2004), noting that $\vartheta < 1$ guarantees the 0-boundary for Z is accessible. [Some errors in the proof of Lemma 4.4 are corrected by Taylor (2007).] Tracing our steps backwards, we have shown that condition (iii) holds provided $0 < T_0(X_{ij}) \leq T < T_{A_2^\complement}$. Since

$$\mathbb{P}(T_{A_1^\complement} > T,\, 0 < T_0(X_{ij}) \leq T < T_{A_2^\complement}) > 0,$$

we conclude $\mathbb{P}(I_T = \infty) > 0$.   □

The conditions given in Theorem 3 simplify our proof, but it seems feasible to substantially weaken them.

3.3 Testing for the presence of recombination

It is possible to use $\hat{\rho}_{\mathrm{MLE}}$ to design a likelihood ratio test for the null hypothesis that $\rho_0 = 0$. Using (17), the appropriate likelihood ratio statistic is, for $I_T < \infty$,

$$\Lambda := 2\log \frac{\mathrm{d}\mathbb{P}^{(T)}_{\hat{\rho}_{\mathrm{MLE}}}}{\mathrm{d}\mathbb{P}^{(T)}_{\rho_0}} = \hat{\rho}^2_{\mathrm{MLE}} I_T, \qquad I_T < \infty.$$

Under standard assumptions, noting that $\rho_0 = 0$ lies on the boundary of Θ, this has an asymptotic distribution which is an equal mixture between a χ^2_1 distribution and a χ^2_0 distribution under the null hypothesis (Self and Liang 1987). Denote the CDF of this distribution by F_m. In particular, to construct a level 5% test one should reject $\rho_0 = 0$ if Λ exceeds the 95th percentile of F_m; equivalently if it exceeds the 90th percentile of a χ^2_1 distribution.

To account for the possibility that $I_T = \infty$ we set

$$\Lambda := \begin{cases} +\infty, & \hat{\rho}_{\mathrm{MLE}} > 0, \\ 0, & \hat{\rho}_{\mathrm{MLE}} = 0. \end{cases}, \qquad I_T = \infty.$$

The asymptotic null distribution for Λ is now less clear, though we note that continuing to assume F_m would be conservative. We study the power of this test empirically in Sect. 5.

3.4 Multiple loci

It is possible to extend the above results to a general multi-locus model. The extension is straightforward and we omit many of the lengthy but straightforward calculations.

In a multi-locus model of ℓ loci with K_j possible alleles at locus j, haplotypes are of the form $\boldsymbol{i} = (i_1, \ldots, i_\ell) \in \prod_{j=1}^{\ell}\{1, \ldots, K_j\} =: E$ in a diffusion on Δ_{d-1} with $d = \prod_{j=1}^{\ell} K_j$ coordinates. Stacking the haplotypes, the diffusion coefficient has entries $V_{\boldsymbol{ik}}(x) = x_{\boldsymbol{i}}(\delta_{\boldsymbol{ik}} - x_{\boldsymbol{k}})$ as usual, and the unknown component of the drift is

$$a_{\boldsymbol{i}}(x; \rho_1, \ldots, \rho_{\ell-1}) = \sum_{j=1}^{\ell-1} \rho_j (x_{\boldsymbol{i}_{\le j}} x_{\boldsymbol{i}_{>j}} - x_{\boldsymbol{i}}),$$

where ρ_j is the recombination rate between locus j and $j + 1$, with each ρ_j to be estimated; $x_{\boldsymbol{i}}$ is the frequency of haplotype $\boldsymbol{i}$; and we marginalize over a contiguous

subset of loci by writing

$$x_{i_{\leq j}} = \sum_{i_{j+1}=1}^{K_{j+1}} \cdots \sum_{i_\ell=1}^{K_\ell} x_{(i_1,\ldots,i_\ell)}, \qquad x_{i_{>j}} = \sum_{i_1=1}^{K_1} \cdots \sum_{i_j=1}^{K_j} x_{(i_1,\ldots,i_\ell)}.$$

From (13) and (14) the joint estimator for $(\rho_1, \ldots, \rho_{\ell-1})$ is $\hat{\varrho} = I_T^{-1} Y$ where I_T is $(\ell-1) \times (\ell-1)$ and Y is $(\ell-1) \times 1$ with elements

$$
\begin{aligned}
I_{jk} &= \int_0^T \sum_{i \in E} \frac{\left(X_{i_{\leq j}}(t) X_{i_{>j}}(t) - X_i(t)\right)\left(X_{i_{\leq k}}(t) X_{i_{>k}}(t) - X_i(t)\right)}{X_i(t)} \, dt \\
&= \int_0^T \sum_{i \in E} \frac{X_{i_{\leq j}}(t) X_{i_{>j}}(t) X_{i_{\leq k}}(t) X_{i_{>k}}(t)}{X_i(t)} \, dt - T, \\
Y_j &= \int_0^T \sum_{i \in E} \frac{X_{i_{\leq j}}(t) X_{i_{>j}}(t) - X_i(t)}{X_i(t)} \, d\widetilde{X}_i(t) = \int_0^T \sum_{i \in E} \frac{X_{i_{\leq j}}(t) X_{i_{>j}}(t)}{X_i(t)} \, d\widetilde{X}_i(t).
\end{aligned}
$$

An alternative model is to set $\rho_j = \rho$ for each $j = 1, \ldots, \ell$ and to construct a single scalar estimator. Then the estimator is

$$\hat{\varrho} = \frac{\displaystyle\sum_{j=1}^{\ell-1} \int_0^T \sum_{i \in E} \frac{X_{i_{\leq j}}(t) X_{i_{>j}}(t)}{X_i(t)} \, d\widetilde{X}_i(t)}{\displaystyle\sum_{j,k=1}^{\ell-1} \int_0^T \sum_{i \in E} \frac{X_{i_{\leq j}}(t) X_{i_{>j}}(t) X_{i_{\leq k}}(t) X_{i_{>k}}(t)}{X_i(t)} \, dt - (\ell-1)^2 T}.$$

These estimators should be corrected as in (26).

4 The effects of natural selection

Methods for inference of recombination can be confounded by natural selection (Reed and Tishkoff 2006; O'Reilly et al. 2008; Peñalba and Wolf 2020). In this section we investigate the effect of selection on $\hat{\rho}_{\mathrm{MLE}}$, for simplicity returning to a two-locus model, though it should be straightforward to extend these results to general multi-locus models.

4.1 Confounding by selection

First consider the following: Suppose that, unknown to the investigator, the two loci are under selection—possibly a complicated type with epistatic interaction. What effect does this have on our estimator for ρ? More precisely, consider a model in which the component of the drift with parameters to be estimated is still $a_{ij}(x; \rho) =$

$\rho(x_i.x._j - x_{ij})$, but the 'known' component of the drift is now

$$
\begin{aligned}
c_{ij}(x) = \ & \frac{\theta_A}{2} \sum_{k=1}^{K} x_{kj}(P_{ki}^A - \delta_{ik}) + \frac{\theta_B}{2} \sum_{l=1}^{L} x_{il}(P_{lj}^B - \delta_{jl}) \\
& + \frac{x_{ij}}{2} \left[\sum_{k=1}^{K}\sum_{l=1}^{L} \left(s_{ij,kl}x_{kl} - \sum_{m=1}^{K}\sum_{n=1}^{L} s_{kl,mn}x_{kl}x_{mn} \right) \right], \\
& i = 1, \dots, K; \ \ j = 1, \dots, L.
\end{aligned}
$$

This is a very general diploid, epistatic model of selection in which the selective advantage of an individual carrying haplotypes (i, j) and (k, l), relative to other individuals, is parametrised by $s_{ij,kl}$. We are interested in the role of selection as a confounder, whereby inference is carried out using the incorrect selection parameters in the dominating measure.

From Eq. (18), $c(\cdot)$ has an effect on $\hat\rho$ only through the term $\mathrm{d}\widetilde{X}_{ij}(t) = \mathrm{d}X_{ij}(t) - c_{ij}(X(t))\,\mathrm{d}t$. Therefore, in a model with selection we should adjust (18) by defining a new estimator

$$
\begin{aligned}
\hat\rho_{\mathrm{sel}} = \ & \hat\rho - \frac{1}{I_T} \int_0^T \sum_{i=1}^{K}\sum_{j=1}^{L} \frac{X_i.(t)X._j(t)}{X_{ij}(t)} \\
& \times \frac{X_{ij}(t)}{2} \left[\sum_{k=1}^{K}\sum_{l=1}^{L} \left(s_{ij,kl}X_{kl}(t) - \sum_{m=1}^{K}\sum_{n=1}^{L} s_{kl,mn}X_{kl}(t)X_{mn}(t) \right) \right] \mathrm{d}t \\
= \ & \hat\rho - \frac{1}{I_T} \int_0^T \sum_{i=1}^{K}\sum_{j=1}^{L} (X_i.(t)X._j(t) - X_{ij}(t)) \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{s_{ij,kl}}{2} X_{kl}(t)\,\mathrm{d}t. \quad (27)
\end{aligned}
$$

The last term, which is linear in the selection parameters, quantifies the error introduced by ignoring selection. However, it demonstrates a remarkable property in the absence of epistasis. In that case we can write $s_{ij,kl} = s_{ik}^A + s_{jl}^B$, where s_{ik}^A is the selection parameter associated with genotype ik at locus A, and similarly for s_{jl}^B. Then Eq. (27) simplifies to

$$
\hat\rho_{\mathrm{sel}} = \hat\rho.
$$

That is, we have an attractive robustness property: if an investigator uses the incorrect model for selection then the estimator $\hat\rho$ is unaffected provided selection is not epistatic. The observed information is also the same. Noting that Theorem 2 continues to hold when $c(\cdot)$ is altered, we conclude that $\hat\rho_{\mathrm{MLE}}$ defined in Sect. 3.2.3 is still the MLE for ρ in the presence of (non-epistatic) selection.

4.2 General confounding

Returning to the general inference problem of Sect. 3.1, we can generalise the previous observations by asking: when does a contribution $c(x)$ to the drift leave the estimator

$\hat{\varphi}$ unchanged? From (5), its contribution to the estimator via $d\widetilde{X}(t)$ will be zero if and only if $a(X(t); \varphi)^\top V(X(t))^{-1} c(X(t))$ does not depend explicitly on φ. When the drift is linear in the parameters as in (11), this requirement becomes

(∗) $\varphi^\top Z^\top V^{-1} c$ does not depend on φ.

If (∗) holds we will say that the estimation problem for $\hat{\varphi}$ is *robust* to the contribution of $c(x)$. In the context of the Wright–Fisher diffusion we have the following result.

Proposition 2 *For a Wright–Fisher diffusion with drift coefficient $\mu(x; \varphi) = c(x) + Z(x)\varphi$ and diffusion coefficient $V = (V_{ij})$, $V_{ij}(x) = x_i(\delta_{ij} - x_j)$, the estimator $\hat{\varphi}$ in (13) is robust to $c(x)$ if and only if*

$$\sum_{i=1}^d \frac{1}{x_i} \frac{\partial a_i}{\partial \varphi_k}(x) c_i(x) = 0, \tag{28}$$

for each $k = 1, \ldots, r$.

Proof We determine (∗) for the first $d-1$ coordinates of the Wright–Fisher diffusion, with $[V^*(x)]^{-1}$ as in (8):

$$\varphi^\top Z(x)^\top [V^*(x)]^{-1} c(x) = \sum_{k=1}^r \varphi_k \sum_{i=1}^{d-1} \left(\frac{1}{x_i} \frac{\partial a_i}{\partial \varphi_k}(x) - \frac{1}{x_d} \frac{\partial a_d}{\partial \varphi_k}(x) \right) c_i(x)$$

$$= \sum_{k=1}^r \varphi_k \sum_{i=1}^d \frac{1}{x_i} \frac{\partial a_i}{\partial \varphi_k}(x) c_i(x),$$

$$c_d(x) := -\sum_{i=1}^{d-1} c_i(x).$$

Since $\frac{\partial a_i}{\partial \varphi_k}(x)$ and $c_i(x)$ do not depend on φ, the above quantity is a linear combination of the φ_k. It does not depend on any φ_k if and only if each of its coefficients is zero, i.e. (28) holds. □

To give another example of the applicability of Proposition 2, consider reversing the roles of selection and recombination, so that we are interested in designing an estimator for (non-epistatic) selection at locus A in the confounding presence of recombination. For simplicity we focus on a genic selection model without mutation:

$$c_{ij}(x) = \rho(x_{i\cdot} x_{\cdot j} - x_{ij}),$$

$$a_{ij}(x; s_1^A, \ldots, s_K^A) = \frac{x_{ij}}{2} \left(s_i^A - \sum_{k=1}^K s_k^A x_{k\cdot} \right), \quad i = 1, \ldots, K; \; j = 1, \ldots, L.$$

In this model we find

$$\sum_{i=1}^K \sum_{j=1}^L \frac{1}{x_{ij}} \frac{\partial a_{ij}}{\partial s_k^A}(x) c_{ij}(x) = \sum_{i=1}^K \sum_{j=1}^L \frac{1}{x_{ij}} \frac{x_{ij}}{2} (\delta_{ik} - x_{k\cdot}) \rho(x_{i\cdot} x_{\cdot j} - x_{ij}) = 0,$$

and (28) holds; by Proposition 2 the estimator $\hat{s}^A$ for $(s_1^A, \ldots, s_K^A)$ is robust to recombination, as we might hope.

Using Proposition 2 it is also possible to show that the following problems are robust:

  (i) Estimation of mutation at locus A when there is selection at locus B,
  (ii) Estimation of genic selection at locus A when there is mutation at locus B;

while the following problems are *not* robust:

 (iii) Estimation of recombination when there is mutation at either locus,
 (iv) Estimation of mutation at locus A when there is mutation at locus B,
  (v) Estimation of mutation at locus A when there is recombination,
 (vi) Estimation of mutation at locus A when there is selection at locus A,
 (vii) Estimation of genic selection at locus A when there is mutation at locus A;

similarly for problems interchanging the two loci. We omit the straightforward calculations. We caution that in the estimation problems above, the likelihood, and thus the estimator $\hat{\varphi}$, will be valid only up to time S as in (25). It is possible to have $I_T = \infty$ even in models without recombination (in particular, we expect two path measures with different mutation parameters to be mutually singular if certain allele frequencies reach 0).

4.3 Joint estimation of recombination and selection

In contrast to Sect. 4.1, one might recognise the possible existence of selection and be interested in constructing a joint estimator for recombination and selection. How does the marginal estimator for ρ from this compare to those already developed? To illustrate the idea, we consider a simple genic selection model at locus A in which only allele k is under selection; that is, $s_i^A = 0$ for $i \neq k$:

$$c_{ij}(x) = \frac{\theta_A}{2} \sum_{k=1}^{K} x_{kj}(P_{ki}^A - \delta_{ik}) + \frac{\theta_B}{2} \sum_{l=1}^{L} x_{il}(P_{lj}^B - \delta_{jl}),$$

$$a_{ij}(x; \rho, s_k^A) = \rho(x_{i.}x_{.j} - x_{ij}) + \frac{x_{ij}}{2}\left(\delta_{ik}s_k^A - s_k^A x_{k.}\right),$$

$$i = 1, \ldots, K; \quad j = 1, \ldots, L.$$

From (13) and (14) we find

$$I_T = \left(\begin{matrix} \int_0^T \sum\limits_{i=1}^{K} \sum\limits_{j=1}^{L} \dfrac{(X_{ij}(t) - X_{i.}(t)X_{.j}(t))^2}{X_{ij}(t)} \, dt & 0 \\[2em] 0 & \dfrac{1}{4} \int_0^T X_{k.}(t)(1 - X_{k.}(t)) \, dt \end{matrix} \right),$$

$$Y = \begin{pmatrix} \int_0^T \sum_{i=1}^K \sum_{j=1}^L \dfrac{X_{i\cdot}(t)X_{\cdot j}(t)}{X_{ij}(t)} \, \mathrm{d}\widetilde{X}_{ij}(t) \\ \dfrac{1}{2} \int_0^T \mathrm{d}\widetilde{X}_{k\cdot}(t) \end{pmatrix},$$

and so $\hat{\varphi} = I_T^{-1} Y$ simplifies to

$$\hat{\varphi} = \begin{pmatrix} \hat{\rho} \\ \hat{s}_k^A \end{pmatrix},$$

where $\hat{\rho}$ is the same estimator as we found in (18) and

$$\hat{s}_k^A = \frac{2(\widetilde{X}(T) - \widetilde{X}(0))}{\int_0^T X_{k\cdot}(t)(1 - X_{k\cdot}(t)) \, \mathrm{d}t}.$$

Setting $\theta_A = \theta_B = 0$ recovers the estimator for selection found by Watterson (1979), up to a choice of timescale. The key point is that the presence of selection as a 'known-unknown' leaves the estimator for ρ unaffected, as is clear from the diagonal nature of I_T. In fact this might have been predicted even earlier: requiring $I_{kl} = 0$ for the off-diagonal entry of an observed information matrix, corresponding to two parameters φ_k, φ_l, is essentially equivalent to the robustness condition (∗). (To see this we identify the $c_i(x)$ term in (∗) with $Z_{il}\varphi_l$, so φ_l parametrises what would have been a confounder in (∗).)

5 Simulation study

In this section we conduct an empirical study of the properties of $\hat{\rho}_{\mathrm{MLE}}$ by simulation. Although there has been recent progress in the development of algorithms for *exact* simulation of certain classes of Wright–Fisher diffusion (Jenkins and Spanò 2017; Griffiths et al. 2018; García-Pareja et al. 2021), these algorithms do not cover the non-reversible diffusions considered in this paper. Instead we resort to simple Euler–Maruyama simulation; that is, to simulate small increments of the diffusion over a fixed, small timestep Δt using the approximation

$$\begin{aligned} X(t + \Delta t) &= X(t) + [c(X(t)) + a(X(t); \varphi)]\Delta t + \sigma(X(t))[W(t + \Delta t) - W(t)], \\ X(0) &= x(0), \end{aligned}$$

where $W(t)$ is the $(d-1)$-dimensional Brownian motion in (23). Integrals involving the sample path of X can be approximated using Riemann sums constructed from the same set of gridpoints.

Because of its singularities at the boundaries of Δ_{d-1}, the Cholesky decomposition of $V(x)$ of Sato (1976) is perhaps not the best choice of $\sigma(x)$ for the purposes of simulation, a point also noted in He et al. (2020). Instead we use a decomposition

found for instance in Pal (2011, 2013):

$$\sigma_{ij}(x) = \sqrt{x_i}(\delta_{ij} - \sqrt{x_i x_j}), \quad i, j = 1, \ldots, d.$$

This formulation has the advantage of being simple, symmetric, bounded in x, and vectorising easily via

$$\sigma(x) = (I_d - \text{diag}(x)1_d)\text{diag}(\sqrt{x}),$$

where I_d is the identity matrix and 1_d is the $d \times d$ matrix of ones. A disadvantage is that it uses a d-dimensional Brownian motion, one dimension more than is necessary for simulation.

Using Euler–Maruyama simulation it is possible to obtain a realisation with $X_{ij}(t + \Delta t) \leq 0$ for some (i, j). If this occurs we set $X_{ij}(t + \Delta t) = 0$, renormalize $X(t + \Delta t)$ so that $X(t + \Delta t) \in \Delta_{d-1}$, and set $I_{t+\Delta t} = \infty$.

In the following we posit a two-locus, diallelic ($K = L = 2$) model with symmetric mutation ($P^A = P^B = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$) and initial condition $X(0) = \begin{pmatrix} 2/5 & 1/5 \\ 1/5 & 1/5 \end{pmatrix}$. The stepsize is set to $\Delta t = 10^{-6}$ and paths are simulated up to a time $T = 1$. We consider two sets of mutation parameters: (i) $\theta_A = \theta_B = 1$, and (ii) $\theta_A = \theta_B = 5$, in order to distinguish models in which the boundaries can or cannot be approached. We explore a variety of recombination parameters, $\rho \in \{0, 0.1, 1, 2.5, 5, 10, 25\}$, and to estimate distributional properties of the estimator we repeat each experiment 100 times.

As an illustration and a check that our implementation is accurate, examples of individual sample paths for $\rho = 5$ are shown in Figs. 1 and 2, together with the accumulated information, I_t, and the evolving error, $\hat{\rho}_{\text{MLE}} - \rho$, as functions of time. As is clear from the Figures, the error is stochastically converging towards 0, with erratic jumps towards 0 in regions where the information accumulates most rapidly. In the second example, in which $\theta_A = \theta_B = 1$, the trajectory for $X_{22}(t)$ wanders sufficiently closely to 0 that $I_t = \infty$ for some $t < T$, whereupon $\hat{\rho}_{\text{MLE}} = \rho$. The distribution of $\hat{\rho}_{\text{MLE}}$ across 100 experiments using these parameters are shown in Fig. 3, with results for further experiments summarised in Table 1.

As is clear from Table 1, $\hat{\rho}_{\text{MLE}}$ is slightly upwardly biased, with the relative bias greater for ρ close to 0. Even with our assumption that the entire sample path is observed, for $\theta_A = \theta_B = 5$ the distribution of $\hat{\rho}_{\text{MLE}}$ is rather flat: for example, when $\rho = 2.5$ the central 90% of its mass is approximately contained in the interval $[0, 15]$. The power to reject the hypothesis $\rho_0 = 0$ at level 5% is consequently poor for small ρ, exceeding 0.5 only for the rows in the table with $\rho \geq 10$.

For $\theta_A = \theta_B = 1$ the picture is very different, demonstrating the sensitivity of $\hat{\rho}_{\text{MLE}}$ to the mutation parameters. We can see that here there is high probability that $\mathbb{P}(\hat{\rho}_{\text{MLE}} = \rho)$, providing very high power to reject $\rho_0 = 0$ even for small $\rho > 0$.

6 Discussion

In this article we have derived an expression for the maximum likelihood estimator of the recombination rate, $\hat{\rho}_{\text{MLE}}$, from a continuously observed diffusion model of
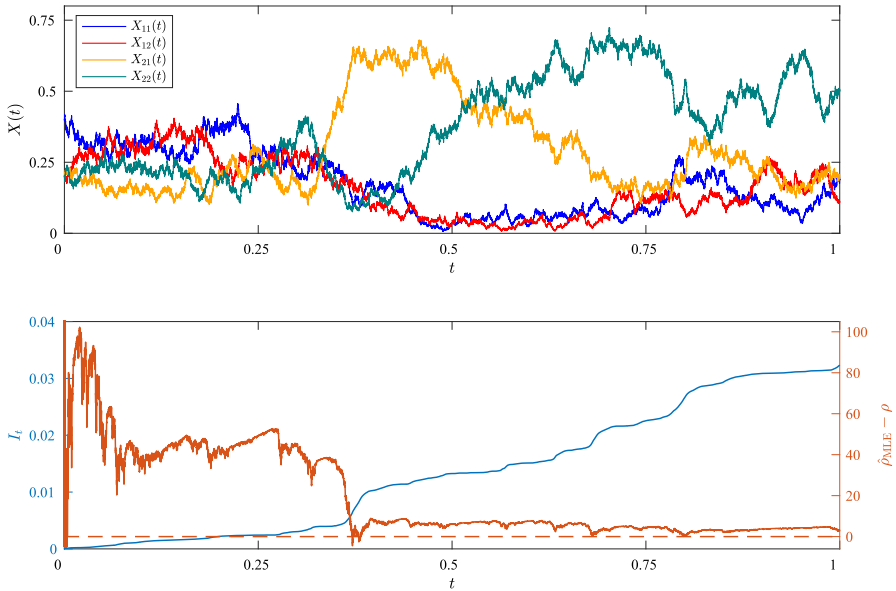
Fig. 1 Example trajectories in a two-locus model with two alleles at each locus, $\rho = 5$, and $\theta_A = \theta_B = 5$. Also shown in the lower plots are the trajectories of $\hat{\rho}_{\mathrm{MLE}} - \rho$ and I_t for this sample path
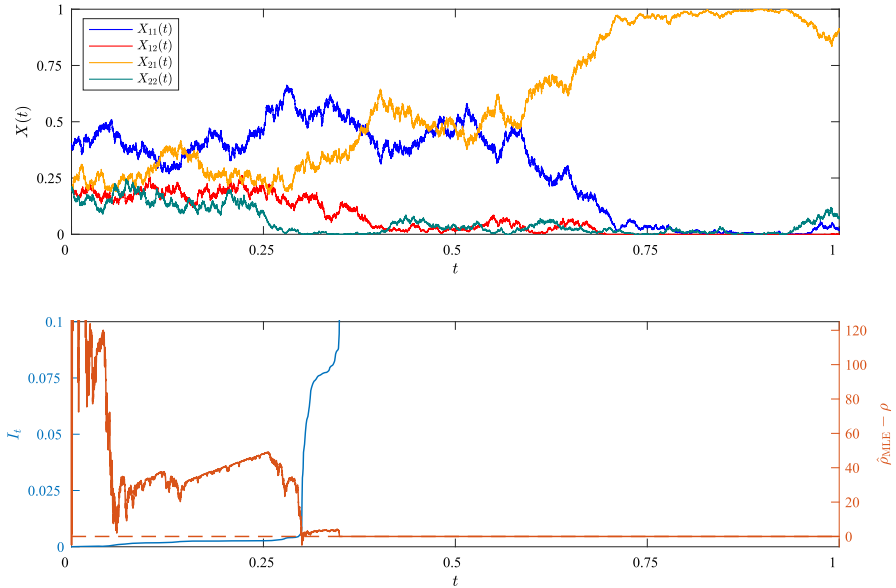

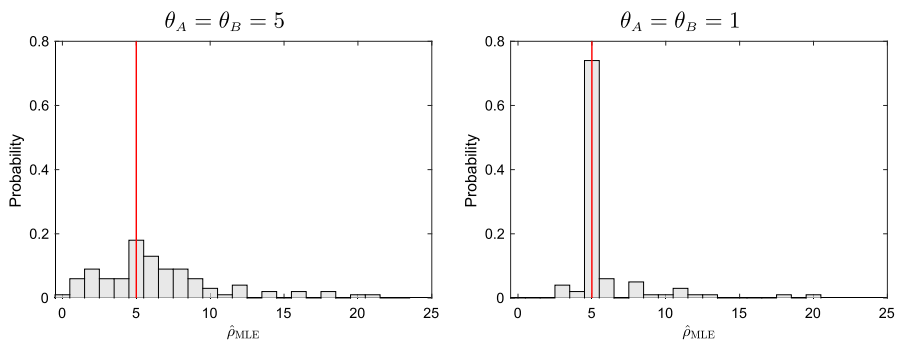**Fig. 2** As Fig. 1 but with $\theta_A = \theta_B = 1$

Fig. 3 Distribution of $\hat{\rho}_{\mathrm{MLE}}$ estimated from 100 replicates. Mutation parameters are $\theta_A = \theta_B = 5$ (left) and $\theta_A = \theta_B = 1$ (right). The true recombination parameter, shown by a red line, is $\rho = 5$

haplotype frequencies. As well as recombination, the diffusion model can incorporate mutation, selection, and genetic drift. We have investigated the empirical properties of the estimator and its robustness to the presence of other processes. We have shown that, contrary to a typical estimator, it is possible to have $\hat{\rho}_{\mathrm{MLE}} = \rho$ with positive probability, and this event is intimately associated with the hitting of the boundary by the diffusion (Theorem 3). Although in that theorem we made some convenient assumptions about the trajectory of $X(t)$, we expect it is possible to refine this result further; indeed we conjecture that $\{I_T = \infty\}$ is equal to the event that one haplotype frequency reaches 0 by time T. This would provide an easy way to check whether $\{I_T = \infty\}$ has occurred.

Although Theorems 2 and 3 are written in statistical language, in terms of estimators and information, we can gain some further intuition by phrasing them in a more fundamental way: it is known that the non-explosion condition (4) holds if and only if $\mathbb{P}_\varphi^{(T)} \ll \mathbb{P}_{\varphi_0}^{(T)}$ (Hobson and Rogers 1998). Thus, when estimating recombination (or mutation, but not selection), hitting a boundary of the diffusion leads to the loss of absolute continuity of one path measure with respect to another. The information I_T provides a natural measure of 'signal-to-noise'. From the point of view of a finite population, although one usually thinks of stochastic effects as being more important when an allele is very rare compared to when it is common, on the contrary what matters in the diffusion limit here is that the variance in offspring distribution (noise) goes to zero at the boundary while the mean detectable effect of recombination (signal) does not. (The qualitatively different behaviour at a boundary between a finite population model and its diffusion limit is also remarked on by Ewens (2004, p180).) This also explains the effects of the mutation rate on estimation of ρ as observed in Sect. 5: higher mutation rates act to push haplotype frequencies toward the interior of the simplex, where the accumulation of information is slower. It also matches biological intuition: if mutation rates are very small, we can reject a null of no recombination by using the four-gamete test on just a sample at a single time point. As mutation rates increase, it is harder to tell apart recurrent mutation from recombination.

Because of the unusual behaviour of the estimator for ρ, we have refrained from providing a detailed description of its asymptotic properties such as local asymptotic

Table 1 Distributional summaries of $\hat{\rho}_{\mathrm{MLE}}$ for $\theta_A = \theta_B = 5$ (top) and $\theta_A = \theta_B = 1$ (bottom): mean, variance, 5th percentile, median, 95th percentile, frequency of zero error, and power to reject $\rho_0 = 0$ at level 5%. Each estimate is based on 100 independent replicates

ρ	$\mathbb{E}(\hat{\rho}_{\mathrm{MLE}})$	$\mathrm{Var}(\hat{\rho}_{\mathrm{MLE}})$	$\mathbb{E}(\hat{\rho}_{\mathrm{MLE},(5)})$	$\mathrm{Median}(\hat{\rho}_{\mathrm{MLE}})$	$\mathbb{E}(\hat{\rho}_{\mathrm{MLE},(95)})$	$\mathbb{P}(\hat{\rho}_{\mathrm{MLE}} = \rho)$	$\mathbb{P}(\Lambda > F_m^{-1}(0.95))$
$\theta_A = \theta_B = 5$							
0	2.69	13.52	0.00	1.06	9.81	0.11	0.10
0.1	2.58	13.57	0.00	0.92	9.99	0.09	0.19
1	2.69	11.15	0.00	1.45	9.55	0.09	0.18
2.5	4.75	20.25	0.00	3.36	14.98	0.08	0.32
5	6.74	20.18	1.41	5.89	16.76	0.07	0.38
10	12.44	39.94	4.26	10.94	26.49	0.04	0.79
25	28.38	69.79	16.89	27.36	44.70	0.02	1.00
$\theta_A = \theta_B = 1$							
0	0.21	1.07	0.00	0.00	0.60	0.95	0.03
0.1	0.34	1.61	0.10	0.10	1.07	0.95	0.98
1	1.22	1.38	1.00	1.00	1.96	0.95	0.98
2.5	3.07	2.78	2.50	2.50	6.69	0.77	0.92
5	5.79	6.55	4.09	5.00	11.00	0.71	1.00
10	10.73	4.88	9.74	10.00	17.29	0.75	1.00
25	26.49	27.72	20.64	25.00	36.30	0.65	1.00

normality (LAN). Using the estimator for the immigration rate of the CBI diffusion as a guide, it should be possible to show that inference for ρ exhibits LAN along the sequence of random times

$$T_n := \inf\{t \in [0, T] : I_t = n\};$$

see Overbeck (1998, §3.4). However, in the case $I_T < \infty$ for each T, finding the asymptotic behaviour of $\hat{\rho}_{\mathrm{MLE}}$ is more involved since we do not know the stationary distribution of X.

Finally, we observe that fundamental quantities appearing throughout this work are

$$\sum_{i=1}^{K}\sum_{j=1}^{L} \frac{(X_{ij} - X_{i.}X_{.j})^2}{X_{ij}}, \quad \text{and} \quad \sum_{i=1}^{K}\sum_{j=1}^{L} \frac{X_{ij} - X_{i.}X_{.j}}{X_{ij}}. \tag{29}$$

The role of X_{ij} in the denominators has been to upweight the importance of those parts of the trajectories where the frequency of haplotype (i, j) is small, since these regions are more informative for ρ. In one sense this is unsatisfactory since the diffusion model is often regarded as an approximation of a discrete population of size N, and behaviour near the boundaries is inappropriate when true frequencies can only be a multiple of $1/N$. One might prefer to replace the estimator $\hat{\rho}$, which integrates each $X_{ij}(t)$ over $[0, T]$, with one that integrates $X_{ij}(t)$ only over some sub-region

$$\{t \in [0, T] : \varepsilon \le X_{ij}(t) \le 1 - \varepsilon\}.$$

Even under this restriction, the quantities in (29) tell us to focus our attention on those regions where the haplotype frequency is far from $1/2$. The normalizations inherent in (29) seem to offer 'natural' new normalizations for the coefficient of linkage disequilibrium, which do not correspond to the usual normalizations found in r^2 and D', for example (Sved and Hill 2018). Exploring the properties of these new summaries of LD will be the subject of future work.

Data availability Data and code used to conduct the simulation study is available at https://github.com/Paul-Jenkins/GriffithsJenkins2023.

Declarations

Conflict of interest The authors have no competing interests to declare.

A Appendix

In this section we study the deterministic mutation-recombination equation

$$\frac{\mathrm{d}x_{ij}}{\mathrm{d}t} = \rho(x_{i\cdot}x_{\cdot j} - x_{ij}) + \frac{\theta_A}{2}(x_{\cdot j}P_i^A - x_{ij}) + \frac{\theta_B}{2}(x_{i\cdot}P_j^B - x_{ij}),$$
$$i = 1, \ldots, K; \ j = 1, \ldots, L. \tag{30}$$

Summing (30) over j yields:

$$\frac{\mathrm{d}x_{i\cdot}}{\mathrm{d}t} = \frac{\theta_A}{2}(P_i^A - x_{i\cdot}), \qquad i = 1, \ldots, K,$$

whose solution is

$$x_{i\cdot}(t) = P_i^A + (x_{i\cdot}(0) - P_i^A)e^{-\frac{\theta_A}{2}t}, \qquad i = 1, \ldots, K.$$

Similarly,

$$x_{\cdot j}(t) = P_j^B + (x_{\cdot j}(0) - P_j^B)e^{-\frac{\theta_B}{2}t}, \qquad j = 1, \ldots, L.$$

Therefore, (30) can be written

$$\frac{\mathrm{d}x_{ij}}{\mathrm{d}t} + \left(\rho + \frac{\theta}{2}\right)x_{ij}(t) = \rho x_{i\cdot}(t)x_{\cdot j}(t) + \frac{\theta_A}{2}x_{\cdot j}(t)P_i^A + \frac{\theta_B}{2}x_{i\cdot}(t)P_j^B =: F(t),$$

with $F(t)$ known and $\theta := \theta_A + \theta_B$. This can be solved via an integrating factor; the solution is

$$x_{ij}(t) = e^{-\left(\rho+\frac{\theta}{2}\right)t}\left[x_{ij}(0) + \int_0^t F(s)e^{\left(\rho+\frac{\theta}{2}\right)s}\,\mathrm{d}s\right]$$

$$= \left(1 - e^{-\left(\rho+\frac{\theta}{2}\right)t}\right)P_i^A P_j^B$$

$$+ (x_{\cdot j}(0) - P_j^B)P_i^A\left(e^{-\frac{\theta_B}{2}t} - e^{-\left(\rho+\frac{\theta}{2}\right)t}\right)$$

$$+ (x_{i\cdot}(0) - P_i^A)P_j^B\left(e^{-\frac{\theta_A}{2}t} - e^{-\left(\rho+\frac{\theta}{2}\right)t}\right)$$

$$+ (x_{i\cdot}(0) - P_i^A)(x_{\cdot j}(0) - P_j^B)\left(e^{-\frac{\theta}{2}t} - e^{-\left(\rho+\frac{\theta}{2}\right)t}\right) + x_{ij}(0)e^{-\left(\rho+\frac{\theta}{2}\right)t}.$$

Substituting the solutions for $x_{i\cdot}(t)$, $x_{\cdot j}(t)$, and $x_{ij}(t)$ into (19) gives an expression for the observed information. After some simplification we arrive at

$$I_T$$
$$= \int_0^T \sum_{i=1}^K \sum_{j=1}^L \frac{D_{ij}(0)^2 e^{-2\left(\rho+\frac{\theta}{2}\right)t}}{D_{ij}(0)e^{-\left(\rho+\frac{\theta}{2}\right)t} + [P_i^A + (x_{i\cdot}(0) - P_i^A)e^{-\frac{\theta_A}{2}t}][P_j^B + (x_{\cdot j}(0) - P_j^B)e^{-\frac{\theta_B}{2}t}]} \, dt.$$

where $D_{ij}(0) = x_{ij}(0) - x_{i\cdot}(0)x_{\cdot j}(0)$. Since the integrand of I_T decays exponentially in t, clearly $I_\infty < \infty$ as before.

References

Barton NH, Etheridge AM, Sturm AK (2004) Coalescence in a random background. Ann Appl Probab 14:754–785

Basawa IV, Prakasa Rao BLS (1980) Statistical inference for stochastic processes. Probability and mathematical statistics. Academic Press, Cambridge

Dehasque M, Ávila Arcos MC, Díez-del-Molino D, Fumagalli M, Guschanski K, Lorenzen ED, Malaspinas A, Marques-Bonet T, Martin MD, Murray GGR, Papadopulos AST, Therkildsen NO, Wegmann D, Dalén L, Foote AD (2020) Inference of natural selection from ancient DNA. Evol Lett 4:94–108

Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley, New York

Ewens WJ (2004) Mathematical population genetics, 2nd edn. Springer-Verlag, New York

García-Pareja C, Hult H, Koski T (2021) Exact simulation of coupled Wright-Fisher diffusions. Adv Appl Probab 53:923–950

Griffiths RC, Marjoram P (1997) An ancestral recombination graph. In: Donnelly P, Tavaré S (eds) Progress in population genetics and human evolution. vol. 87. Springer-Verlag, Berlin, pp 257–270

Griffiths RC, Jenkins PA, Spanò D (2018) Wright-Fisher diffusion bridges. Theor Popul Biol 122:67–77

Hayman E, Ignatieva A, Hein J (2022) Recoverability of ancestral recombination graph topologies. bioRxiv. https://doi.org/10.1101/2021.10.10.463724

He Z, Dai X, Beaumont M, Yu F (2020) Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. Genetics 216:521–541

Hobson DG, Rogers LCG (1998) Complete models with stochastic volatility. Math Finance 8:27–48

Ikeda N, Watanabe S (1977) A comparison theorem for solutions of stochastic differential equations and its applications. Osaka J Math 14:619–633

Jenkins PA, Song YS (2009) Closed-form two-locus sampling distributions: accuracy and universality. Genetics 183:1087–1103

Jenkins PA, Spanò D (2017) Exact simulation of the Wright-Fisher diffusion. Ann Appl Probab 27:1478–1509

Kendall MG, Stuart A, Ord JK, Arnold SF, O'Hagan A (1994) Kendall's advanced theory of statistics. Volume I: distribution theory. Edward Arnold, 6th edition

Kloeden PE, Platen E, Schurz H (2003) Numerical solution of SDE through computer experiments, 3rd edn. Springer, Berlin

Kutoyants YA (2004) Statistical inference for ergodic diffusion processes. Springer-Verlag, London

Liptser RS, Shiryaev AN (2001) Statistics of random processes: I. General theory. Springer, Berlin, Heidelberg

Mijatović A, Novak N, Urusov M (2012) Martingale property of generalized stochastic exponentials. In: Séminaire de Probabilités XLIV, Springer. pp 41–59

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

Ohta T, Kimura M (1969) Linkage disequilibrium due to random genetic drift. Genet Res 13:47–55

O'Reilly PF, Birney E, Balding DJ (2008) Confounding between recombination and selection, and the Ped/Pop method for detecting selection. Genome Res 18:1304–1313

Overbeck L (1998) Estimation for continuous branching processes. Scand J Stat 25:111–126

Pal S (2011) Analysis of market weights under volatility-stabilized market models. Ann Appl Probab 21:1180–1213

Pal S (2013) Wright-Fisher diffusion with negative mutation rates. Ann Probab 41:503–526

Peñalba JV, Wolf JBW (2020) From molecules to populations: appreciating and estimating recombination rate variation. Nat Rev Genet 21:476–492

Reed FA, Tishkoff SA (2006) Positive selection can create false hotspots of recombination. Genetics 172:2011–2014

Revuz D, Yor M (1999) Continuous martingales and Brownian motion, 3rd edn. Springer, Berlin

Sant J, Jenkins PA, Koskela J, Spanò D (2022) Convergence of likelihood ratios and estimators for selection in nonneutral Wright-Fisher diffusions. Scand J Stat 49:1728–1760

Sato K (1976) Diffusion processes and a class of Markov chains related to population genetics. Osaka J Math 13:631–659

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Sved JA, Hill WG (2018) One hundred years of linkage disequilibrium. Genetics 209:629–636

Taylor J (2007) The common ancestor process for a Wright-Fisher diffusion. Electron J Probab 12:808–847

Watterson GA (1979) Estimating and testing selection: the two-alleles, genic selection diffusion model. Adv Appl Probab 11:14–30

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.