Asymptotic behaviour of sampling and backward transition probabilities of the coalescent with parent dependent mutations

Martina Favero\(^1\)\(^*\) and Henrik Hult\(^1\)

\(^1\)KTH, Royal Institute of Technology, Stockholm

Abstract: Several relevant quantities related to the Kingman coalescent are not explicitly known when mutations are parent dependent. Examples include the probability, referred to as the sampling probability, that the typed coalescent, evolving forwards in time, hits a certain configuration, and the backward transition probabilities of its block counting process. In this paper, an asymptotic result is derived for these quantities, as the sample size goes to infinity. It is shown that the sampling probabilities decay polynomially in the sample size with multiplying constant depending on the stationary density of the Wright-Fisher diffusion; and that the backward transition probabilities of the normalised jump chain converge to the limit of frequencies of types in the sample.

MSC2020 subject classifications: Primary 60J90; secondary 60J70, 92D15.

Keywords and phrases: coalescent, population genetics, parent dependent mutations, Wright-Fisher diffusion.

1. Introduction

Given a sample of genetic material from some individuals in a population, the Kingman coalescent [7] models the genealogy of the individuals. When mutations are parent independent (PIM), i.e. the genetic type of the mutated offspring does not depend on the type of its parent, the coalescent can be simulated forward and backward in time, as both its forward and backward transition probabilities are explicitly known. Furthermore, its likelihood function is known as well as several other quantities. However, when mutations are parent dependent, the model becomes more complex and several of the expressions that are explicit in the PIM case become implicit.

The coalescent and its numerous generalisations have been extensively used for inference on genetic data sets, in combination with Monte Carlo methods used to approximate its implicit likelihood function in the non-PIM case, see e.g. [9] for an

\(^*\)corresponding author: mfavero@kth.se
overview. The asymptotic behaviour of coalescent models in relation to samples of large size has recently gained attention due to the large size of modern study samples. While the size of samples continues to increase, due to advancements in DNA sequencing technology, inference methods based on coalescent models struggle to provide reliable results, since the coalescent does not scale well, see e.g. [6]. Moreover, there is a distortion of some of the properties of the coalescent, which is a suitable approximation of certain models, e.g. Wright-Fisher model, provided the sample size is sufficiently smaller than the effective population size, see e.g. [1]. Studying the asymptotic behaviour of the coalescent for large sample size could likely serve as a support in the analysis of these problems.

This paper contributes to the analysis of the asymptotic behaviour of the coalescent by proving an asymptotic result for its sampling and transition probabilities as specified in the following. A large sample is considered to be of the form \( ny^{(n)} \), with \( n \in \mathbb{N} \setminus \{0\} \) and \( y^{(n)} \in \frac{1}{n} \mathbb{N}^d \setminus \{0\} \). As \( n \to \infty \), it is assumed that \( y^{(n)} \to y \in \mathbb{R}^d > 0 \). An example is \( y^{(n)} = \frac{1}{n} \lfloor ny \rfloor \) for some \( y \in \mathbb{R}^d > 0 \). Let \( p(ny^{(n)}) \) be the probability that the block counting process of the Kingman coalescent, evolving forward in time, hits state \( n y^{(n)} \). This corresponds to the likelihood function when a sample \( n y^{(n)} \) from the population at the present time is observed. A recursion formula is known for \( p \), see e.g. [4], but an explicit formula is unknown in the general case of parent dependent mutations. When the sample size is large, it is computationally too expensive to compute \( p \) using the recursion formula. In this paper the asymptotic behaviour of the sampling probabilities \( p(ny^{(n)}) \) is studied, as \( n \) tends to infinity and \( y^{(n)} \) tends to some \( y \in \mathbb{R}^d > 0 \).

When mutations are parent independent, the sampling probabilities are explicitly known and it is easy to show that their asymptotic decay is polynomial of degree \( d - 1 \), where \( d \) is the number of possible allele types, see Section 2 for details. In particular, the degree of the polynomial does not depend on the mutation parameters, which hints that the same behaviour may apply in the general case when mutations are parent dependent. To verify this intuition the following strategy is adopted. First, in Section 2, the sampling probabilities are represented as expectations of multinomial draws from the stationary distribution of the Wright-Fisher diffusion. The resulting expectations are then interpreted as expectations with respect to a sequence of Dirichlet distributions. Finally, a local limit theorem for the sequence of Dirichlet distributions is derived, in Section 3, and used, in Section 4, to prove that the asymptotic behaviour of the sampling probabilities is indeed decaying polynomially with degree \( d - 1 \),

\[
p(ny^{(n)}) \sim \tilde{p}\left(\frac{y}{\|y\|_1}\right)\|y\|_1^{-d} n^{1-d}, \quad \text{as } n \to \infty, y^{(n)} \to y,
\]

where \( \tilde{p} \) is the stationary density of the Wright-Fisher diffusion, \( \|y\|_1 = |y_1| + \cdots + |y_d| \),
and \( \sim \) denotes asymptotic equality in the sense that \( a_n \sim b_n \) if \( \lim_{n \to \infty} a_n/b_n = 1 \).

Establishing the asymptotic decay of \( p(ny^{(n)}) \), enables the study of the asymptotic behaviour of the backward transition probabilities of the normalised jump chain of the block counting process of the Kingman coalescent. Let \( \rho^{(n)}(v|y^{(n)}) \) be the probability of transition from state \( y^{(n)} \) to state \( y^{(n)} - \frac{1}{n}v \), where \( v \) can be equal to \( e_j \), \( j = 1, \ldots, d \), the unit vector, which corresponds to a coalescence of type \( j \), or to \( e_j - e_i \), \( i, j = 1, \ldots, d \), which corresponds to a mutation from type \( i \) to type \( j \) forward-in-time (type \( j \) to \( i \) backward-in-time). In Section 5 it is proved that \( \lim_{n \to \infty} \rho^{(n)}(e_i|y^{(n)}) = \frac{y_i}{\|y\|_1} \) and \( \lim_{n \to \infty} n\rho^{(n)}(e_j - e_i|y^{(n)}) = \frac{\theta y_i P_{ij}}{\|y\|_1^2} \), \( i, j = 1, \ldots, d \), where \( \theta \) and \( P_{ij} \) are mutation parameters.

2. Interpreting the sampling probabilities

Consider the jump chain of the block counting process of the Kingman coalescent with mutations, possibly parent dependent, \( H = \{H(r)\}_{r \in \mathbb{N}} \in \mathbb{N}^d \setminus \{0\} \), evolving forward in time. Then the probability of observing sample \( n \in \mathbb{N}^d \setminus \{0\} \) is given by

\[
p(n) = \mathbb{P}(H(r) = n, \text{for some } r \in \mathbb{N}).
\]

Let \( d \in \mathbb{N}_{>0} \) be the number of possible types, \( \theta \) be the mutation rate and \( P = (P_{ij})_{i,j=1}^d \) be the mutation probability matrix, which is assumed to be irreducible. Assume, as customary, that the probability of the coalescent starting with one individual of type \( i \) is \( \pi_{i} \), where \( \pi = (\pi_i)_{i=1}^d \) is the invariant measure of \( P \), i.e. \( p(e_i) = \pi_i \).

The aim of this section is to provide a representation of the sampling probabilities that is convenient for the study of their asymptotic behaviour. This is a key step, in fact, once the representation is identified, the outline of the proof of the asymptotic behaviour becomes intuitively clear.

To obtain the desired representation, a well-known relationship between the coalescent process and the Wright-Fisher diffusion is used. Let \( X = \{X(t)\}_{t \geq 0} \subset \mathcal{S} = \{x \in [0,1]^d : \sum_{i=1}^d x_i = 1\} \) be the Wright-Fisher diffusion with mutations, that is the strong solution to the following stochastic differential equation,

\[
dX(t) = \mu(X(t))dt + \sigma(X(t))^{1/2}dW(t), \quad t \geq 0,
\]

where \( W = \{W(t)\}_{t \geq 0} \) is a \( d \) dimensional Wiener process, the diffusion coefficient is \( \sigma_{ij}(x) = x_i(\delta_{ij} - x_j) \), \( i, j = 1, \ldots, d \), and the drift is \( \mu_i(x) = \theta \sum_{j=1}^d x_j P_{ji} - \theta x_i \), \( i = 1, \ldots, d \) where \( \theta \) is the mutation rate and \( P \) the mutation probability matrix of the Kingman coalescent.

The two processes, \( H \) and \( X \), are closely related. The sampling probability, \( p(n) \), can be expressed as the expectation of a multinomial draw from the stationary distribution.
of the Wright-Fisher diffusion. Let \( \tilde{X} \) be distributed according to be the stationary distribution of the Wright-Fisher diffusion, then, see e.g. [4],

\[
p(n) = \left( \frac{\|n\|_1}{n} \right) E \left[ \prod_{i=1}^{d} \tilde{X}_i^{n_i} \right].
\] (3)

Obviously, the relation above only holds if the Wright-Fisher diffusion admits a stationary distribution. The existence of a stationary distribution is related to the structure of the mutation mechanism, more precisely, to the existence of an invariant measure for the mutation probability matrix \( P \). When mutations are parent independent, i.e. \( P_{ij} = Q_j > 0, i, j = 1, \ldots, d \), the stationary distribution, not only exists but also has an explicitly known density, a Dirichlet density with parameters \( \theta_Q = (\theta_{Q1}, \ldots, \theta_{Qd}) \), due to [11].

Unfortunately, the PIM case is the only case where the stationary distribution is explicitly known. Furthermore, the Wright-Fisher diffusion has a degenerate elliptic generator, since the diffusion matrix is zero on the boundary of the domain. Thus the classical theory for the study of stationarity and for the study of solutions to the Fokker-Planck equation does not apply and a specific analysis is needed, see [8] for the analysis of the stationary distribution and density of the Wright-Fisher diffusion.

In this paper, the unknown stationary density is used to analyse the asymptotic behaviour of the sampling probabilities in the non-PIM case. In order for the stationary distribution to exist, it is assumed that the mutation transition matrix \( P \) is irreducible. In this case, as shown in [8], the invariant measure of \( P \) uniquely exists, the stationary distribution of the Wright-Fisher diffusion is absolutely continuous with respect to the Lebesgue measure and its probability density function, \( \tilde{p} \), defined on \( \Delta = \{ \mathbf{x} \in [0,1]^{d-1} : \sum_{i=1}^{d-1} x_i \leq 1 \} \), is smooth on \( \Delta^* = \{ \mathbf{x} \in \Delta : x_i > 0, i = 1, \ldots, d - 1 \} \).

Before focusing on the general case, the asymptotic behaviour of the sampling probabilities in the PIM case is analysed, in order to get an insight on what kind of asymptotic decay is to be expected in the general case. Using (3) and the Dirichlet stationary density for the PIM case, an explicit expression for the sampling probabilities is given by

\[
p(n) = \left( \frac{\|n\|_1}{n} \right) \frac{B(n + \theta Q)}{B(\theta Q)} = \frac{1}{B(\theta Q)} \frac{\Gamma(\|n\|_1 + 1)}{\Gamma(\|n\|_1 + \theta)} \prod_{i=1}^{d} \frac{\Gamma(n_i + \theta Q_i)}{\Gamma(n_i + 1)},
\]

where \( B \) is the Beta function and \( \Gamma \) is the Gamma function. Applying Stirling’s formula
to the Gamma functions yields

\[ p(ny^{(n)}) \sim \frac{1}{B(\theta Q)} \left( \left\| ny^{(n)} \right\|_1 + 1 \right)^{\frac{n}{2}} e^{-\left( \left\| ny^{(n)} \right\|_1 + 1 \right)} \]

\[ \cdot \prod_{i=1}^{d} \frac{\left( y_i^{(n)} + \theta Q_i \right)^{n y_i^{(n)} + \theta Q_i} \cdot e^{-\left( n y_i^{(n)} + \theta Q_i \right)}}{\left( n y_i^{(n)} + 1 \right)^{n y_i^{(n)} + \theta Q_i} \cdot e^{-\left( n y_i^{(n)} + 1 \right)}} \]

\[ \sim 1 \frac{1 - \theta}{B(\theta Q)} \left( 1 + \frac{1 - \theta}{\left\| ny^{(n)} \right\|_1 + \theta} \right)^{\frac{n}{2}} \left( \left\| ny^{(n)} \right\|_1 + \theta \right)^{1-\theta} e^{\theta-1} \]

\[ \cdot \prod_{i=1}^{d} \left( 1 + \frac{\theta Q_i - 1}{n y_i^{(n)} + \theta Q_i} \right)^{n y_i^{(n)} + \frac{1}{2}} \left( n y_i^{(n)} + \theta Q_i \right)^{\theta Q_i - 1} e^{1-\theta Q_i} \]

\[ \sim n^{1-d} \left\| y \right\|_1^{1-d} \frac{1}{B(\theta Q)} \prod_{i=1}^{d} \left( \frac{y_i}{\left\| y \right\|_1} \right)^{\theta Q_i - 1} \]

Note that the sampling probabilities decay polynomially with degree \( d - 1 \) independently on the mutation parameters. In the general case one may expect the same type of decay, which is indeed correct, as shown in Section 4. Despite the ease of the intuition, the proof of an asymptotic result for \( p(ny^{(n)}) \) in the general case is more involved, because of the lack of an explicit form for the stationary density of the Wight-Fisher diffusion. It is based on the following key observation. Denote the expectation in (3) by \( k(n) \), i.e., for \( n \in \mathbb{N}^d \setminus \{0\} \),

\[ k(n) = \mathbb{E} \left[ \prod_{i=1}^{d} \tilde{X}_i^{n_i} \right] = \int_{\mathbb{R}^d} \prod_{i=1}^{d} x_i^{n_i} \tilde{p}(x) dx, \quad (4) \]

so that, for \( y^{(n)} \in \frac{1}{n} \mathbb{N}^d \setminus \{0\} \),

\[ p(ny^{(n)}) = \left( \frac{n \left\| y^{(n)} \right\|_1}{ny^{(n)}} \right) k(ny^{(n)}). \]

The function \( k \) admits an interpretation as an expectation with respect to a Dirichlet distribution, provided it is divided by the appropriate normalising constant. That is,

\[ \frac{k(ny^{(n)})}{B(ny^{(n)} + 1)} = \int_{\Delta} f_{D^{(n)}}(x) \tilde{p}(x) dx = \mathbb{E} \left[ \tilde{p}(D^{(n)}) \right], \quad (5) \]
where $D^{(n)}$ is Dirichlet distributed with concentration parameters equal to $ny^{(n)} + 1$, $f_{D^{(n)}}$ its probability density function, and $\mathbf{1}$ is the vector of ones in $\mathbb{R}^d$.

A clarification about notation and state spaces is needed. The space $S$ is used as a state space for the Wright-Fisher diffusion, $X$, and the Dirichlet distributed random vectors, $D^{(n)}$, while their densities are integrated over $\Delta$. When the density functions, which are defined on $\Delta$, are evaluated at a point in $S$, it is implicitly understood that the first $d-1$ components of the vectors are used, the last component being a function of the first $d-1$ components. And vice versa, for a vector in $x \in \Delta$, $x_d$ stands for $1 - \sum_{i=1}^{d-1} x_i$.

Interpreting the function $k$ as in (5), turns out to be an effective tool for analysing the asymptotic behaviour of the sampling probabilities. In fact, it can be proved that the expectation in (5) converges to the constant $\tilde{p} \left( \frac{y}{\|y\|_1} \right)$, while the remaining factors, $\binom{n}{ny^{(n)}} B(ny^{(n)} + 1)$, give rise to the polynomial decay of $p(ny^{(n)})$. In particular, the proof can be divided into two parts. First, the asymptotic behaviour of the Dirichlet random vectors $D^{(n)}$ is studied, using classical tools, see Section 3. Secondly, the function $\tilde{p}$ is applied to $D^{(n)}$ and the asymptotic behaviour of the expectations (5) is studied in Section 4, the main difficulty being that the stationary density is unknown and possibly unbounded near the axes.

3. Central and local limit theorems for a sequence of Dirichlet random vectors

This section is devoted to the study of the asymptotic behaviour of a sequence of Dirichlet distributed random vectors. Although a classical topic, central and local limit theorems are derived to make the paper self contained.

Let $D^{(n)} = (D^{(n)}_1, \ldots, D^{(n)}_d) \in S$ be a Dirichlet distributed random vector with concentration parameters $\alpha^{(n)} \in \mathbb{R}^d_{>0}$ such that

$$\lim_{n \to \infty} \frac{\alpha^{(n)}}{n} = \alpha \in \mathbb{R}^d_{>0}. \quad (6)$$

The sequence of Dirichlet vectors in the previous section corresponds to $\alpha^{(n)} = ny^{(n)} + 1$. A central limit theorem for the sequence $D^{(n)}$ is obtained using the well-known, see e.g. [3], relationship between Dirichlet and Gamma distributions

$$D^{(n)} \overset{d}{=} \frac{G^{(n)}}{\|G^{(n)}\|_1}, \quad (7)$$

where $G^{(n)} = (G^{(n)}_1, \ldots, G^{(n)}_d)$ is a vector of independent Gamma random variables with shape parameters $\alpha^{(n)}_i, i = 1, \ldots, d$ and rate parameter $\beta \in \mathbb{R}_{>0}$. Note that $\beta$
is irrelevant in the transformation from $G^{(n)}$ to $D^{(n)}$. Furthermore, since the Gamma distribution is infinitely divisible, it is possible to write each of the Gamma random variables as a sum of independent Gamma random variables, and thus for each variable $G_i^{(n)}$ the following central limit theorem holds

$$\sqrt{n} \frac{\beta}{\alpha} \left( \frac{1}{n} G_i^{(n)} - \frac{\alpha_i^{(n)}}{\alpha} \right) \xrightarrow{d} N(0, 1), \quad i = 1, \ldots, d.$$ 

Consequently, by (6) and independence of the components of the random vectors $G^{(n)}$, the following central limit theorem for $G^{(n)}$ holds

$$\sqrt{n} \left( \frac{1}{n} G^{(n)} - \frac{\alpha}{\beta} \right) \xrightarrow{d} N_d \left( 0, \frac{1}{\beta^2} \text{diag}(\alpha) \right),$$

where $\text{diag}(\alpha)$ is the diagonal matrix with $\alpha$ on the diagonal. Since $D^{(n)}$ can be written as a function of $G^{(n)}$, as in (7), applying the multivariate delta method yields

$$\sqrt{n} \left( D^{(n)} - \frac{\alpha}{\|\alpha\|_1} \right) \xrightarrow{d} N_d \left( 0, \Sigma(\alpha) \right),$$

with

$$\Sigma(\alpha) = J \left( \frac{\alpha}{\beta} \right) \frac{1}{\beta^2} \text{diag}(\alpha) J \left( \frac{\alpha}{\beta} \right)^T,$$

where $J$ is the Jacobian matrix associated to the transformation in (7). Calculating the Jacobian and multiplying the matrices, which is omitted, yields

$$\Sigma_{ij}(\alpha) = \frac{\alpha_i}{\|\alpha\|_1^3} (\delta_{ij} \|\alpha\|_1 - \alpha_j), \quad i, j = 1, \ldots, d.$$ 

As expected, the covariance matrix above does not depend on the auxiliary rate parameter $\beta$.

Note that, if $\|\alpha\|_1 = 1$, then $\Sigma(\alpha) = \sigma(\alpha)$, where $\sigma$ is the diffusion matrix in (2). Consequently, the Wright-Fisher diffusion matrix can be interpreted as the covariance matrix of the Gaussian limit of a sequence of Dirichlet random vectors.

The Gaussian limiting vector, being the limit of a sequence of Dirichlet random vectors, has a degenerate distribution, as its last component can be expressed in terms of the first $d - 1$ components. Therefore, in order to work with density functions and obtain a local limit result, the state space is restricted to $\mathbb{R}^{d-1}$ by excluding the last component of each vector in the remaining part of this section.
Let \( \phi \) be the pdf of the \( d - 1 \) dimensional centred Gaussian vector with covariance matrix \( \Sigma_{d-1}(\alpha) \), the restriction of \( \Sigma(\alpha) \) to the first \( d - 1 \) components. Let
\[
f_{D^{(n)}}(x) = \frac{1}{B(\alpha^{(n)})} \prod_{i=1}^{d} x_i^{(n)-1}, \quad x \in \Delta,
\]
be the pdf of (the first \( d - 1 \) components) of \( D^{(n)} \), where \( x_d \) stands for \( 1 - \sum_{i=1}^{d-1} x_i \). Then the pdf of \( \sqrt{n} \left( D^{(n)} - \frac{\alpha}{\|\alpha\|_1} \right) \) is given by
\[
\phi_n(u) = n^{-\frac{1}{2}(d-1)} f_{D^{(n)}} \left( \frac{1}{\sqrt{n}} u + \frac{\alpha^{(n)}}{\|\alpha^{(n)}\|_1} \right), \quad \text{for } u \in -\sqrt{n} \left( \Delta - \frac{\alpha^{(n)}}{\|\alpha^{(n)}\|_1} \right),
\]
and equal to 0 otherwise. In general, convergence in distribution does not imply convergence of probability density functions, however, under the conditions of the converse to Scheffe’s theorem, see [2], it does. More precisely, if the sequence \( \phi_n \) is bounded and uniformly equicontinuous, then \( \phi_n \to \phi \), as \( n \to \infty \), uniformly on \( \mathbb{R}^{d-1} \). In the remaining part of this section, the conditions for the convergence of densities are verified, again, it is fundamental that the parameters of the Dirichlet vectors grow to infinity linearly, as assumed in (6).

In order to show boundedness of \( \phi_n \), first notice that, for a fixed \( n \), the maximum of \( \phi_n \) is reached at \( \sqrt{n} \left( \frac{\alpha^{(n)}}{\|\alpha^{(n)}\|_1} - \frac{\alpha^{(n)}}{\|\alpha^{(n)}\|_1} \right) \), and thus
\[
s_n = \sup_{u \in \mathbb{R}^{d-1}} \phi_n(u) = n^{-\frac{1}{2}(d-1)} \frac{1}{B(\alpha^{(n)})} \prod_{i=1}^{d} \left( \frac{\alpha^{(n)}_i - 1}{\|\alpha^{(n)}\|_1 - d} \right)^{\alpha^{(n)}_i - 1}.
\]

Using Stirling’s formula for the Beta function,
\[
s_n \sim \left( \frac{(2\pi)^{d-1}}{\prod_{i=1}^{d} \|\alpha\|_1} \right)^{-\frac{1}{2}} \left( \frac{\|\alpha^{(n)}\|_1}{n} \right)^{\frac{1}{2}(d-1)}.
\]
From the expression above, it is clear that if \( \|\alpha^{(n)}\|_1 \) grows faster than linearly, the sequence is not bounded and if it grows slower it converges to zero. By assumption (6), \( \|\alpha^{(n)}\|_1 \) grows linearly and therefore the supremum \( s_n \) converges to
\[
\left( \frac{(2\pi)^{d-1}}{\prod_{i=1}^{d} \|\alpha\|_1 \|\alpha\|_1^{d-1}} \right)^{-\frac{1}{2}}.
\]
as $n \to \infty$, and thus the sequence of density functions $\phi_n$ is bounded. Note that from the expression in the last display, the determinant of $\Sigma_d^{-1}(\alpha)$ can be identified

$$\det(\Sigma_d^{-1}(\alpha)) = \prod_{i=1}^d \frac{\alpha_i}{\|\alpha\|_1^{d-1}}.$$  

The reasoning leading up to the expression for $\det(\Sigma_d^{-1}(\alpha))$ provides an alternative way to compute the determinant of the diffusion matrix of the Wright-Fisher diffusion, $\det(\sigma_d^{-1}(x)) = \prod_{i=1}^d x_i$.

Following the same type of argument as for the sequence of density functions, it is straightforward to show that also each sequence of their first order partial derivatives is bounded. Furthermore, the density functions $\phi_n$ are smooth on a compact set. Therefore, the sequence of densities is uniformly Lipschitz continuous, which verifies the uniform equicontinuity condition, proving the local limit result as stated in the following proposition.

**Proposition 3.1.** Let $D^{(n)} \sim \text{Dirich}(\alpha^{(n)})$, $\alpha^{(n)} \in \mathbb{R}_d > 0$ such that $\lim_{n \to \infty} \alpha^{(n)}_n = \alpha \in \mathbb{R}_d > 0$. Then the following central limit theorem holds

$$\sqrt{n} \left( D^{(n)} - \frac{\alpha}{\|\alpha\|_1} \right) \xrightarrow{n \to \infty} N_d(0, \Sigma(\alpha)),$$

with

$$\Sigma_{ij}(\alpha) = \frac{\alpha_i}{\|\alpha\|_1^3} (\delta_{ij} \|\alpha\|_1 - \alpha_j), \quad i, j = 1, \ldots, d.$$

Furthermore, a local limit theorem for the corresponding probability density functions, $\phi_n$ and $\phi$, holds

$$\lim_{n \to \infty} \sup_{u \in \mathbb{R}^{d-1}} |\phi_n(u) - \phi(u)| = 0.$$

### 4. Asymptotic behaviour of the sampling probabilities

The goal of this section is to prove an asymptotic result for the sampling probabilities (1). In Section 2, the sampling probabilities have been written in terms of the function $k$ such that $\frac{k(n\gamma^{(n)})}{B(n\gamma^{(n)} + 1)} = \mathbb{E} \left[ \tilde{p}(D^{(n)}) \right]$. The asymptotic behaviour of this expectation is analysed in the following theorem using the results for the sequence of Dirichlet random vectors of the previous section.

**Theorem 4.1.** Let $k$ be defined as in (4). Let $\tilde{p}$ be the stationary density of the Wright-Fisher diffusion (2), assuming the mutation probability matrix, $P$, is irreducible. Let
\( y^{(n)} \in \frac{1}{n} \mathbb{N}^d \setminus \{0\} \) such that \( y^{(n)} \to y \in \mathbb{R}^d > 0 \), as \( n \to \infty \). Then

\[
\lim_{n \to \infty} \frac{k \left( n y^{(n)} \right)}{B \left( n y^{(n)} + 1 \right)} = \tilde{p} \left( \frac{y}{\|y\|_1} \right).
\]

**Proof.** The stationary density \( \tilde{p} \) is smooth on \( \Delta = \{ x \in \Delta : x_i > 0 \} \), so it is bounded on any compact set contained in \( \Delta \). It could, however, explode on the axes. In order to deal with this problem, the domain is divided in two parts. For \( \varepsilon > 0 \), define \( \Delta^\varepsilon = \{ x \in \Delta : x_i \geq \varepsilon, i = 1, \ldots, d-1 \} \). Since \( \frac{y}{\|y\|_1} \in \Delta \), it follows that \( \frac{y}{\|y\|_1} \in \Delta^\varepsilon \), for all \( 0 < \varepsilon \leq \varepsilon_y \), with \( \varepsilon_y = \frac{1}{\|y\|_1} \min_{i=1,\ldots,d-1} y_i \). Fixing \( 0 < \varepsilon \leq \varepsilon_y \), consider

\[
\frac{k \left( n y^{(n)} \right)}{B \left( n y^{(n)} + 1 \right)} = \int_{\Delta} f_{D^{(n)}}(x) \tilde{p}(x) dx = \int_{\Delta^\varepsilon} f_{D^{(n)}}(x) \tilde{p}(x) dx + \int_{\Delta \setminus \Delta^\varepsilon} f_{D^{(n)}}(x) \tilde{p}(x) dx.
\]

To show convergence for the first term in (8), a change of variable \( s \) yields

\[
\int_{\Delta^\varepsilon} f_{D^{(n)}}(x) \tilde{p}(x) dx = \int \frac{1}{\sqrt{n}} \left( \Delta^\varepsilon - \frac{y^{(n)}}{\|y^{(n)}\|_1} \right) \phi_n(u) \tilde{p} \left( \frac{1}{\sqrt{n}} u + \frac{y^{(n)}}{\|y^{(n)}\|_1} \right) du.
\]

By Proposition 3.1 and continuity of \( \tilde{p} \) on \( \Delta \), the integrand above converges pointwise to \( \phi(u) \tilde{p} \left( \frac{x}{\|y\|_1} \right) \). Furthermore, since \( \tilde{p} \) is bounded in \( \Delta^\varepsilon \) by some \( c_\varepsilon \), the sequence is dominated by the sequence \( \phi_n(u) c_\varepsilon \), the integral of which is equal to \( c_\varepsilon \), for all \( n \). Therefore, by the generalized dominated convergence theorem,

\[
\int_{\Delta^\varepsilon} f_{D^{(n)}}(x) \tilde{p}(x) dx \quad \lim_{n \to \infty} \quad \int_{\mathbb{R}^{d-1}} \phi(u) \tilde{p} \left( \frac{y}{\|y\|_1} \right) du = \tilde{p} \left( \frac{y}{\|y\|_1} \right).
\]

It remains to show that the the second term in (8) converges to zero. First note that, if \( x \in \Delta \setminus \Delta^\varepsilon \), then \( x_j < \varepsilon \) for some \( j \), and, since \( x_i \leq 1, i = 1, \ldots, d \),

\[
\prod_{i=1}^{d} x_i^{y_i^{(n)}} < \varepsilon^{y_j^{(n)}} \leq \varepsilon^{y_{min}^{(n)}},
\]

where \( y_j^{(n)} = \min_{i=1,\ldots,d-1} y_i^{(n)} \). Using the inequality above, the fact that the
Theorem 4.2. Let \( \bar{\mathbf{p}} \) be the sampling probability (1) of the block counting process of the Kingman’s coalescent. Let \( \bar{\rho} \) be the stationary density of the Wright-Fisher diffusion (2), assuming the mutation probability matrix \( P \) is irreducible. Let \( \mathbf{y}^{(n)} \rightarrow \mathbf{y} \in \mathbb{R}^d_{>0} \), as \( n \rightarrow \infty \). Then

\[
\lim_{n \rightarrow \infty} n^{d-1} p(n \mathbf{y}^{(n)}) = \left\| \mathbf{y} \right\|_1^{1-d} \bar{\rho} \left( \frac{\mathbf{y}}{\left\| \mathbf{y} \right\|_1} \right).
\]

Proof. Since \( p(n \mathbf{y}^{(n)}) = \left( \left\| \mathbf{y}^{(n)} \right\|_1 \right) k(n \mathbf{y}^{(n)}) \), rewrite

\[
n^{d-1} p(n \mathbf{y}^{(n)}) = n^{d-1} \left( \frac{n \left\| \mathbf{y}^{(n)} \right\|_1}{n \mathbf{y}^{(n)}} \right) B \left( n \mathbf{y}^{(n)} + 1 \right) \frac{k(n \mathbf{y}^{(n)})}{B \left( n \mathbf{y}^{(n)} + 1 \right)}.
\]

Then, note that

\[
n^{d-1} \left( \frac{n \left\| \mathbf{y}^{(n)} \right\|_1}{n \mathbf{y}^{(n)}} \right) B \left( n \mathbf{y}^{(n)} + 1 \right) = n^{d-1} \frac{\Gamma \left( n \left\| \mathbf{y}^{(n)} \right\|_1 + 1 \right)}{\Gamma \left( n \left\| \mathbf{y}^{(n)} \right\|_1 + d \right)} \rightarrow \left\| \mathbf{y} \right\|_1^{1-d},
\]
as \( n \rightarrow \infty \), whereas, by Theorem 4.1, \( \frac{k(n \mathbf{y}^{(n)})}{B(n \mathbf{y}^{(n)} + 1)} \) converges to \( \bar{\rho} \left( \frac{\mathbf{y}}{\left\| \mathbf{y} \right\|_1} \right) \). This completes the proof. \( \square \)
5. Asymptotic behaviour of the transition probabilities

Theorem 4.1 does not only imply Theorem 4.2, but also enables the direct analysis of the asymptotic behaviour of the backward transitions probabilities of the normalised jump chain of the block counting process of the Kingman coalescent. Let \( \mathbf{H}^{(n)} \), \( n \in \mathbb{N} \), be independent copies of \( \mathbf{H} \), and \( \mathbf{Y}^{(n)} = \frac{1}{n} \mathbf{H}^{(n)} \subset \frac{1}{n} \mathbb{N}^d \setminus \{\mathbf{0}\} \). The focus in this section is on the asymptotic behaviour of the transition probabilities

\[
\rho^{(n)}(\mathbf{v}|\mathbf{y}^{(n)}) = \mathbb{P} \left( \mathbf{H}^{(n)}(1) = n\mathbf{y}^{(n)} - \mathbf{v} | \mathbf{H}^{(n)}(0) = n\mathbf{y}^{(n)} \right) = \mathbb{P} \left( \mathbf{Y}^{(n)}(1) = \mathbf{y}^{(n)} - \frac{1}{n}\mathbf{v} | \mathbf{Y}^{(n)}(0) = \mathbf{y}^{(n)} \right),
\]

for \( \mathbf{v} = \mathbf{e}_j, \mathbf{e}_j - \mathbf{e}_i, i, j, = 1, \ldots, d, \) as \( n \to \infty \) and \( \mathbf{y}^{(n)} \to \mathbf{y} \in \mathbb{R}^d_{>0} \). As for the sampling probabilities, the difficulty here is that, when mutations are parent dependent, an explicit expression for the backward transition probabilities is not available. Nevertheless, they can be written in terms of the probabilities \( \pi \), defined below, by applying the Bayes’ formula to the known forward transition probabilities and using a symmetry condition, see [10, 5] for more details,

\[
\rho^{(n)}(\mathbf{v}|\mathbf{y}^{(n)}) = \begin{cases} 
\frac{\pi_j(\mathbf{y}^{(n)} - \mathbf{v})}{\pi_j(\mathbf{y}^{(n)})}, & \text{if } \mathbf{v} = \mathbf{e}_j, \ j = 1 \ldots d, \\
\frac{\pi_{i,j}(\mathbf{y}^{(n)} - \mathbf{e}_j)}{\pi_{i,j}(\mathbf{y}^{(n)})}, & \text{if } \mathbf{v} = \mathbf{e}_j - \mathbf{e}_i, \ i, j = 1 \ldots d, \\
0, & \text{otherwise.}
\end{cases}
\]

The probability \( \pi[j|n], n \in \mathbb{N}^d \setminus \{\mathbf{0}\} \), is the probability of sampling an individual of type \( j \) after sampling \( \|n\|_1 \) individuals with types given by \( n \). Equivalently, it can be written, in terms of the sampling probabilities, \( p \), as

\[
\pi[i|n] = \frac{n_i + 1}{\|n\|_1 + 1} \frac{p(n + e_i)}{p(n)}.
\]

In the PIM case, \( P_{ij} = Q_j, i, j = 1, \ldots, d, \) the sampling probabilities are known, and

\[
\pi[i|n] = \frac{n_i + \theta Q_i}{\|n\|_1 + \theta},
\]

see e.g. [10]. In this case thus \( \pi[i|n\mathbf{y}^{(n)}] \) converges to \( \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \). It turns out that in the general case the limit of the probabilities \( \pi \) is the same. To prove this, \( \pi \) is written in terms of the function \( k \),

\[
\pi[i|n] = \frac{k(n + e_i)}{k(n)}.
\]

The results of the previous section, combined with the expression above, make it straightforward to study the asymptotic behaviour of \( \pi[i|n\mathbf{y}^{(n)}] \), and, consequently, of the backward transition probabilities.
Proposition 5.1. Let \( \pi \) be defined as in (10). Assume the mutation probability matrix \( P \) is irreducible. Let \( \mathbf{y}^{(n)} \in \frac{1}{n} \mathbb{N}^d \setminus \{0\} \), such that \( \mathbf{y}^{(n)} \to \mathbf{y} \in \mathbb{R}_{>0}^d \), as \( n \to \infty \). Then, for \( i = 1, \ldots, d \),

\[
\lim_{n \to \infty} \pi[i|n\mathbf{y}^{(n)}] = \frac{y_i}{\|\mathbf{y}\|_1}.
\]

Proof. Rewrite

\[
\pi[i|n\mathbf{y}^{(n)}] = \frac{k(n\mathbf{y}^{(n)} + e_i)}{B(n\mathbf{y}^{(n)} + e_i + 1)} \frac{B(n\mathbf{y}^{(n)} + 1)}{B(n\mathbf{y}^{(n)} + e_i + 1)} \frac{B(n\mathbf{y}^{(n)} + e_i + 1)}{B(n\mathbf{y}^{(n)} + 1)}.
\]

By Theorem 4.1, \( \frac{k(n\mathbf{y}^{(n)} + e_i)}{B(n\mathbf{y}^{(n)} + e_i + 1)} \) and \( \frac{k(n\mathbf{y}^{(n)} + e_i)}{B(n\mathbf{y}^{(n)} + e_i + 1)} \) both converge to \( \tilde{p}\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_1}\right) \), as \( n \to \infty \). Calculating,

\[
\frac{B(n\mathbf{y}^{(n)} + e_i + 1)}{B(n\mathbf{y}^{(n)} + 1)} = \frac{\Gamma(ny_i^{(n)} + 2)}{\Gamma(ny_i^{(n)} + 1)} \frac{\Gamma(n\|\mathbf{y}\|_1 + d)}{\Gamma(n\|\mathbf{y}\|_1 + 1 + d)} = \frac{ny_i^{(n)} + 1}{n\|\mathbf{y}\|_1 + d},
\]

and letting \( n \to \infty \) concludes the proof. \( \square \)

Knowing the asymptotic behaviour of \( \pi \) directly solves the problem of analysing the asymptotic behaviour of the backward transition probabilities.

Corollary 5.2. Let \( \rho^{(n)} \) be the transition probabilities in (9). Under the assumptions of Proposition 5.1, for \( i, j = 1, \ldots, d \),

\[
\lim_{n \to \infty} \rho^{(n)}(e_j|\mathbf{y}^{(n)}) = \frac{y_j}{\|\mathbf{y}\|_1} \quad \text{and} \quad \lim_{n \to \infty} n\rho^{(n)}(e_j - e_i|\mathbf{y}^{(n)}) = \frac{\theta P_{ij} y_i}{\|\mathbf{y}\|_1^2}.
\]

References

[1] Bhaskar, A., Clark, A. G. and Song, Y. S. (2014). Distortion of genealogical properties when the sample is very large. Proceedings of the National Academy of Sciences of the United States of America 111 2385–2390.
[2] Boos, D. D. (1985). A Converse to Scheffe’s Theorem. The Annals of Statistics 13 423–427.
[3] Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York.
[4] Griffiths, R. C. and Tavare, S. (1994). Simulating Probability Distributions in the Coalescent. Theoretical Population Biology 46 131–159.
[5] De Iorio, M. and Griffiths, R. C. (2004). Importance sampling on coalescent histories. I. Advances in Applied Probability 36 417–433.
[6] Kelleher, J., Etheridge, A. M. and McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology* **12**.

[7] Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications* **13** 235 - 248.

[8] Shiga, T. (1981). Diffusion Processes in Population Genetics. *J. Math. Kyoto Univ. (JMKYAZ)* **21** 133–151.

[9] Stephens, M. (2007). Inference under the coalescent. In *Handbook of Statistical Genetics* (D. Balding, M. Bishop and C. Cannings, eds.) 26, 878-908. Whiley, Chichester, UK.

[10] Stephens, M. and Donnelly, P. (2000). Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62** 605-635.

[11] Wright, S. (1949). Adaption and selection. In *Genetics, Paleontology, and Evolution* (G. L. Jepson, E. Mayr and G. G. Simpson, eds.) 365-389. Princeton Univ. Press.